Shahid Rajaee
Teacher Training University

Journal of
Electrical and Computer Engineering Innovations (JECEI)

Vol. 12 No. 2, Summer-Fall 2024

Journal of Electrical and Computer Engineering Innovations Vol. 12 No. 2, Summer-Fall 2024

JECEI

Shahid Rajaee
Teacher Training University

pISSN 2322-3952
eISSN 2345-3044

Journal of
Electrical and Computer Engineering Innovations (JECEI)

Semiannual Publication

Volume 12, Issue 2, Summer-Fall 2024

Journal of Electrical and Computer Engineering Innovations
http://jecei.sru.ac.ir
EISSN: 2345-3044
ISSN: 2322-3952
JECEI
SRTTU
JECEI

License Holder: Shahid Rajaee Teacher Training University (SRTTU)

Address: Lavizan, 16788-15811, Tehran, Iran.

Journal of Electrical and Computer Engineering Innovations

Vol. 12; Issue 2: 2024

Contents

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Using GPT-2 Model and Hazm Library for Persian Text Generation

M. Solouki, Z. Askarinejadamiri *, N. Zanjani

Computer science of Refah University college, Tehran, Iran.

Article Info

Article History:
Received 10 September 2023
Reviewed 20 October 2023
Revised 18 December 2023
Accepted 25 December 2023

Keywords:
Persian text generation
Hazm library
GPT-2 model

*Corresponding Author's Email Address:
Askarinejad@refah.ac.ir

Abstract

Background and Objectives: This article explores a method for generating Persian texts using the GPT-2 language model and the Hazm library. Researchers and writers often require tools that can assist them in the writing process and even think on their behalf in various domains. By leveraging the GPT-2 model, it becomes possible to generate acceptable and creative texts, which increases writing speed and efficiency, thus mitigating the high costs associated with article writing.

Methods: In this research, the GPT-2 model is employed to generate and predict Persian texts. The Hazm library is utilized for natural language processing and automated text generation. The results of this study are evaluated using different datasets and output representations, demonstrating that employing the Hazm library with input data exceeding 1000 yields superior outcomes compared to other text generation methods.

Results: Through extensive experimentation and analysis, the study demonstrates the effectiveness of this combination in generating coherent and contextually appropriate text in the Persian language. The results highlight the potential of leveraging advanced language models and linguistic processing tools for enhancing natural language generation tasks in Persian. The findings of this research contribute to the growing field of Persian language processing and provide valuable insights for researchers and practitioners working on text generation applications in similar languages.

Conclusion: Overall, this study showcases the promising capabilities of the GPT-2 model and Hazm library in Persian text generation, underscoring their potential for future advancements in the field This research serves as a valuable guide and tool for generating Persian texts in the field of research and scientific writing, contributing to cost and time reduction in article writing.

Introduction

Text generation is a field within natural language processing (NLP) that combines computational linguistics and artificial intelligence to automatically create written text to meet specific communication needs. Advances in AI have greatly improved our ability to generate targeted texts for various purposes. The study of text generation dates back to the 1970s, with early work by Golding focusing on natural language generation (NLG) based on deep conceptual understanding. Substantial progress was made in this field during the 1980s.

McDonald viewed text generation as a decision-making problem [1]. Later, the concept of Natural Language Generation (NLG) was introduced, referring to the software process that converts data into human-like language. In 2019, OpenAI released an AI engine called GPT-2, capable of generating text. GPT-2 could process text based on specific guidelines and even had an experimental online tool called Talk To Transformer. This

AI engine possessed powerful text generation and prediction capabilities, particularly suited for smartphones. To improve GPT-2, a machine learning system was trained using around 8 million web pages [2]. This system can identify missing words in text and replace them appropriately.

Text production is essential in any language, and text generation serves various applications. GPT-2 allows or assists writers and typists by completing sentences and generating ideas across different domains. GPT-2 has been used for text generation in multiple languages, including English, German, and Bengali.

GPT-2's language modeling algorithm demonstrates coherent semantic capabilities and possesses a unique generative feature found in human language. Language generation involves producing new linguistic expressions based on existing language rules, and GPT-2 excels in this aspect [3]. The training of this language model involved utilizing over 40 gigabytes of web data and nearly 1.5 billion parameters of text structures.

GPT-2 exhibits a wide range of capabilities, such as generating high-quality conditional text samples when given input. It outperforms other language models in specific domains like Wikipedia, news, or books without the need for extensive training datasets [4], [5]. GPT-2 can learn. In tasks such as comprehension, summarization, and translation, GPT-2 learns these tasks from raw text without relying on specific training data. By using GPT-2, we can facilitate the writing process, reduce energy consumption, increase speed, and improve efficiency. For instance, when writing abstracts or articles that would otherwise take hours, GPT-2 helps reduce time, cost, and energy consumption, highlighting the necessity and importance of this field. GPT-2 also contributes to cost reduction in scientific production, which can be a costly endeavor [6], [7]. A similar example is the auto-complete feature when individuals search on Google. It helps users reach their intended search queries faster by suggesting word completions, reducing costs and time spent on multiple searches.

Guessing the next word, which may occur billions of times, has enabled this AI-based system to produce text that appears to be written by a human. The generated texts from this text processing engine are grammatically correct, have a consistent and coherent theme, and possess an engaging style of expression. Therefore, intelligent text generation has been performed in different languages using specific algorithms and artificial intelligence algorithms. In this research, we aim to enable text generation in the Persian language.

Related Work

A. Language Modelling

Using a starting word to create meaningful sentences is crucial in natural language processing. The question of whether machines can think and be creative like humans is what we aim to answer in sentence construction. We train systems for specific tasks and apply them in natural language processing to tackle challenges in sentence generation, such as text summarization, machine translation, and automated question answering.

Language modeling (LM) involves predicting the next word in a given sequence of data. Researchers have recently taken a keen interest in LM in natural language processing. Language models can be divided into count-based models and continuous space models [8].

Count-based models use statistical formulas to describe the language model and construct the joint probability distribution of word sequences. An example is the n-gram model, which predicts one word at a time based on the Markov assumption. The probability of a word sequence is calculated as the product of word probabilities based on previous words, with a history limited to a certain number of words.

Continuous space models include Neural Language Models (NLMs). There are two main types of NLMs: feed-forward neural networks and recurrent neural networks. Feed-forward neural networks address data sparsity, while recurrent neural networks overcome the limitation of context. Recurrent neural networks have shown advanced performance. Feed-forward neural networks have hierarchical probabilistic models and HLBL models that improve training speed.

Various methods exist for text generation, including Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) models, encoder-decoder frameworks, and sequence-to-sequence models. RNNs and LSTMs generate sentences starting from a given word, while encoder-decoder models encode sentences into fixed-length vectors and then use the vectors to generate sentences. However, Transformer models like OpenAI GPT-2 and BERT have also proven effective in text generation.

GPT-2 language models consist of three components. "Generative" means the model is trained to predict or generate the next element in a sequence, even non-sequentially [9]. Prediction is a crucial task in scientific and industrial applications. GPT-2 can generate diverse and meaningful text when provided with raw textual data, without human intervention. It excels in language tasks like machine translation. GPT-2 is based on the Transformer architecture, which is known for language modeling, translation, and classification tasks, as well as computational efficiency [10]. Transformers differ from previous technologies like RNNs and LSTMs by not relying on previous states, resulting in faster computation. Machine translation involves converting a sentence from one language to another, and Transformers, with their

encoder and decoder sections, facilitate this process. GPT-2 has been widely used for text generation in various languages, including Chinese, Japanese, German, and Persian (referred to as "Bolbol-e-Zaban" in the presented work for generating poems and verses using AI).

B. Reviewing Text Generation in Different Language

In Chinese text generation, words are masked in sentences, and the BERT model predicts the masked labels [11]. GPT-2 is built with a Transformer decoder module that takes a starting token as input and generates tokens one at a time, creating the sequence. BERT, on the other hand, uses a bidirectional encoder based on the Transformer and predicts masked labels using its Masked Language Model (MLM) task. Generating sentences based on starting words is the main task when using BERT.GPT-2, an advanced transformer-based model, was used to generate artificial text samples by training on various user prompts as input. The original version had 1.5 gigabytes of parameters, and a smaller version with 117 megabytes of parameters was later released for fine-tuning on custom text datasets. In this particular language generation task, Jpop text was used as the training data, with each text delimited by <I endoftext I>.

In paper under title of “Toward Russian Text Generation Problem Using OpenAI’s GPT-2” discuss around text generation problem in Russian language. This work focuses on Natural Language Generation (NLG) and explores modern approaches using deep neural networks. Specifically, it examines popular NLG solutions based on the Transformers architecture with pre-trained models like GPT-2 and BERT. The challenge lies in the limited availability of Russian language models that can generate text within specific subject areas. The objective of this study is to develop a model capable of generating contextually coherent narrow-profile text in Russian. As part of the study, a model was trained to generate coherent articles in Russian within a specific subject area, along with a software application for interacting with it [12]-[14]. In this study, we try do same process to generate Persian text [12].

The field of question answering (QA) and text generation using BERT and GPT-2 transformers is a specialized area within information retrieval systems. It involves formulating queries in natural language and retrieving the most accurate or closest answer. The primary goal of QA systems is to provide concise answers to questions, rather than presenting a list of relevant documents. On the other hand, text generation focuses on producing coherent English text by predicting the next sentence or generating content based on previous words [15], [16].

The latest version (3.1.0) of Huggingface transformers was used to generate German text. The new Trainer class was utilized, and the GPT-2 model was fine-tuned using German recipes from chefkoch.de. German text data for GPT-2 training was obtained from the Huggingface model hub, specifically from German recipe datasets with metadata from chefkoch.de [11]. The aim was to fine-tune GPT-2 using recipe descriptions and generate new recipes [14], [17], [18].

In another study, researcher focus t generate Sepedi text. This study focuses on developing and analyzing a language model called Generative Pre-Trained Transformer 2 (GPT-2) to generate phrases in the Sepedi language. Sepedi is a language that lacks resources and poses challenges due to its orthographic representation. The GPT-2 transformer requires large datasets and advanced computational resources. The researchers used the National Centre for Human Language Technology (NCHLT) Sepedi text dataset, which is unstructured. Despite working with a small dataset, the developed text generation model achieved a low loss value of 2.36. The generated text produced by this model is syntactically correct but contains some grammatical errors. Overall, the model outperformed previous Sepedi text generation models that used transformer-based techniques [19]-[21].

Recently, The Guardian published an article about GPT-3, stating that it can write a 500-word article using simple and clear language especially in English. The article emphasized that there is no need to fear artificial intelligence. The potential impact of machine learning across various fields, such as medicine, law, sociology, and communications, in the coming decade is expected to be unprecedented [22], [23]. GPT-3 generates text that is statistically appropriate. For example, the initial description of a car accident resembles the opening sentence of Jane Austen's Sanditon. This is the latest achievement of GPT-3, as described in the original text (Fig. 1 & Fig. 2).

A gentleman and a lady travelling from Tunbridge towards that part of the Sussex coast which lies between Hastings and Eastbourne, being induced by business to quit the high road and attempt a very rough lane, were overturned in toiling up its long ascent, half rock, half sand. The accident happened just beyond the only gentleman’s house near the lane—a house which their driver, on being first required to take that direction, had conceived to be necessarily their object and had with most unwilling looks been constrained to pass by. He had grumbled and shaken his shoulders and pitied and cut his horses so sharply that he might have been open to the suspicion of overturning them on purpose (especially as the carriage was not his master’s own) if the road had not indisputably become worse than before.

Fig. 1: text generation of description of a car accident.

We provided GPT-3 with the first sentence and compared the previous text draft and text generation

using GPT-2 by referring to Fig. 1 and Fig. 2. Both GPT-3 and GPT-2 function like search engines, where the generated text by GPT-3 is an additional line. Similar to how Google provides relevant answers to our queries without understanding them, GPT-2 follows the given text and seamlessly continues the sequence of our words, extending it to a specified length [24]-[26].

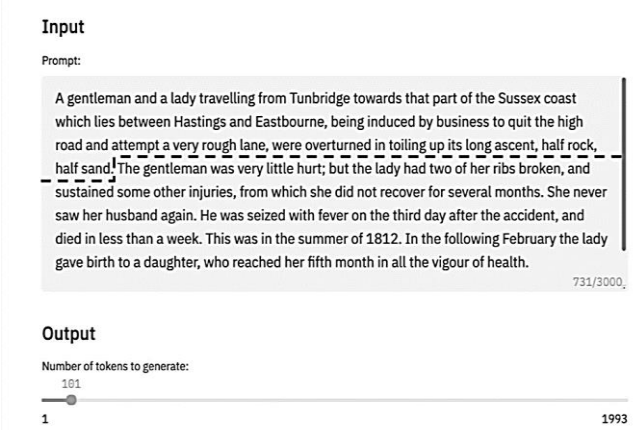


Fig. 2: Text generated by gpt-2.

This study involved focus of GPT-2 in generating text across various languages. Through extensive experimentation and evaluation, it was observed that larger datasets produced more accurate and meaningful samples compared to smaller ones. Moreover, the application of specific language-specific libraries, such as the Hazm library for Persian, significantly improved the quality of generated text, ensuring correct spelling and adherence to linguistic rules. The study successfully demonstrated the capability of GPT-2 in generating content-rich text in Persian, highlighting its potential as a valuable tool for text generation in the Persian language. With the advancements in language models like GPT-2, the field of text generation is continuously evolving and promising, providing new possibilities for researchers and language enthusiasts alike.

C. Language Model in Different Languages

Having explored text generation in different languages and discussed various language models, let's now delve into the discussion of text generation using different decoding methods.

Several algorithms are used for text generation, including [27]-[30]:

1. Greedy search: Choosing the next word based on the highest probability, which is the most probable word.

2. Beam search: Maintaining a set number of hypotheses at each step and selecting the hypothesis with the highest probability. This method reduces the risk of losing highly probable word sequences. Transformers can utilize beam search by setting parameters like num_beams > 1 and early_stopping = True to stop generation when all beam hypotheses reach the end of sentence (EOS) token.

Recently, beam search has been criticized for not being the optimal choice due to the following reasons:

- Beam search works well when the desired generation length is predictable, such as in machine translation or text summarization tasks.

- Beam search struggles with generating repetitive text, and controlling it through measures like n-grams or penalties is challenging.

To address these limitations, Holtzman et al. proposed an alternative method known as sampling. In its simplest form, sampling involves randomly selecting the next word based on its conditional probability distribution. One sampling scheme called Top-K sampling was introduced, where the next words are filtered based on the top-K highest probabilities, and the probability mass is redistributed only among those K words. GPT-2 adopted this sampling scheme, contributing to its success in generating stories.

The word domain used in the example above was expanded from 3 words to 10 words to better illustrate Top-K sampling. One issue with traditional sampling is that it lacks dynamic adaptability to the number of words filtered from the probability distribution of the next word (P(w | w1:t-1)). To address this, Top-p sampling is employed. Instead of sampling solely from the top-K probable words, Top-p sampling selects from a smaller set of words whose cumulative probability exceeds a certain threshold, denoted as p. The probability mass is then distributed among this set of words. This approach allows the size of the word set to dynamically change based on the probability distribution of the next word [9].

In general, when using a text generation algorithm, it is important to first determine the objectives of the text generation task and then experiment with different algorithms to select the most suitable one. In this study, various language models were explored in different languages, including English, German, Chinese, Russian and Sepedi. Through extensive analysis, it was evident that GPT-2 stands out as a highly effective language model for text generation in Persian as well. By fine-tuning GPT-2 with the Hazm library, the generated Persian text exhibited coherence, accuracy, and adherence to linguistic rules. The successful application of GPT-2 in Persian text generation opens new horizons for researchers, writers, and content creators in the Persian language domain. The integration of language-specific libraries, like Hazm, proved crucial in achieving superior results, ensuring that the generated text reflects the linguistic nuances and intricacies unique to Persian. Overall, the study's findings demonstrate the immense potential of GPT-2 in generating high-quality and contextually relevant text in Persian, showcasing its versatility as a language model for diverse linguistic applications.

Proposed Method

For this study, our goal was to explore relevant articles on text generation from resources like Google Scholar, ScienceDirect, and websites such as GitHub, Hugging Face, and Medium. The aim was to find existing libraries for text generation in the Persian language. To achieve this, we utilized the GPT-2 tool and online resources like Google Colab. We also made use of libraries such as regex, Transformer, and Hazm. Text generation was performed using the GPT-2 tool, where sentences were initiated with a starting word and tokenization was applied using GPT-2 and various libraries and tools.

As machine learning algorithms require input data, it was necessary to prepare a dataset. Google Colab, a free cloud service provided by Google, was used for implementation and text generation. It allows programming in Python and offers the capability to install and work with various Python packages. The proposed method of this paper combines various deep learning libraries and text processing tools to build and evaluate language models for text generation in the Persian language.

1. Pipeline: The method utilizes the Pipeline feature available in libraries like Hugging Face's Transformers. The Pipeline offers a convenient API for various natural language processing tasks, including text generation. It allows for easy integration of pre-trained language models and tokenizers, simplifying the process of building language models for specific tasks.

2. PyTorch: To build one of the language models, the proposed method employs PyTorch, a flexible and powerful deep learning library. PyTorch enables the creation and training of custom neural network architectures tailored for Persian text generation. The model is trained on a sizable dataset of Persian text using PyTorch's efficient optimization and GPU capabilities.

3. Tokenizer: The proposed method incorporates tokenization libraries compatible with PyTorch and other deep learning frameworks. Tokenizers are essential for breaking down text into smaller units and encoding them as input for language models. These libraries facilitate efficient tokenization and data preparation for training the models.

4. Regex: The method also utilizes regular expressions (regex) for text preprocessing tasks. Regex is valuable for pattern matching and replacing specific strings, allowing for data cleaning and normalization before feeding the text into the language models. This step ensures that the input data is in the appropriate format for training.

5. Transformer: The proposed method focuses on using the Transformer architecture for language modeling. Transformers have proven to be highly effective in natural language processing tasks, including text generation. The Transformer's self-attention mechanism enables the model to capture contextual dependencies effectively, leading to more coherent and contextually relevant text generation.

6. Hazm Library: For language-specific preprocessing and normalization tasks in Persian, the Hazm library is utilized. Hazm offers features such as sentence and word segmentation, stemming, and handling half-space characters, specifically designed for Persian text processing. The integration of the Hazm library ensures that the language model can handle the unique linguistic characteristics of Persian.

By combining the Pipeline, PyTorch, tokenizer libraries, regex, Transformer architecture, and the Hazm library, the proposed method aims to create comprehensive language models for Persian text generation. This multi-tool approach will enable a thorough evaluation of the models' performance and their effectiveness in generating coherent and contextually appropriate text in the Persian language. The paper's findings will provide valuable insights into the strengths and limitations of each tool for text generation tasks in Persian, guiding researchers and developers in building powerful language models for Persian natural language processing.

Average consumed resources were calculated in the middle of Google Colaboratory after many launches of each of the original models. The results can be seen in Table 1.

Table1: consume source

Number of articles	Occupied disk space, GB	RAM, GB	GPU memory, GB
Gpt-2 100	0.7	2.63	2.52
Gpt-2 500	1.32	3.56	4.63
Gpt-2 1000	4.5	6.53	8.3
Gpt-2 5000	5.3	8.34	9.33
Gpt-2 20000	8.26	11.3	12.5

Implementation Results and Discussion

Generating Persian text with artificial intelligence tools is feasible, and this section examines models with different datasets and sizes. In this study, the execution method involves removing HTML tags using the Beautiful Soap library from a dataset of 100 articles, leaving only the Persian text. Next, the data is normalized using the Hazm library. Certain characters in the dataset are replaced using the regex library, such as "ی" and "ک". Tags like <Num> and <LAT> are used to represent English letters and numbers. The <eos> tag indicates the end of sentences, while the [STARTABS] and [ENDABS] tags mark the beginning and end of phrases to improve data structure.

Model Training

After installing the required libraries (transformer, tokenizer, pipeline, etc.), the model is retrained. It is then trained with different initial inputs, such as "پژوهشگران", "در این کار تجربی ابتدا" and "مطالعه برخی خواص" to generate samples and perform text generation (shown in Table 2 and Table 3). There are also several launches of all models with different input phrases such as "پژوهشگران "،"دراین کارتجربی ابتدا" and "مطالعه برخی خواص".

Table 1: comparing beam search and to-k search result

Beam search	Top-k search
آب مروارید کودکان شایع می باشد. گروه آزمایشی در دو گروه آزمایشی و کنترل، مکمل ویتامین دریافت کردند و گروه کنترل نیز مکمل ویتامین دریافت کردند. گروه آزمایشی در دو گروه آزمایشی و کنترل، مکمل ویتامین دریافت کردند و گروه کنترل نیز مکمل ویتامین دریافت کردند. در این مطالعه مقطعی، از کودکان مبتلا به در دو گروه آزمایشی و کنترل، خون گیری به عمل آمد. کودکان گروه آزمایشی در دو گروه آزمایش و کنترل، به طور تصادفی به دو گروه آزمایش و کنترل تقسیم شدند. کودکان گروه آزمایشی در دو گروه آزمایش و کنترل، به طور تصادفی به دو گروه آزمایش و کنترل تقسیم شدند	آب مروارید کودکان با روش توصیفی تحلیل کوواریانس از دید ساله شهر Nکودکان زیر تبریز را مورد بررسی قرار داده اند. جمع آوری داده ها بر اساس روش توصیف کوواریانس در دو مرحله تجزیه شده و با استفاده از نرم افزار و آزمون های تی تست انجام گردید. یافته ها: نتایج نشان داد که بین سبک های دلبستگی والدین و هوش هیجانی کودکان رابطه مثبت و معنادار وجود دارد؛ اما در سطح معناداری هوش هیجانی، هوش عاطفی رابطه معنی دار را با هوش هیجانی کودکان برقرار نمی کند. نتیجه گیری: ها بیانگر نقش یافته میانجی گری والدین و هوش هیجانی است و می تواند حاکی از نقش والدین در رابطه عاطفی بهتر کودکان باشد.به منظور تعیین ارزش ویژه هوش هیجانی در کودکان تیزهوش، نتایج یک پژوهش توصیفی تحلیلی شهر تبریز انجام بر روی شد. جامعه آماری پژوهش، کودکان تیزهوش و کم شنوای منطقه۳ ساله شهر تبریز بودند.

In order to compare suitable algorithm, in one example, the beam search algorithm with the num_beams parameter did not produce satisfactory results, as it led to word repetitions in the generated samples (Table 1). Subsequently, the top-k and top-p algorithms were used, resulting in samples with lower error rates. Table 1 demonstrates the output obtained using the top-k algorithm.

Initially, a 500 parameter-model was evaluated with different inputs, and the results were analyzed. The 500-model, when given that input produced the samples shown in Table 3. The section compares the 500-model with and without the Hazm library applied. The samples indicate that the 500-model with the Hazm library produces more accurate sentences in terms of grammar and appearance (Table3).

Table3: Comparing Model with and without Hazm

Start word	Model without Hazm	Model with Hazm
مطالعه برخی خواص500	مطالعه برخی خواص شربت سنجد و سنجد سنجد سنجد و سنجد سنجد	مطالعه برخی خواص منحصر بفرد گل میخک می توانید از آن به عنوان یک ماده ضد التهاب و ضد سرطان نام ببرید . بررسی ها مشخص کردند میخک منبع خوبی از آنتی اکسیدان به عنوان یک ضد التهاب قوی است که التهاب را کاهش می دهد و از ابتلا به بلاک های خونی جلوگیری می کند. همچنین، از آن به عنوان یک گیاه ضد انعقاد خون ، نوعی آنتی بیوتیک طبیعی نام می برند. میخک یکی از موثر ترین و مفید ترین داروهای طبیعی شناخته شده است و همچنین، از آن به عنوان یک ماده ضد حساسیت و کاهش دهنده التهابات بدن نام برده میشود. این گیاه یکی از بهترین عصاره های گیاهی شناخته شده در درمان آلرژی های غذایی و اگزما میباشد . ماده ای که در این گیاه می توان یافت .
پژوهشگران 500	پژوهشگران با بررسی و مطالعه بر روی تعداد مشخصی از میکروب ها موفق شدند به نتایج جالب توجهی برسند . در حالی که با بررسی میکروبی به نام شOه و مقایسه باکتری های Gه و میکروسینیتی دریافتندe به طور مرتب آب و مایعات بیشتری در دهان خود میبلعد. به گفته کارشناسان میکروب شbه در برابر Bه شXه با قرار گرفتن در معرض آب کمتری مینوشد. در حالیmه آب بیشتری در دهان خود میسوزاند شه٪ آب بیشتری در دهان میبلعد.	پژوهشگران با استفاده از تکنیک هایی خاص، به بررسی رفتار ژنتیکی باکتری های روده پرداختند. این باکتری ها به شکل خاصی نسبت به مواد آنتی باکتریایی واکنش نشان می دهند و به طور معمول باعث بروز علائم مسمومیت می شوند. آزمایش آنها توانست با انجام مطالعاتی بر روی باکتری های روده باعث بررسی دقیق آن بیماری می شود. در طی این تحقیق باکتری های روده ای مورد بررسی قرار گرفتند. نتایج این بررسی در نشریه های معتبر علوم مولکولی دنیا منتشر شد و به عنوان یکی از مقالات مهمی که با موفقیت چاپ گردید، در نظر گرفته شد. نتایج این تحقیق نشان داد باکتری های روده ای نسبت به مواد آنتی باکتریالی مقاومند. هدف از این مطالعه اندازه گیری آنتی باکتریال و آنزیمی بود که در باکتری های روده وجود دارد. میکروب های روده در طی سالهای متوالی نسبت به این مواد حساسیت نشان داده و به باکتری های گرم منفی واکنش نشان داده و موجب بروز علایم مسمومیت میشوند. اکثر باکتری های روده مقاوم هستند. لذا برای کاهش و یا دفع باکتری های مفید روده اثر میگذارند. هدف از این تحقیق مطالعه تأثیر میکروبیتاسیون بر باکتری های روده بود.

Next, the 1000 parameter -model was considered in two scenarios: with and without the Hazm library. The model was tested with different inputs. Ordered and disordered English and Persian letters in the generated sample, leading to some words and sentences being nonsensical. However, this issue is resolved in the subsequent example by applying the Hazm library. The 1000-model, when not normalized and without the Hazm library, produced nonsensical words and fragmented English words in the samples. In contrast, the 1000-model with the Hazm library applied generated samples with correct spelling and linguistic aspects. It's important to note that the generated samples effectively convey their meanings. They consist of coherent and relevant words and sentences related to the discussed topics, indicating accurate content. For example, we present a sample from the10000 and 50,000 parameter-model. We then compare the 10,000_Hazm model with the Bolbolzaban model as a persian text generator model. Table 4 illustrate the output of the Bolbolzaban model for generating text with the different inputs (Table 4). Initially, the "Bolbolzaban" model was used for training the data, and the generated samples from the Bolbolzaban model were compared to the 10,000-case data model as described earlier.

Table 4: comparing Hazm Model with Bolbolzaban

	Hazm	Bolbolzaban
دراین کارتجربی 10000	دراین کارتجربی ، لایه های نازک اکسید مس- و روی در آب تهیه شده است. در این مقاله ، بررسی مراحل انجام فرایند جوشکاری سیمی، در قالب طرح توجیهی و عملی ارایه شده است. این مطالعه به منظور بررسی ویژگیها و مراحل مختلف مراحل جوشکاری سیمی در قالب طرح توجیهی و عملی ارائه شده است تا میزان استفاده از فلزات رایج و متداول در فرایند جوشکاری را مورد اندازه گیری قرار داده و مزایای استفاده از این روش در صنایع مختلف مورد بررسی قرار گیرد. به طور کلی، طی این مطالعه، عوامل مختلفی که باعث تسریع و توسعه فرایند جوزایی در صنایع مختلف می گردد، مورد بررسی قرار گرفت و در نهایت در صورت نیاز، روش جوشکاری اکسید مس با روش جوشکاری القایی همراه شده است. با توجه به اینکه اجرای این روش به روش الکترودینامیک محدود بوده و هزینه بسیار بالایی را نیز در بر دارد، تحقیقات لازم در رابطه با پارامترهایی نظیر طول و عمق، مقاومت الکتریکی و طول در جوشکاری انجام شد. روش های مختلف مختلف جوشکاری از قبیل پیچها، جوشکاری لایه ای و برشی در این مطالعه مورد بررسی قرار گرفت که نتایج آن در این مقاله ارایه گردیده است.	دراین کارتجربی، لایه های نازک است تا نگردد پاره ، این جان پرست

The 10,000_Hazm model shows better results compared to the Bolbolzaban model, as depicted in Table 4. The 10,000_Hazm model generates more meaningful samples and longer sentences.

Table 5 the output of the 50,000_Hazm model with the initial inputs "مطالعه برخی خواص", indicating the generation of suitable text.

Table 5: 50000 parameters with Hazm Model

Start Word	Output
مطالعه برخی خواص	مطالعه برخی خواص روغن خردل بر روی پوست صورت، مطالعات زیادی در حوزه سلامت پوستی انجام شده است. هدف از این مطالعه بررسی تاثیر ترکیبات موجود در ترکیب روغن خردل بر روی پوست صورت با استفاده از سه مدل تجربی و انجام چند آزمایش به روش کارآزمایی بالینی با ترکیبی استریل ترکیبی و بدون روغن خردل می باشد. پژوهش حاضر به منظور بررسی ارتباط اسیدهای چرب و عصاره گیاه خردل و میزان بروز آکنه در دو گروه آزمایشی و کنترل دانشجو دانشگاه آزاد واحد کرمانشاه صورت پذیرفت. مواد و روش ها: در این مطالعه با استفاده از آنالیز واریانس، از گروه تجربی استفاده شد. از روش کارآزمایی بالینی، به منظور تعیین میزان بروز آکنه صورت در دو گروه آزمایشی و کنترل، سه گروه شاهد به طور تصادفی، نمونه های پوست افراد گروه تجربی در سالن ورزشی با اسید nمیلی گرم انجام گرفت. آزمایش ها: نتایج حاصل از آنالیز واریانس به عنوان نمونه ها برای آزمون هادر قالب طرح کاملا تصادفی در دو گروه نمونه و شاهد با استفاده از آزمون های استاندارد انجام گرفته است.
پژوهشگران	پژوهشگران توانستند برای نخستین بار اثر ضدقارچی های گیاهی را بر روی سطح برگ از نظر خواص مکانیکی بررسی کنند. از آن به بعد، در هر آزمایش علاوه بر ویژگی های بیوشیمیایی آن، میزان عصاره به کار رفته برای مبارزه با میکروب های عامل بیماری زا بررسی شد. نتایج این تحقیق نشان داد عصاره های گیاهی تولید شده بر سطح برگ درختان لیموترش در مقایسه با عصاره بر روی برگ گیاهان دیگر با درصد تتراکریت کمتر، قابل استفاده بوده است. در این تحقیق اثر ضدقارچ های گیاهی برای گیاه لیموترش موردبررسی قرار گرفت. میزان عصاره مورد بررسی از مخلوط عصاره برگ و پودر عصاره مرکبات و آب لیموترش به دست آمد. پس از گذشت سه ماه، نتایج این آزمایش نشان داد عصاره گیری با عصاره های گیاهی امکان پذیر نبوده است. هم چنین غلظت عصاره های تولیدی از نظر مقاومت به قارچ زایی، نسبت به عصاره گیری به روش های شیمیایی غیر مجاز بیشتر بوده است. نتایج این پژوهش با مقایسه عصاره گیری شیمیایی و خالص، حاکی از آن است که عصاره گیری شیمیایی با عصاره گیری از اسانس لیموترش می تواند باعث کاهش اندوکرین و در نتیجه کاهش مقاومت به این باکتری شود.

The pre-trained model works reasonably well, generating grammatically correct texts while maintaining context. The 50000 of parameter model is expected to have more coherent text than the 20000 of parameter model, but to run a larger model, more resources are needed, and they work longer. For the model for correct

work it's required to train it on a sufficiently large amount of text.

Thus, the primary task before training the model was the search for Persian-language resources with a large database of articles on different topics. Due to the number of articles and the approximate amount of text in them,

Thus, the samples generated using the libraries introduced in this study demonstrate that GPT-2, in addition to text generation and prediction, can also complete text continuations and generate content-rich text. It can be a valuable tool for researchers in providing article abstracts.

Experiments to Explore Model Training

After a series of experiments, studies were carried out on various devices on the basis of which the training took place such as GPU, CPU, TPU. It was decided to configure the server with a connection to Google Colud GPU. The chosen subject area is assistance to researcher and copywriters.

Train The Model to Generate Whole Texts

When the articles become longer, the coherence of the text within them increases. An example can be seen in the figure below with 50,000 parameters.

Table 6: 50000 parameters with Hazm Model

Start Word	Output
پژوهشگران	پژوهشگران ژاپنی تصمیم گرفتند با استفاده از یک روش جدید ریاضی ، هندسه دانان را در پی پاسخ این سوال برانگیزند که میزان دقت دستگاه اعداد صفر و یک در مقایسه با دقت دستگاه اعداد طبیعی چگونه اندازه گیری می شود ؟ آیا آن ها با اندازه گیری این ارقام می توانند عملکرد ریاضی دستگاه اعداد طبیعی را بهبود بخشند؟ شیوه جدید ریاضی این دو ریاضیدان ژاپنی نتیجه آزمایش های اخیر آن ها درباره اندازه گیری دقت دستگاه اعداد طبیعی به وسیله دستگاه اعداد طبیعی است روشی که مدت هاست پژوهشگران علوم محاسباتی ، ریاضی دان ها و فیلسوفان و فلاسفه سراسر جهان در پی آن ها به جستجو و بررسی و کاوش در تئوری های حل مسایل ریاضی درباره اندازه گیری دقت دستگاه های اعداد طبیعی پرداخته اند و تلاش های آنان تاکنون بی نتیجه مانده است بررسی کارایی دستگاه های اعداد طبیعی در ارزیابی و سنجش دقت دستگاه های اعداد طبیعی نتایجی را در پی داشت که در پی آن نقش روش اعداد طبیعی در حل مسایل ریاضی آشکار شده و در واقع روش یادگیری برای کودکان و نوجوانان امروزی تضمین گردیبی است که به آن ها کمک می کند تا درک خود از مسایل ریاضی را نسبت به مسایل تجربی افزایش دهند روش یادگیری دانش آموزان امروزی به دو عامل اصلی « سرعت» و « دقت» وابسته است.
پژوهشگران	

Conclusion

In conclusion, this study focused on generating Persian text using GPT-2, a language model based on the Transformer architecture. Various research studies on text generation in different languages were explored to assess the success rates of machine learning algorithms. Larger datasets yielded more accurate, meaningful, and grammatically correct samples compared to smaller datasets. The system reads and learns from input sentences and data, generating text that is not merely a copy of the content. Libraries like Hazm and regex were used for tasks such as HTML tag removal, data normalization, and string pattern matching. Other libraries like Transformer and tokenizer were also utilized. Different algorithms and parameters were experimented with to improve results. The created models outperformed the bolbolzaban model in text generation. The 50000 model served as the parent model. The results showed that the generated text had the highest correlation with the given words and phrases, resulting in coherent and content-rich texts with correct spelling and adherence to grammatical rules. GPT-2 and GPT-3 technologies have the potential to significantly impact future jobs, streamlining workflows through artificial intelligence. Although these models have their strengths and weaknesses, such as input size limitations and text generation constraints, advancements in language models are expected to address these challenges.

Author Contributions

Z. Askarinejadamiri: Supervision, Project administration, Conceptualization, Methodology, Visualization, Investigation, Writing Reviewing and Editing, Programmer, Writing - Original draft preparation.

M. Solouki: Programmer, Validation, Conceptualization, validation, Investigation, collected the dataset, Writing-Reviewing and Editing, Writing-Original draft preparation.

N. Zanjani: Writing Reviewing and Editing. All authors discussed the results.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

GPT	Generative pre-trained transformer
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Networks

References

[1] D. McDonald, P. Proctor, W. Gill, S. Heaven, J. Marr, J. Young, "Increasing early childhood educators' use of communication-facilitating and language-modelling strategies: Brief speech and language therapy training," Child Lang. Teach. Ther., 31(3): 305-322, 2015.

[2] E. S. Jo, T. Gebru, "Lessons from archives: Strategies for collecting sociocultural data in machine learning," in Proc. the 2020 Conference on Fairness, Accountability and Transparency: 306-316 2020.

[3] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," Internet Things Cyber-Phys. Syst., 3: 121-154, 2023.

[4] A. Tamkin, M. Brundage, J. Clark, D. Ganguli, "Understanding the capabilities, limitations, and societal impact of large language models," arXiv preprint arXiv:2102.02503, 2021.

[5] Y. Yibin, C. Na, X. Chaoqian, Y. Junjian, "Accuracy assessment and analysis for GPT2," Acta Geod. Cartographica Sin., 44(7):726, 2015.

[6] A. S. George, A. H. George, "A review of ChatGPT AI's impact on several business sectors," Partners Univers. Int. Innovation J., 1(1): 9-23, 2023.

[7] X. Zheng, C. Zhang, P. C. Woodland, "Adapting GPT, GPT-2 and BERT language models for speech recognition," in Proc. 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU): 162-168, 2021.

[8] Z. Liu, R. A. Roberts, M. Lal-Nag, X. Chen, R. Huang, W. Tong, "AI-based language models powering drug discovery and development," Drug Discovery Today, 26(11): 2593-2607, 2021.

[9] N. De Cao, T. Kipf, "MolGAN: An implicit generative model for small molecular graphs," arXiv preprint arXiv:1805.11973, 2018.

[10] M. Mars, "From word embeddings to pre-trained language models: A state-of-the-art walkthrough," Appl. Sci., 12(17): 8805, 2022.

[11] Y. Qu, P. Liu, W. Song, L. Liu, M. Cheng, "A text generation and prediction system: pre-training on new corpora using BERT and GPT-2," in Proc. 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC): 323-326, 2020.

[12] O. Shatalov, N. Ryabova, "Towards Russian text generation problem using OpenAI's GPT-2," in Proc. CEUR Workshop 2021.

[13] N. Alexandr, O. Irina, K. Tatyana, K. Inessa, P. Arina, "Fine-tuning gpt-3 for russian text summarization," in Data Science and Intelligent Systems: Proceedings of 5th Computational Methods in Systems and Software, 2: 748-757, 2021.

[14] T. Goyal, J. J. Li, G. Durrett, "News summarization and evaluation in the era of gpt-3," arXiv preprint arXiv:2209.12356, 2022.

[15] S. Kumari, T. Pushphavati, "Question answering and text generation using BERT and GPT-2 model," in Proc. ICCMDE 2021 Computational Methods and Data Engineering: 93-110, 2022.

[16] A. H. George, A. S. Hameed, A. S. George, T. Baskar, "Study on quantitative understanding and knowledge of farmers in trichy district," Partners Univer. Int. Res. J., 1(2): 5-8, 2022.

[17] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[18] A. Kraft, "Triggering models: Measuring and mitigating bias in german language generation," Universität Hamburg, 2021.

[19] M. Moila, T. Modipa, J. Manamela, "The analysis of a GPT-based Sepedi text generation model," in Proc. International Conference on Intelligent and Innovative Computing Applications: 144-152, 2022.

[20] S. P. Ramalepe, T. I. Modipa, M. H. Davel, "The development of a sepedi text generation model using transformers," in Proc. Southern Africa Telecommunication Networks and Applications Conference (SATNAC)," 2022.

[21] L. Butgereit, A. van Staden, "Supporting home-language education in africa with multi-lingual mathematics tutoring using GPT-4," in Proc. International Conference on Artificial Intelligence and its Applications: 44-49, 2023.

[22] S. Yang, D. Feng, L. Qiao, Z. Kan, D. Li, "Exploring pre-trained language models for event extraction and generation," in Proc. the 57th Annual Meeting of the Association for Computational Linguistics: 5284-5294, 2019.

[23] J. Mellon, J. Bailey, R. Scott, J. Breckwoldt, M. Miori, " Do AIs know what the most important issue is? Using language models to code open-text social survey responses at scale," SSRN Electron. J., 2022.

[24] A. Radford et al., "Better language models and their implications. OpenAI," ed, 2019.

[25] R. Dale, "GPT-3: What's it good for?," Nat. Lang. Eng., 27(1): 113-118, 2021.

[26] M. Zhang, J. Li, "A commentary of GPT-3 in MIT Technology Review 2021," Fundam. Res., 1(6): 831-833, 2021.

[27] B. Zhu, Z. Gu, Y. Qian, F. Lau, Z. Tian, "Leveraging transferability and improved beam search in textual adversarial attacks," Neurocomputing, 500: 135-142, 2022.

[28] D. Kim, J. Lee, "Designing an algorithm-driven text generation system for personalized and interactive news reading," Int. J. Human–Computer Interact., 35(2): 109-122, 2019.

[29] X. He et al., "Cater: Intellectual property protection on text generation apis via conditional watermarks," Adv. Neural Inf. Process. Syst., 35: 5431-5445, 2022.

[30] J. Li, Z. Li, L. Mou, X. Jiang, M. Lyu, I. King, "Unsupervised text generation by learning from search," Adv. Neural Inf. Process. Sys., 33: 10820-10831, 2020.

Biographies

Mohadese Solouki was born 1999 in Tehran, Iran. She received her B.S. degree in Software Engineering in2020. She is currently a M.Sc. student in artificial intelligence at Islamic Azad University (Science and Research Branch) and a research fellow at the Image Processing and Natural Language Processing. Her research interests are machine learning, Deep Learning.

- Email: m.soluki1999@gmail.com
- ORCID: 0009-0008-6502-945X
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Zahra Askarinejad is a faculty member of the Computer Department at the Refah College University. She holds a Master's degree in Software Engineering from APU University and a PhD in computer science in field of Software Engineering from the University Putra Malaysia. Her areas of expertise include software engineering, requirement engineering, and HCI.

- Email: askarinejad@refah.ac.ir
- ORCID: 0000-0002-7204-1384
- Web of Science Researcher ID: JDM-5453-2023
- Scopus Author ID: NA
- Homepage: https://refah.ac.ir/cv

Nastaran Zanjani is a faculty member of the Computer Department at the Refah College University. She holds a Bachelor's degree in Electrical Engineering with a specialization in Electronics from the Shahid Beheshti University. She also has a Master's degree in Telecommunications Engineering from Khaje Nasir Toosi University and a Ph.D. in Information Technology from the Queensland University of Technology in Australia. Her areas of expertise include information technology, and HCI.

- Email: m.soluki1999@gmail.com
- ORCID: 0000-0002-5307-683X
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: https://refah.ac.ir/cv

How to cite this paper:
M. Soluki, Z. Askarinejadamiri, N. Zanjani, "Using GPT-2 model and Hazm library for persian text generation," J. Electr. Comput. Eng. Innovations, 12(2): 285-303, 2024.

DOI: 10.22061/jecei.2023.10101.690

URL: https://jecei.sru.ac.ir/article_2023.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

A Node-Centric Approach for Community Detection in Dynamic Networks

M. Sabzekar [1,*], *S. Baradaran Nezhad* [2], *M. Khazaeipoor* [2]

[1]*Department of Computer Engineering, Birjand University of Technology, Birjand, Iran.*

[2]*Department of Computer Engineering, Birjand Branch, Islamic Azad University, Birjand, Iran*.

Article Info

Article History:
Received 25 September 2023
Reviewed 13 October 2023
Revised 03 January 2023
Accepted 15 January 2023

Keywords:
Social networks
Dynamic networks
Community detection
Node influence
Overlapping communities.

*Corresponding Author's Email Address:
sabzekar@birjandut.ac.ir

Abstract

Background and Objectives: Nowadays, social networks are recognized as significant sources of information exchange. Consequently, many organizations have chosen social networks as essential tools for marketing and brand management. Communities are essential structures that can enhance the performance of social networks by grouping nodes and analyzing the information derived from them. This subject becomes more important with the increase in information volume and the complexity of relationships in networks. The goal of community identification is to find subgraphs that are densely connected internally but loosely connected externally.

Methods: While community detection has mostly been studied in static networks in the past, this paper focuses on dynamic networks and the influence of central nodes in forming communities. In the proposed algorithm, the network is captured through multiple snapshots. The initial snapshot calculates the influence of each node. Then, by selecting *k* nodes with higher influence, network communities are formed, and other nodes belong to the community with the most common edges. In the second step, after receiving the next snapshot, communities are updated. Then, *k* nodes with higher influence are selected, and their associated community is created if needed. If the previous community centers are not among the newly selected *k* nodes, the community is dissolved, and the nodes within it belong to other communities.

Results: Based on the results obtained, the proposed algorithm has managed to achieve better results in most cases compared to the compared algorithms, especially in terms of modularity metrics. The reason behind this success could be attributed to the utilization of influential nodes in community formation.

Conclusion: Drawing from the outcomes attained, the suggested algorithm has effectively outperformed the contrasted algorithms in a majority of instances, particularly concerning metrics related to modularity. This accomplishment can potentially be ascribed to the incorporation of influential nodes during the process of community formation.

Introduction

Online social networks have become the most popular interactive medium on the internet due to the possibility of connecting thousands of individuals via the internet [1]. In these networks, users can establish various forms of social relationships such as liking, following, trusting, and more, with each other [2]. These relationships can pave the way for new forms of communication and

sharing emotions and experiences [3], [4]. Nowadays, many organizations view social networks as primary tools for marketing and product management [5]. Each network can be considered as a graph. In this graph, nodes represent individuals, and edges between them represent friendship, interaction, or connection [6]. One of the most important characteristics of networks is the structure of communities within them. Identifying community structure is a significant and challenging topic in social networks, with the aim of finding subgraphs that are internally dense but externally sparsely connected [5].

One of the crucial challenges in community detection is the issue of overlapping communities [7]. The shared membership of some group members is referred to as community overlap in networks [8]. Considering such a concept, each node can belong to multiple groups based on its attributes. When dealing with large-scale data networks, identifying communities with overlaps presents a challenging and computationally complex problem. As a result, many research efforts attempt to minimize the consideration of overlaps [6].

Various studies have been addressed thr community detection problem. However, online social networks possess a dynamic nature, constantly changing and evolving. A dynamic network (DN) can be understood as a network that changes and evolves over time. These changes can be summarized in four types of operations: node creation, node deletion, edge creation, and edge deletion. In the past, community detection was primarily studied for static networks, with the dynamic nature of networks often overlooked. Nowadays, due to the substantial growth in network size and the evolving nature of network structures, researchers have shifted their focus to DNs [9]. Community detection in DNs is essential for crucial applications such as social network analysis [10].

In many social networks, users can express their opinions and feelings about specific products, services, or topics based on their experiences and share them with others. These opinions and feelings are expressed by real users and customers and are observed by their friends or followers. If an opinion is presented by a familiar individual such as a friend or a celebrity, it holds a more significant influence on the user's decision regarding that topic. This phenomenon is considered a new concept in social networks. In social networks, nodes that possess a higher capacity for disseminating information hold more significance and are known as influential nodes or leaders. Identifying influential nodes or leaders in a network can be perceived as ranking nodes in terms of importance within the network [11]. Despite studies indicating the correlation between the behavior of a node and the behavior of its neighboring nodes in the network [12], rarely have the behavioral aspects been incorporated into the problem of community detection.

Alongside the problem of community detection in dynamic social networks, this topic can introduce various challenges such as the speed and formation of communities, dynamics within communities, and the ability to update them, which have not yet received suitable solutions.

In this paper, an attempt has been made to provide an algorithm for community detection in social networks that aims to efficiently identify existing communities in the network based on the influence of nodes on each other and taking into account the possibility of overlapping communities. Furthermore, we strive to address the issue of identifying influential nodes in dynamic networks and enhance current criteria for effectiveness in dynamic networks. Moreover, recognizing that real-world network communities often exhibit overlaps, our aim was to devise a method addressing such overlaps. Our method aims to preserve communities with minimal overlap, allowing their members to participate in multiple overlapping communities. To our knowledge, no prior method exists that combines influential node utilization for community determination in DNs while also accounting for overlapping communities. Furthermore, our proposed method accommodates various changes in dynamic networks, including node and edge additions or removals. The algorithm achieves these objectives through six distinct phases: network snapshot acquisition, influential node identification via local and global information, initialization, community expansion, evaluation and merging of communities, and finally, community updates.

Basic Concepts

A. Dynamic Networks

Social networks are one of the most common types of networks, characterized by a connected structure of entities formed for social interactions [13]. To encode networks and represent their adjacencies, adjacency matrices are also utilized. To express a network in the form of an adjacency matrix, an *n*×*n* matrix is considered, where *n* corresponds to the number of graph vertices [14]:

$$A_{i,j} = \begin{cases} 1 & if\ there\ is\ an\ edge\ between\ i\ and\ j \\ 0 & otherwise \end{cases} \quad (1)$$

It is evident that interactions among members in a network, particularly in social networks, change over time. Networks evolve through the joining or departure of members from a network and the establishment or termination of relationships. This evolution not only alters the fundamental structures of networks but also adjusts the community structures in various snapshots of the network over time [15]. Diverse approaches have treated

social networks as sets of static graphs, each representing entities and their relationships in a momentary snapshot of the network [15], [16]. In Fig. 1, an example of a snapshots of a network is presented.

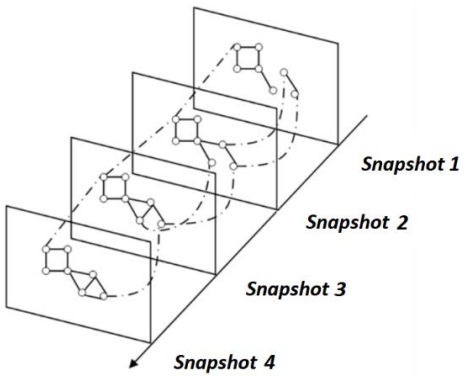


Fig. 1: An example of snapshots in a dynamic network.

B. Community Detection

In general, community detection is an unsupervised learning technique for clustering nodes, considering the network's structure, and is also a key feature that can be used to extract useful information from networks [17]. Internal communications within each community are dense, while external communications between communities are sparse [5]. Each community is a group of network nodes in such a way that the connections between nodes within the group are stronger than their connections with other network nodes [14].

With the introduction and prevalence of dynamic networks, the concept of dynamic communities has also emerged [18]. The objective of dynamic community detection is to identify a set of all the existing communities within a dynamic network in a way that the described partitions by it can have overlaps [19].

The problem of community detection in large-scale networks is computationally infeasible and NP-hard. Numerous techniques have been proposed to find optimal communities relatively quickly. Most of these techniques are based on optimizing objective functions, with modularity optimization being one of the most widely used techniques among them. Nevertheless, this problem is still NP-hard [17]. Researchers have developed various techniques that usually start from pre-defined small communities and implement an algorithm that first expands these communities, then identifies newly formed communities (in all cases), and finally uncovers overlaps between communities [20]. To measure and evaluate community quality, different metrics have been proposed. One of the most important metrics is modularity, which is widely used to assess the quality of the network community structure. Communities with high modularity have denser connections among nodes within similar communities but sparser connections with nodes in other communities. Another metric in this field is the normalized mutual information [21].

Numerous algorithms for community detection using different techniques and tools have been developed. However, due to the wide spectrum of networks, a single community detection algorithm cannot perform well and effectively in all types of networks and have good and suitable performance [22].

C. Influential Nodes in Networks

Identifying influential users in social networks has extensive applications in marketing, politics, disease control, and more [13]. Nodes in such networks have the ability to influence their neighboring nodes, and an influenced node can acquire a behavior or attribute from its neighboring nodes. Finding nodes with the highest influence has drawn the attention of social network managers and analysts. Marketing managers might be interested in identifying influential individuals and offering them discounts or free products, hoping that these individuals will encourage their friends to purchase these products [23].

The importance of a node in a network can be assessed using metrics available in graph theory. These metrics rely on the topology of the network. The impact of social networks relates to a user's ability to change the emotions, attitudes, or behaviors of other users in a network. The strength of the link between two nodes in a network depends on the overlap of their neighbors. Influential individuals are significantly associated with more groups compared to ordinary individuals. However, in online social networks, this criterion may not always be applicable for identifying influential users. Various methods for identifying influential nodes have been proposed, with the most common being [24]: degree centrality, centrality, closeness centrality, eigenvector centrality, and leader benefit. Each of these methods identifies influential nodes by examining the nodes and their connections in a certain way.

Literature Review

Numerous studies have been presented that address the problem of community detection using influential nodes. Most of them are proposed to detect communities in the static networks [5], [25], and [26]. Static research methods encompass topological analysis of complex networks, identification of key nodes or community leaders, knowledge-community discovery, and community-structure discovery.

However, real-world networks, particularly prevalent online social networks like Facebook, LinkedIn, and Twitter, are inherently dynamic and continually expanding in size and complexity. Therefore, developing effective and efficient algorithms to detect communities in dynamic networks is a formidable challenge.

An efficient dynamic community detection algorithm should adaptively and incrementally update communities based on changes in the network structure. Redundant computations need to be avoided for computational efficiency. Designing such an algorithm that maintains effectiveness similar to static algorithms solely through historical community structures and incremental changes is challenging. Additionally, there is ongoing uncertainty regarding categorizing incremental changes in dynamic networks and evaluating their influence on community structure updates, which is crucial for an effective and efficient dynamic algorithm.

Recent research has proposed several methods for detecting communities in dynamic networks, broadly categorized into three classes [21]: instant-optimal, temporal trade-off, and incremental approaches. Firstly, instant-optimal methods involve applying static algorithms independently to each network snapshot to detect communities, subsequently matching these communities with those detected in previous snapshots. As an example, the authors in [27] provide a localized modularity optimization approach where only the communities that underwent changes are examined, leaving the rest of the communities untouched. According to the claims of the paper, this algorithm exhibits greater effectiveness compared to similar algorithms.

Secondly, temporal trade-off approaches assume that communities at a certain time are influenced not only by the current network topology but also by past topology or identified partitions. These approaches strike a balance between an optimal solution at the current time and information from the past without considering future changes. As an example of the algorithms in this category, in [15], a method for detecting overlapping community structures is presented. This method considers the task of community detection as a non-negative matrix factorization problem. The proposed approach utilizes a probabilistic model to account for the dynamic nature of community structures and employs a block coordinate descent technique to solve the objective function of the matrix factorization model. This solution introduces a non-negative hidden factor to estimate gradients for faster computation. The results obtained indicate that the proposed method outperforms previous algorithms in terms of well-known evaluation metrics for evolving networks. In another work [28], the adopted approach involves a multi-objective optimization strategy. Initially, the method incorporates the probability fusion technique and employs two distinct approaches, namely neighbor diversity and neighbor crowd. These approaches facilitate the rapid and precise formation of appropriate communities. Also, the utilization of a progression metric enables the authors to identify similarities between the communities formed in two consecutive snapshots.

Finally, cross-time algorithms aim to discover communities that are relevant across the entire network evolution, where communities identified at a given time depend on both past and future network topologies. For example, in [21], a dynamic community detection algorithm based on modularity is introduced. This method aims to identify communities in dynamic networks through the repeated use of static algorithms, but in a more efficient manner. This approach is an adaptive and incremental algorithm designed to maximize incremental modularity during the update of dynamic network community structures. In this paper, the dynamic network is modeled as a sequence of gradual changes, and for each gradual change, an operation was designed to maximize modularity. An influence-based community detection in dynamic networks was proposed in [29] that formulate the problem as a combinatorial optimization problem that aims at partitioning a given social network into disjoint *m* communities. The objective is to maximize the sum of influence propagation of a social network through maximizing it within each community. In another study [30] a community detection method for dynamic networks was represented that is based on tracking of backbones and bridges. They applied the “backbones” to reflect the critical edges of communities and the “bridge” edges to describe the key connections between communities. Table 1 summarizes some of the proposed methods for addressing the problem.

In this section, an attempt has been made to introduce the latest articles in this field. Based on conducted studies, there are numerous challenges that need to be addressed as open research issues. Community detection itself is a computationally expensive and complex problem. The existence of various snapshots of the network necessitates re-computation to update or create communities. Therefore, methods are required to be as cost-effective and computationally efficient as possible. Additionally, for community detection in dynamic networks, it is necessary to compare two momentary images of the past and present. This comparison is often time-consuming and requires moderate to high computations. Having rules or methods to expedite this process can be the key to success in speeding up community detection algorithms in dynamic networks.

Furthermore, influential nodes can serve as a foundation for generating many communities. This aspect has been overlooked in many studies. Moreover, the identification and updating of influential nodes in a network is a subject that has received less attention. Ultimately, many research efforts have disregarded the issue of community overlap. In other words, attempts have been made to develop algorithms for non-overlapping communities. Yet, in today's world, networks, especially social networks, exhibit overlapping communities.

Table 1: Overview some of the reviewed studies

Method	Description	Strengths/Weaknesses
LGIEM [5]	A community detection algorithm for static networks utilizing influential nodes to find communities.	Strengths: • Using a suitable criterion to reduce overlap between communities Weaknesses: • time-consuming and computationally expensive
LPA_NI [25]	Detection of communities based on label propagation Conducting label propagation operations based on node importance and influence.	Strengths: • Utilizing node importance for label propagation can identify more appropriate communities. • The influence of nodes on each other is well modeled. • The detected communities have good quality. Weaknesses: • Calculating node importance and influence increases computational load.
NANI [26]	Utilizing group influence for identifying communities Using various metrics to determine the influence of each node on others Utilizing a method similar to hierarchical agglomerative clustering for community detection	Strengths: • High simplicity of the proposed algorithm using a wide range of metrics to determine influence and compare node similarities for community creation. Weaknesses: • No specific corrective mechanism for refining clusters and finding optimal clusters overlap.
DynaMo [21]	Proposing a community detection algorithm for dynamic networks Considering six categories of changes in communities and designing strategies for each	Strengths: • Appropriate speed for introducing changes in communities Weaknesses: • Repetitive computations and operations
PODCD [15]	Using a probabilistic method for identifying overlapping communities	Strengths: • Taking overlap into account in communities. Weaknesses: • High computational load in large networks.
D-Louvain [27]	Based on the modularity optimization algorithm (Louvain algorithm), which is one of the strongest algorithms in this field.	Strengths: • Examining changing communities instead of all communities, leading to reduced computations and increased algorithm speed. Weaknesses: • The algorithm is stochastic, resulting in unstable results. Unable to cope with overlap in communities.
MOCCD [28]	Based on characteristics fusion of dynamic social networks	Strengths: • Utilize multi-objective optimization that fuses the characteristics of dynamic network communities. • Fast convergence and high accuracy.
Sandwich [29]	formulate the problem as a combinatorial optimization problem based on sandwich approximation framework.	Strengths: • influence maximization • develop a lower bound and an upper bound of the objective function.
BBTA [30]	Introducing two concepts, backbones and bridges for edges.	Strengths: • novel incremental algorithm to detect dynamic communities based on the network change rate and changes on the backbones or bridges. Weaknesses: • Stability and robustness

The Proposed Method

The proposed method consists of six main phases, and in the following, we will explain each phase. Fig. 2 illustrates the overall phases of the proposed method. The details of each phase will be discussed at the following of the section.

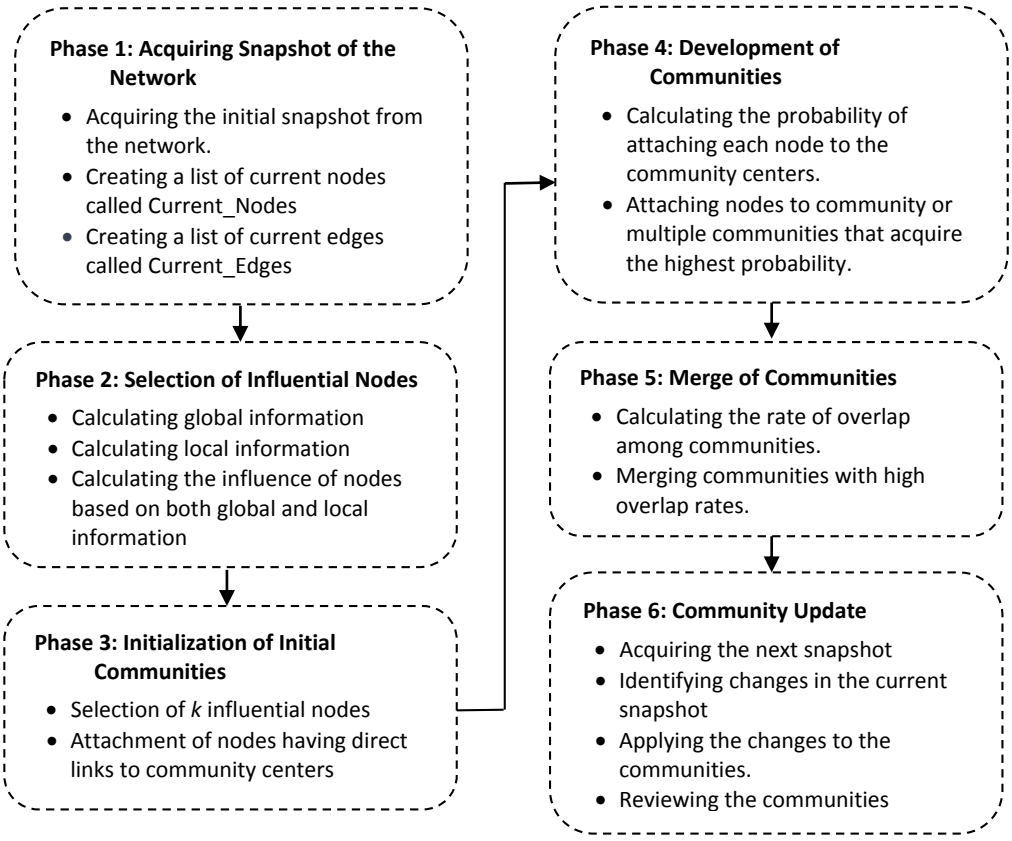


Fig. 2: Phases of the proposed method.

A. Phase 1: Acquiring Snapshot of the Network

In this phase, multiple snapshot images of the network are taken at different time intervals. Each image represents a set of nodes and the links between them, divided based on the time interval. Each image is independently given to the algorithm for community identification. The first image of the network is used for community detection, and the subsequent images are used to update the communities. Additionally, the nodes and edges present in the current network snapshot are extracted and encoded. For this purpose, two sets will be available: one containing the nodes and the other containing the edges. It is assumed that each node has a unique identifier, and the network can identify a node using this identifier. Therefore, based on this identifier, the sets of nodes and edges are generated.

The set of nodes in the current network snapshot is stored in an array called *Current_Nodes*. To store the edges, a matrix of size $m \times 2$ named *Current_Edges* is used, where m represents the number of edges, and each row contains the identifiers of the nodes forming the edge.

B. Phase 2: Selection of Influential Nodes

In this phase, following the approach in [5], three steps are taken to find influential nodes based on their local and global information. In the first step, to identify the global information of a node, the k-shell network decomposition algorithm is employed. Various metrics exist for calculating node importance in the network, but only node degree and clustering coefficient can indicate local network information. The k-shell is a connected subgraph of the maximum possible size in graph G where each vertex has a degree of at least k. The k-shell value for node i denoted as Ks(i) indicates that node i belongs to shell k but not to any other (k+1) shell. The k-shell decomposition method is often used to identify core nodes and peripheral nodes of the network. It starts by removing all nodes with only one link until no nodes remain and assigns them to shell 1. Likewise, it recursively removes all nodes with degree 2 or less and creates shell 2. This process continues until all nodes of the network are assigned to a single shell. Shells with higher indices are located in the core or center of the network. The k-shell decomposition method can be efficiently implemented with a linear time complexity of O(m), where m represents the number of edges in the network. An example of the k-shell algorithm's operation is depicted in Fig. 3.

In the second step, both global and local information of each node is computed. Global information indicates the node's status within the entire network. A node with high centrality has a higher *k*-shell value. The global

information of a node *i*, denoted as GI_i, indicates the dependency strength of other nodes in the network on node *i*. In other words, GI_i is calculated based on the average shells of the neighboring nodes. Thus, for a node like node *i*, it is computed according to (2).

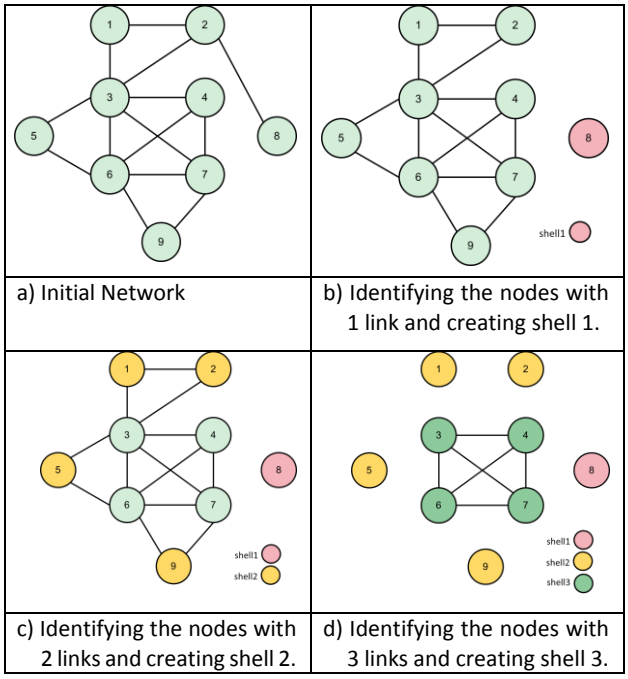


Fig. 3: An example of the k-shell decomposition algorithm's.

In this equation, *NumShell* represents the number of layers created by *k*-shell decomposition. Moreover, *neighbor(i,j)* is the number of neighbors of node *i* that belong to layer *j* of the *k*-shell decomposition.

$$GI_i = \frac{\sum_{j \in \text{NumShell}} |neighbor(i,j)| \times j}{\text{NumShell}} \quad (2)$$

After obtaining the global information of nodes, the measurement of local information follows. For measuring local information, the number of neighbors of each node is utilized. Thus, based on (2), the value of LI_i, representing the local information of node *i*, is derived.

$$LI_i = |neighbor(i)| \quad (3)$$

In the final step, to calculate the node's influence in the network, global and local information are combined according to (4). In this equation, α and β are coefficients for global and local information, respectively.

$$influence(i) = \alpha(GI_i) + \beta(LI_i) \quad (4)$$

Nodes with higher influence are considered as community centers. The pseudocode corresponding to the algorithm of this phase is provided in Algorithm 1.

C. Phase Three: Initialization of Initial Communities

Based on Phase Two, the list of influential nodes is obtained and sorted in descending order according to their influence level. In the next step, *k* nodes with higher influence are chosen as the cluster centers. Among the remaining nodes in the network, those having a direct link to cluster center nodes form the basis of communities. If two cluster centers have a direct link to each other, this link is disregarded, and these two centers will not be part of each other's communities.

Algorithm 1: Calculating nodes' influences

Input: Graph
Output: find influence of each node

```
initialize V = all nodes in G
for i = 1: n
    compute k-shell by k-shell decomposition algorithm
end for
compute k-shell for each node
calculate the number of neighbors of each node
for each node like i:
    calculate GI(i) by formula (2)
    calculate LI (i) by formula (3)
    calculate influence(i) by formula (4)
end for
```

D. Phase Four: Development of Communities

In this phase, nodes that are not yet part of any community are integrated into existing communities. To achieve this, attention is given to the nodes' neighbors. Essentially, a node joins a cluster where the majority of its neighbors are already affiliated. If the number of neighbors is the same for multiple clusters, the node becomes a member of all those clusters. The degree of association with a community is determined using (5):

$$dependency(i) = \max_{j \in ClusterHeads} (|cluster(neighbor\ (i),j)|), \quad (5)$$

where, $|clulouvainster(neighbor\ (i),j)|$ represents the count of neighbors of node *i* belonging to cluster *j*. If a node has neighbors to which no cluster has been assigned yet, those neighbors are disregarded in the aforementioned equation. Algorithm 2 outlines the steps related to community creation and expansion.

In line 1 of Algorithm (2), nodes are sorted based on their influence levels. In line 2, k nodes with higher influence are selected as seed nodes and stored in a list named "HeadCluster," responsible for maintaining the community centers. To better manage node processing for community expansion, a list called "Candidate" is created in line 3. This list is responsible for keeping track of nodes that have at least one neighbor belonging to a community. This ensures that nodes chosen for community extension are those which are guaranteed to have at least one neighboring node with an assigned

cluster. Otherwise, the node will remain without a community affiliation. Consequently, in line 3, all neighbors of the community centers are added to the Candidate list. Additionally, another list named *"assign_Nodes"* is created in line 4, tasked with holding nodes that have joined at least one community.

Algorithm 2: Algorithm for community creation and development

Input: Get Graph snapshot
Output: Output: Communities

```
1. Sort nodes in a descending order
2. HeadCluster ← select k top nodes
3. Candidate ← Neighbors(HeadCluster )
4. assign_Nodes←assign_Nodes ∪ HeadCluster
5. for j in Candidate
6.     flag ← false
7.     for HeadCluster like h
8.         if j is neighbors h
9.             Communities (h) ← j
10.            flag ← true
11.        end if
12.    end for
13.    if (flag == false)
14.        maxCom ← calculate dependency(j)
15.        for each community in maxCom
16.            Communities (community) ← j
17.        end for
18.    end if
19.    for each node in Neighbars(j ) like t
20.        if t ∉ assign_Nodes
21.            Candidate ← Candidate ∪ t
22.        end if
23.    end for
24.    assign_Nodes ← assign_Nodes ∪ j
25. end for
26. while (there is node in Graph that don't exist in
        assign_Nodes like j)
27.    neighbor ← Neighbars(j ) ∪ j
28.    HeadCluster ← maxinflunce (neighbor )
29.    for each node in neighbor like t
30.        Communities (maxinfluence(neighbor)) ← t
31.        assign_Nodes ← assign_Nodes ∪ t
31.    end for
32. end while
```

Continuing, all nodes selected as seed nodes are added to this list to prevent them from reappearing during the clustering process (line 5). In line 6, a variable named "flag" is initialized to identify cases where a node is a neighbor of a community center. Line 7 iterates through each community center to check whether node j is a neighbor or not. If node j is a neighbor of a community center, it is integrated into that community, and the flag is set to true (lines 7-12). If node j is not a neighbor of any community center, its association with communities is calculated using (5), and it is added to the possible communities (lines 13-18). Subsequently, all neighbors of node j that do not belong to any community are added to the Candidate list, and node *j* is also added to the *assign_Nodes* list to exclude it from the community creation process (lines 19-24). This process continues until the Candidate list contains at least one member (lines 26-30). Finally, lines 26-30 address cases of community dispersion. If nodes without a community exist, the most influential node among them and its neighbors are selected as the community center, and other nodes join this cluster as members.

E. Phase Five: Evaluation and Merge of Communities

In the process of selecting influential nodes, k nodes with high influence and their neighbors form initial communities. Therefore, if a node is connected to multiple community centers, it can be assigned to different communities. During the community expansion process, if a node of equal priority belongs to multiple communities, it is assigned to all of them. Consequently, many nodes exhibit overlapping membership. However, excessive overlapping between communities can increase algorithm complexity. Thus, controlling community overlap is essential for optimizing community structure. To control overlap, the overlap rate is employed, calculated based on (6):

$$\delta = \frac{|C_i \cap C_j|}{\min\{|C_i|,|C_j|\}}, \tag{6}$$

where C_i and C_j are two clusters or communities and $|C_i|$ and $|C_j|$ denote the number of their nodes. As δ increases, more nodes share membership in overlapping communities, making two communities more susceptible to merging. Nevertheless, in real-world scenarios, although two communities might have high overlap, practical considerations such as the community's subject matter might prevent their merger. Therefore, this paper also introduces another criterion called "community fitness". Community fitness indicates the significance of a community in the network and is calculated according to (7):

$$\begin{aligned} fitness(C_i) &= \frac{density_i + unique_i}{2}, \\ density_i &= \frac{|C_i|}{totalNodes}, \\ unique_i &= \frac{unique\_member_i}{member_i} \end{aligned} \tag{7}$$

Here, $density_i$ represents the density of the community which is calculates as the result of dividing the number of community members by the total number of nodes in the network. Furthermore, $unique_i$ shows the ratio of all nodes belonging to the same community i to the total nodes present in the community. Consequently,

communities that can achieve a fitness value exceeding a certain threshold persist, while others merge with different communities. In other words, for the community fitness criterion, lower values increase the likelihood of cluster merging.

To merge two communities, the node with the highest influence is selected as the new community center. All members of both clusters, including the old community center, are considered members of the new merged community. Algorithm 3 outlines the procedure for merging communities.

Algorithm 3: Algorithm for Community Merge

Input: communities
Output: final communities

for each community like C_i
  for each community like $C_j \neq C_i$
    calculate δ according to (6) and (7)
    if (δ > threshold$_\delta$ and fitness < threshold$_{fitness}$)
      $C_{new} \leftarrow max(seed_{influence,i}, seed_{influence,j})$
      $C_{new} \leftarrow \{C_i \cup C_j\}$
      $community \leftarrow community - C_i - C_j$
      $community \leftarrow C_{new}$
    end if
  end for
end for

F. Phase Six: Community Update

This phase consists of four steps, including: obtaining a new snapshot and identifying changes, incorporating changes into communities, reviewing communities, determining communities for orphan nodes, and evaluating and merging communities. Algorithm 4 illustrates the steps involved in performing this task. Assume that the designated nodes in the previous network are placed in the list *Current_Node* and the corresponding edges are in the list *Current_Edge*.

In the first step, a new snapshot of the network is acquired, and based on the existing communities, the communities are updated. To update communities, considering t as the current time, changes occurring in the network at time t compared to time *t*-1 (the previous network) must be determined. Network changes may involve adding or removing a node to/from the network or adding/removing an edge to/from the network. Set operators, including union and intersection, are utilized to identify network changes.

In the second step, following the identification of changes, the changes are applied to the network and the existing community structure from the previous network snapshot. The network can provide five types of changes, including adding or removing a node, or adding/removing an edge. In each of these scenarios, the influence levels of nodes and node membership in communities might undergo changes, necessitating decisions regarding them.

After implementing changes in the previous step, a reevaluation of the communities is carried out. During this stage, all communities that have undergone at least one change are reviewed. In this phase, the community centers and the nodes present within them are examined. If necessary, a community might be reassigned to a different node. Furthermore, if a community cannot meet the threshold for influence, it is dissolved. In this case, the dependency value (dependency(i)) is calculated for the members of the dissolved community, and decisions are made concerning orphan nodes.

In the final step, the degree of overlap and the fitness of nodes within communities are evaluated and determined. If necessary, communities are merged. In the next section, a detailed discussion will be presented regarding the effectiveness of the proposed method.

Algorithm 4: Algorithm for Network Change Detection

Input:
New_{Graph} //Get new Graph snapshot
$Current_{\mathrm{Node}}$ // Nodes in pervious network
$Current_{Edge}$ // Edges in pervious network
Output: $Current_{\mathrm{Node}}$, $Current_{Edge}$

Step 1: Initialization
1. $old_{\mathrm{Node}} \leftarrow Current_{\mathrm{Node}}$, $old_{Edge} \leftarrow Current_{Edge}$
2. get New_{Graph} and create New_{Node} , New_{Edge}

Step 2: find node changes
3. $NodeList \leftarrow old_{\mathrm{Node}} \cap New_{Node}$
4. $Omited_{Node} = old_{\mathrm{Node}} - NodeList$
5. $Added_{Node} = New_{\mathrm{Node}} - NodeList$
6. $Current_{\mathrm{Node}} = Current_{\mathrm{Node}} - Omited_{Node}$
7. $Current_{\mathrm{Node}} = Current_{\mathrm{Node}} + Added_{Node}$

Step 3: find edge intersect
8. $EdgeList \leftarrow old_{Edge} \cap New_{Edge}$
9. $Omited_{Edge} = old_{Edge} - EdgeList$
10. $Added_{Edge} = New_{Edge} - EdgeList$
11. $Current_{Edge} = Current_{Edge} - Omited_{Edge}$
12. $Current_{Edge} = Current_{Edge} + Added_{Edge}$

Step 4: Calculate influence
13. **for** each Node that adds or changes
14. Calculate influence according (Algorithm 1).
15. **end for**

Results and Discussion

In this section, we focus on the implementation and evaluation of the proposed algorithm (DIC). For implementation, the Python programming language is utilized. To execute the algorithm, relevant datasets are loaded. The datasets employed in this study are presented in Table 2.

To assess the performance of the proposed algorithm, we compare it with five recent methods introduced in studies DynaMo [21], D-Louvain [27], BBTA [30], DPC-DLP [31], ECD [32], IncNSA [33].

Table 2: Datasets used for the experiment

Dataset	#Nodes	#Edges	#Snapshots
Cit-Hep Ph	34546	421578	11
sx-mathoverflow	24818	506550	8
CollegeMsg	1899	59835	7

The evaluation metrics employed in this article consist of Newman's modularity, modularity with split penalty, and modularity density. The Newman's modularity (Q), applicable to undirected and unweighted networks, is defined as the difference ratio between the actual and expected number of edges within a community, as shown in (8).

$$Q = \sum_{c_i \in C} \left[\frac{\left|E_{c_i}^{in}\right|}{|E|} - \left(\frac{2\left|E_{c_i}^{in}\right| + \left|E_{c_i}^{out}\right|}{2|E|} \right)^2 \right], \tag{8}$$

where C represents the set of all communities, c_i denotes the i-th community, $\left|E_{c_i}^{in}\right|$ is the number of edges within the community, $\left|E_{c_i}^{out}\right|$ is the number of edges outside the community, and |E| is the total number of edges in the network. A higher value of this metric indicates the suitability of the communities. It's worth noting that due to the complexity of social networks' interactions and the possibility of links between any two edges, a value around 0.5 is considered appropriate for modularity. Therefore, if the value of this metric is around 0.5, it signifies the appropriateness of the community structure in a social network. Table 3 to 5 display the results of this metric for the compared methods across various datasets.

Table 3: Comparison of results using the Newman's modularity metric for the cit-Hep Ph dataset

TS*	DIC	D-Louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.655**	0.628	0.534	0.538	0.566	0.575	0.631
2	**0.088**	0.083	0.021	0.087	0.0722	0.059	0.078
3	0.040	0.036	0.034	0.034	0.0508	**0.061**	0.049
4	**0.022**	0.017	0.013	0.015	0.012	0.016	**0.025**
5	0.012	0.005	0.010	0.010	0.0018	**0.015**	0.013
6	0.011	0.005	0.010	0.011	0.0006	0.010	**0.012**
7	**0.199**	0.111	0.174	0.123	0.195	0.118	0.163
8	**0.218**	0.201	0.215	0.211	0.213	0.209	0.200
9	**0.374**	0.223	0.116	0.238	0.263	0.292	0.374
10	**0.405**	0.103	0.157	0.365	0.280	0.351	0.295
11	**0.412**	0.225	0.261	0.303	0.255	0.252	0.359

* TS: Time stamp

As observed, according to the results obtained from Table 3 to 5, the proposed method has shown better performance compared to the other compared methods in most time intervals for each dataset.

Table 4: Comparison of results using the Newman's modularity metric for the CollegeMsg dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.380**	0.329	0.331	0.278	0.322	0.354	0.348
2	0.146	0.149	0.122	0.129	0.122	0.108	**0.151**
3	0.494	0.477	0.431	0.378	0.443	0.359	**0.510**
4	**0.253**	0.244	0.232	0.245	0.202	0.210	0.211
5	**0.137**	0.117	0.110	0.104	0.111	0.108	0.110
6	0.167	0.150	0.181	0.171	**0.195**	0.154	0.181
7	**0.288**	0.206	0.209	0.238	0.183	0.239	0.213

* TS: Time stamp

Table 5: Comparison of results using the Newman's modularity metric for the sx-mathoverflow dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.042**	0.214	0.190	0.209	0.216	0.151	0.041
2	0.000	0.068	0.001	0.001	0.007	0.057	**0.074**
3	0.000	0.218	0.188	0.213	0.176	0.160	**0.221**
4	0.001	0.004	0.032	0.084	0.020	**0.114**	0.046
5	**0.100**	0.054	0.069	0.049	0.008	0.062	0.074
6	**0.118**	0.096	0.086	0.106	0.106	0.104	0.110
7	**0.124**	0.071	0.107	0.111	0.114	0.061	0.101
8	**0.110**	0.016	0.074	0.108	0.064	0.102	0.079

* TS: Time stamp

Modularity with split penalty (Q_s) is another evaluation criterion used to assess the quality of community structures. It is calculated based on (9). In this equation, Q represents the modularity measure, and SP is the number of shared edges between communities, calculated according to (10).

$$Q_s = Q - \mathcal{SP} \tag{9}$$

$$\mathcal{SP} = \sum_{c_i \in C} \left[\sum_{\substack{c_j \in C \\ c_j \neq c_i}} \frac{\left|E_{c_i,c_j}\right|}{2|E|} \right] \tag{10}$$

In (10), $\left|E_{c_i,c_j}\right|$ represents the number of edges that connect community c_i to c_j. In this criterion as well, a higher value indicates a more suitable community structure. The results obtained for this criterion, separated by dataset, are presented in Table 6 to 8.

According to the results obtained from Table 6 to 8, for most datasets and time intervals, the proposed method based on the modularity with penalty division criterion has performed better.

Both Newman's modularity (Q) and modularity with split penalty (Qs) are independent of the number of nodes within communities. Modularity density (Q_{ds}) examines the nodes' density and their compactness within communities. For undirected networks, modularity density (Q_{ds}) is defined by (11).

Table 6: Comparison of results using Qs criterion for the cit-Hep Ph dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.196**	0.120	0.136	0.195	0.145	0.130	0.156
2	0.433	0.437	0.328	0.412	0.403	0.456	**0.555**
3	0.4663	**0.4695**	0.446	0.426	0.424	0.427	0.410
4	**0.488**	0.485	0.433	0.384	0.384	0.383	0.419
5	0.489	**0.495**	0.406	0.388	0.417	0.474	0.436
6	**0.490**	0.485	0.452	0.450	0.459	0.462	0.450
7	0.143	0.104	**0.176**	0.101	0.173	0.134	0.160
8	0.492	0.498	0.389	0.422	0.391	0.395	**0.515**
9	0.491	0.486	0.410	0.446	0.446	0.465	**0.532**
10	**0.491**	0.486	0.457	0.446	0.397	0.398	0.364
11	**0.153**	0.015	0.087	0.131	0.161	0.096	0.140

* TS: Time stamp

Table 7: Comparison of results using Qs criterion for the CollegeMsg dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.499**	0.491	0.143	0.399	0.296	0.399	0.368
2	**0.500**	0.500	0.493	0.490	0.410	0.415	0.432
3	**0.499**	0.399	0.408	0.430	0.433	0.484	0.360
4	**0.498**	0.499	0.397	0.324	0.313	0.424	0.460
5	0.040	**0.048**	0.162	0.034	0.027	0.011	0.042
6	**0.074**	0.054	0.040	0.0339	0.041	0.038	0.055
7	**0.068**	0.058	0.141	0.136	0.141	0.132	0.059

* TS: Time stamp

Table 8: Comparison of results using Qs criterion for the sx-mathoverflow dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	0.5660	0.1829	0.2727	0.2722	0.3167	0.3268	**0.6245**
2	0.0002	0.0002	0.1279	0.0417	0.1514	0.1212	**0.1540**
3	0.0005	**0.0008**	0.0001	0.0004	0.0005	0.0005	0.0006
4	0.0018	**0.0031**	0.0013	0.0013	0.0014	0.0012	0.0025
5	**0.1750**	0.1321	0.1067	0.1590	0.1644	0.1600	0.1236
6	**0.2080**	0.0101	0.0155	0.1747	0.1495	0.1090	0.1608
7	**0.1577**	0.0146	0.1404	0.1393	0.1341	0.0188	0.1554
8	**0.0109**	0.0104	0.0107	0.0023	0.0102	0.0104	0.0105

* TS: Time stamp

$$Q_{ds} = \sum_{c_i \in C} \left[\frac{|E_{c_i}^{in}|}{2|E|} d_{c_j} - \left(\frac{2|E_{c_i}^{in}| + |E_{c_i}^{out}|}{2|E|} \right)^2 - \sum_{\substack{c_j \in C \\ c_j \neq c_i}} \frac{|E_{c_i,c_j}|}{2|E|} d_{c_i,c_j} \right], \quad (11)$$

where d_{ci} represents the internal density of cluster c_i, and $d_{ci,cj}$ represents the between-community density between communities c_i and c_j, calculated using (12).

$$d_{c_i} = \frac{2|E_{c_i}^{in}|}{|c_i|(|c_i|-1)}, \quad d_{c_i,c_j} = \frac{|E_{c_i,c_j}|}{|c_i||c_j|} \quad (12)$$

Similar to the previous criteria, in this criterion as well, a higher value indicates a more suitable community structure. The results obtained for this criterion, separated by dataset, are presented in Table 9 to 11.

Table 9: Comparison of results using Qds criterion for the cit-Hep Ph dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.513**	0.474	0.399	0.479	0.433	0.472	0.480
2	0.070	0.064	0.1229	0.140	0.116	**0.197**	0.165
3	0.095	0.030	0.122	0.060	0.116	**0.128**	**0.120**
4	**0.121**	0.015	0.066	0.108	0.119	0.118	0.110
5	0.111	0.005	0.138	0.059	**0.166**	0.063	0.160
6	0.128	0.001	0.123	0.126	0.072	0.058	**0.148**
7	**0.257**	0.181	0.229	0.150	0.231	0.239	0.249
8	**0.167**	0.100	0.169	0.141	0.153	0.146	0.127
9	**0.208**	0.101	0.143	0.043	0.126	0.070	0.183
10	**0.210**	0.103	0.127	0.100	0.045	0.131	0.115
11	**0.223**	0.133	0.138	0.216	0.192	0.169	0.186

* TS: Time stamp

Table 10: Comparison of results obtained using Qds criterion for the CollegeMsg dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	0.200	**0.219**	0.186	0.170	0.166	0.171	0.180
2	**0.003**	0.001	0.001	0.000	0.000	0.000	0.003
3	**0.002**	0.001	0.001	0.000	0.001	0.001	0.001
4	**0.001**	0.000	0.000	0.000	0.000	0.000	0.000
5	0.339	**0.373**	0.190	0.243	0.269	0.195	0.345
6	**0.208**	0.129	0.120	0.125	0.136	0.222	0.208
7	**0.286**	0.206	0.204	0.162	0.237	0.284	0.274

* TS: Time stamp

Table 11: Comparison of results obtained using Qds criterion for the sx-mathoverflow dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.566**	0.182	0.272	0.272	0.316	0.326	0.410
2	0.000	0.000	0.127	0.041	0.151	0.121	**0.160**
3	0.001	**0.001**	0.000	0.00	0.000	0.000	0.001
4	0.002	**0.003**	0.001	0.001	0.001	0.001	0.002
5	0.075	**0.132**	0.106	0.059	0.064	0.060	0.124
6	**0.208**	0.0001	0.195	0.074	0.049	0.009	0.197
7	**0.157**	0.014	0.140	0.109	0.034	0.018	0.136
8	0.010	0.016	0.001	0.003	0.010	0.004	**0.016**

* TS: Time stamp

Considering the obtained results, it can be observed that the proposed method has demonstrated better performance compared to the compared methods on the introduced datasets, as well as with various evaluation criteria.

Finally, it is interesting to compare the execution time of different methods. Table 12 to 14 summarize the results for different datasets.

Table 12: Comparison of execution time (seconds) for the cit-Hep Ph dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**0.06**	0.07	1.3	2.5	4.4	3.2	1.3
2	**3.6**	4.4	6.6	8.7	9.1	6.2	4.6
3	**17.1**	59.8	63.3	63.1	63.0	64.9	46.7
4	**51.9**	91.6	93.6	93.5	96.7	94.2	68.7
5	**95.7**	126.5	129.1	128.6	130.8	129.1	140.3
6	**146.6**	182.6	183.9	187.1	185.8	185.9	168.1
7	250.2	**246.6**	248.0	250.7	251.0	248.1	269.3
8	**145.6**	181.5	184.6	184.4	186.3	183.5	191.9

* TS: Time stamp

Table 13: Comparison of execution time (seconds) for the CollegeMsg dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	0.8	**0.3**	0.6	1.3	0.6	1.6	0.4
2	**18.3**	19.6	20.4	21.0	20.8	11.1	16.4
3	**37.6**	40.9	41.5	42.1	42.1	41.6	40.6
4	**13.6**	14.4	14.8	14.8	15.5	18.4	16.5
5	3.3	**3.1**	4.0	3.6	3.8	4.8	3.6
6	1.4	**1.1**	2.1	1.7	1.5	2.6	2.0
7	0.7	**0.2**	1.6	0.7	0.8	1.5	0.6

* TS: Time stamp

Table 14: Comparison of execution time (seconds) for the sx-mathoverflow dataset

TS*	DIC	D-louvain	DynaMo	IncNSA	ECD	DPC-DLP	BBTA
1	**14.0**	21.9	22.7	23.0	22.9	23.4	20.6
2	**300.8**	460.5	472.7	519.9	541.2	491.8	521.6
3	**997.6**	1515.3	1291.2	1371.3	1487.2	1871.8	1418.6
4	**1198.7**	1804.8	1336.0	1425.1	1745.3	1409.8	1245.7
5	**1048.2**	1573.9	1741.5	1575.0	1322.8	1612.8	1510.0
6	**1002.1**	1503.9	1505.0	1504.8	1504.6	1505.3	1469.1
7	**986.0**	1479.4	1480.1	1480.2	1479.8	1480.9	1341.1
8	**789.4**	1184.1	923.2	1012.5	964.2	811.1	794.0

* TS: Time stamp

According to the obtained results, the proposed method has been reported lower execution time in comparison with other methods.

Conclusion and Future Works

In this paper, we have presented a method for detecting communities in dynamic networks based on influential nodes. Considering that communities and groups formed in a network in the real world can overlap, another goal of this paper is to provide a method to address community overlaps. Through this method, an attempt is made to preserve communities that have low overlap and allow their members to belong to both overlapping communities. To the best of our knowledge, a method for detecting communities in dynamic networks that simultaneously utilizes influential nodes to determine communities while considering overlapping communities has not been presented so far. Furthermore, the proposed method supports all possible changes in dynamic networks, including the addition and removal of nodes as well as the addition and removal of edges in dynamic networks. The proposed algorithm accomplishes the desired objectives through six phases: obtaining a snapshot of the network, selecting influential nodes based on local and global information, initialization, community expansion, evaluation and merging of communities, and finally updating communities. To evaluate the performance of the proposed method, we compared it with five recently proposed methods using three modularity metrics.

Based on the results obtained, the proposed algorithm has managed to achieve better results in most cases compared to the compared algorithms, especially in terms of modularity metrics and execution time. The reason behind this success could be attributed to the utilization of influential nodes in community formation. In the proposed algorithm, two metrics, namely node degree and k-shell decomposition, are used to determine influential nodes, both of which focus on node degree. As a result, the algorithm aims to shape the community around nodes that have the most connections to other nodes. On the other hand, the proposed method attempts to form a community around nodes that have the potential to establish a community, meaning they have high influence and lack connections with current communities. Ultimately, communities capable of merging based on the degree of congruence and overlap rate are merged with each other. Consequently, the proposed method retains only those communities that, on the one hand, have the potential to form a community and, on the other hand, cannot merge with other communities. In contrast, in other methods, attempts are made to use baseline modularity as a criterion for deciding whether to merge two communities or form a new one. Moreover, no clear and coherent idea exists regarding which nodes should be recognized as the center of a community. Only if the algorithm determines that separating one or more nodes would improve the

baseline modularity, will the operation of forming a new community take place. Furthermore, in the evaluation and community merging phase, the algorithm addresses inter-community edges and merges communities with overlapping members as much as possible. Therefore, in terms of modularity metrics, the proposed algorithm has achieved more favorable results compared to other methods.

Despite all the advantages of the proposed method, there are still challenges that are posed as open research issues. Firstly, we should emphasize that rapid identification of a change in the network and its application to communities in real-time might be closer to real-world applications. Thus, it is interesting that extend our method to detect the communities in real-time. Secondly, in this paper, we used two metrics, k-shell and node degree, for identifying influential nodes. The k-shell metric itself has high computational and time complexity. Therefore, considering other metrics and examining their results in the created communities could be another subject in this field.

Author Contributions

S. Baradaran Nejad implemented the methods and evaluated their performance. M. Sabzekar wrote the paper, coordinated the study and contributed to the analysis of the results. M. Khazaeipoor edited the manuscript. All authors revised and discussed the results and approved the final manuscript.

Acknowledgment

This work is completely self-supporting, thereby no any financial agency's role is available.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

LI	local information
DN	Dynamic Network
SN	Social Network

References

[1] B. Yang, D. Liu, J. Liu, "Discovering communities from social networks: Methodologies and applications," Handb. Soc. Netw. Technol. Appl.: 331–346, 2010.

[2] S. Souravlas, S. D. Anastasiadou, T. Economides, S. Katsavounis, "Probabilistic community detection in social networks," IEEE Access, 11: 25629–25641, 2023.

[3] A. Abbasi, H. Chen, A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums," ACM Trans. Inf. Syst., 26(3): 1–34, 2008.

[4] R. K. Bakshi, N. Kaur, R. Kaur, G. Kaur, "Opinion mining and sentiment analysis," in Proc. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 16: 452-455, 2016.

[5] T. Ma, Q. Liu, J. Cao, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan, "LGIEM: Global and local node influence-based community detection," Futur. Gener. Comput. Syst., 105: 533–546, 2020.

[6] N. Chen, Y. Liu, J. Cheng, Q. Liu, "A novel parallel community detection scheme based on label propagation," World Wide Web, 21: 1377–1398, 2018.

[7] Y. Niu, D. Kong, L. Liu, R. Wen, J. Xiao, "Overlapping community detection with adaptive density peaks clustering and iterative partition strategy," Expert Syst. Appl., 213: 119213, 2023.

[8] A. Reihanian, M. R. Feizi-Derakhshi, H. S. Aghdasi, "An enhanced multi-objective biogeography-based optimization for overlapping community detection in social networks with node attributes," Inf. Sci. (Ny)., 622: 903-929, 2023.

[9] P. Agarwal, R. Verma, A. Agarwal, T. Chakraborty, "DyPerm: Maximizing permanence for dynamic community detection," in Proc. Pacific-Asia conference on knowledge discovery and data mining: 437-449, 2018.

[10] X. Zeng, W. Wang, C. Chen, G. G. Yen, "A consensus community-based particle swarm optimization for dynamic community detection," IEEE Trans. Cybern., 50(6): 2502–2513, 2020.

[11] S. Ahajjam, M. El Haddad, H. Badir, "A new scalable leader-community detection approach for community detection in social networks," Soc. Networks, 54: 41-49, 2018.

[12] K. Dasgupta et al., "Social ties and their relevance to churn in mobile telecom networks," in Proc. 11th international conference on Extending database technology: Advances in database technology: 668–677, 2008

[13] M. A. Al-Garadi et al., "Analysis of online social network connections for identification of influential users: Survey and open research issues," ACM Comput. Surv., 51(1): 1-37, 2018.

[14] Y. Zhao, "A survey on theoretical advances of community detection in networks," Wiley Interdiscip. Rev. Comput. Stat., 9(5): 1403, 2017.

[15] S. Bahadori, H. Zare, P. Moradi, "PODCD: Probabilistic overlapping dynamic community detection," Expert Syst. Appl., 174: 114650, 2021.

[16] N. Chen, B. Hu, Y. Rui, "Dynamic network community detection with coherent neighborhood propinquity," IEEE Access, 8: 27915-27926, 2020.

[17] B. S. Khan, M. A. Niazi, "Network community detection: A review and visual survey," arXiv Prepr. arXiv1708.00977, 2017.

[18] R. Cazabet, G. Rossetti, F. Amblard, "Dynamic community detection," In: Alhajj, R., Rokne, J. (eds) Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY., 2017.

[19] G. Rossetti, R. Cazabet, "Community discovery in dynamic networks: a survey," ACM Comput. Surv., 51(2): 1-37, 2018.

[20] S. Souravlas, S. Anastasiadou, S. Katsavounis, "A survey on the recent advances of deep community detection," Appl. Sci., 11(16): 7179, 2021.

[21] D. Zhuang, J. M. Chang, M. Li, "DynaMo: Dynamic community detection by incrementally maximizing modularity," IEEE Trans. Knowl. Data Eng., 33(5): 1934-1945, 2019.

[22] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, A. Baig, "Community detection in networks: A multidisciplinary review," J. Netw. Comput. Appl., 108: 87-111, 2018.

[23] J. Scripps, "Discovering influential nodes in social networks through community finding.," in Proc. WEBIST: 403-412, 2013.

[24] J. Dai et al., "Identifying influential nodes in complex networks based on local neighbor contribution," IEEE Access, 7: 131719-131731, 2019.

[25] X. K. Zhang, J. Ren, C. Song, J. Jia, Q. Zhang, "Label propagation algorithm for community detection based on node importance and label influence," Phys. Lett. A, 381(33): 2691-2698, 2017.

[26] C. Li et al., "NANI: an efficient community detection algorithm based on nested aggregation of node influence," in Proc. the ACM Turing 50th Celebration Conference-China: 1-10, 2017.

[27] M. Cordeiro, R. P. Sarmento, J. Gama, "Dynamic community detection in evolving networks using locality modularity optimization," Soc. Netw. Anal. Min., 6: 1-20, 2016.

[28] W. Li, X. Zhou, C. Yang, Y. Fan, Z. Wang, Y. Liu, "Multi-objective optimization algorithm based on characteristics fusion of dynamic social networks for community discovery," Information Fusion, 79: 110–123, 2022.

[29] Q. Ni, J. Guo, W. Wu H. Wang, "Influence-Based community partition with sandwich method for social networks," IEEE Trans. Comput. Social Syst., 10: 2: 819-830, 2023.

[30] H. Long, X. Li, X. Liu, X. et al., "BBTA: Detecting communities incrementally from dynamic networks based on tracking of backbones and bridges," Applied Intelligence, 53:1084–1100, 2023.

[31] S. A. Seyedi, A. Lotfi, P. Moradi, N. N. Qader, "Dynamic graph-based label propagation for density peaks clustering," Expert Syst. Appl., 115: 314–328, 2019.

[32] F. Liu, J. Wu, C. Zhou, J. Yang, "Evolutionary community detection in dynamic social networks," in Proc. 2019 International Joint Conference on Neural Networks (IJCNN): 1–7, 2019.

[33] X. Su, J. Cheng, H. Yang, M. Leng, W. Zhang, X. Chen, "IncNSA: Detecting communities incrementally from time-evolving networks based on node similarity," Int. J. Mod. Phys. C, 31(7): 2050094, 2020.

Biographies

Mostafa Sabzekar received the B.S. degree in computer engineering from Kharzami University of Tehran, Iran, in 2007, and the M.S. and Ph.D. degrees in Computer Engineering from Ferdowsi University of Mashhad, Iran, in 2009, and 2017, respectively. He is currently an assistant professor at Birjand University of Technology, Birjand, Iran. His research focuses mainly on machine learning, evolutionary computation, and Bioinformatics.

- Email: Sabzekar@birjandut.ac.ir
- ORCID: 0000-0002-6886-1240
- Web of Science Researcher ID: NA
- Scopus Author ID: 35796344600
- Homepage: https://cv.birjandut.ac.ir/sabzekar/en

Shima Baradaran Nezhad was born in Birjand, Iran. She received the B.Sc. degree in Computer Engineering, from University of Birjand, Iran, in 2007. She received the M.Sc. dgree in Computer Engineeringtion from Birjand Branch, Islamic Azad University, Birjand, Iran, in 2022. She is currently an employee of the General Department of Natural Resources and Watershed Management of South Khorasan, Iran. Her research interests include Pattern Recognition and Algorithms.

- Email: sh.baradaran@frw.ir
- ORCID: NA
- Web of Science Researcher ID: NA
- Scopus Author ID: 35796344600
- Homepage: NA

Mahdi Khazaiepoor recived the B. Sc. Degree in Software Engineering from Islamic Azad university of mashhad, iran, in 2003 and M.Sc. degree in Software Engineering from Islamic Azad University Science and Research Branch, Tehran, Iran, in 2008, and Ph. D. degree in Software Engineering- Development of software systems from Islamic Azad University kerman branch, kerman, Iran. He is He is currently an assistant professor and the director of the computer department of the Islamic Azad University, Birjand branch, Iran. His research interest includes Data Mining, Software Effort Estimation and swarm intelligence algorithms.

- Email: mkhazaeipoor@iaubir.ac.ir
- ORCID: NA
- Web of Science Researcher ID: NA
- Scopus Author ID: 35796344600
- Homepage: NA

How to cite this paper:
M. Sabzekar, S. Baradaran Nezhad, M. Khazaeipoor, "A node-centric approach for community detection in dynamic networks," J. Electr. Comput. Eng. Innovations, 12(2): 305-318, 2024.

DOI: 10.22061/jecei.2024.10202.687

URL: https://jecei.sru.ac.ir/article_2043.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

WHOFWA: An Effective Hybrid Metaheuristic Algorithm Based on Wild Horse Optimizer and Fireworks Algorithm

A. Rouhi [*], ***E. Pira***

Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran.

Article Info

Article History:
Received 30 November 2023
Reviewed 30 December 2023
Revised 25 January 2024
Accepted 12 February 2024

Keywords:
Wild horse optimizer
Fireworks algorithm
CEC 2005
CEC 2019
CEC 2022
Convergence speed
Hit rate

[*]Corresponding Author's Email Address: *rouhi@azaruniv.ac.ir*

Abstract

Background and Objectives: This paper explores the realm of optimization by synergistically integrating two unique metaheuristic algorithms: The Wild Horse Optimizer (WHO) and the Fireworks Algorithm (FWA). WHO, inspired by the behaviors of wild horses, demonstrates proficiency in global exploration, while FWA emulates the dynamic behavior of fireworks, thereby enhancing local exploitation. The goal is to harness the complementary strengths of these algorithms, achieving a harmonious balance between exploration and exploitation to enhance overall optimization performance.

Methods: The study introduces a novel hybrid metaheuristic algorithm, WHOFWA, detailing its design and implementation. Emphasis is placed on the algorithm's ability to balance exploration and exploitation. Extensive experiments, featuring a diverse set of benchmark optimization problems, including general test functions and those from CEC 2005, CEC 2019, and 2022, assess WHOFWA's effectiveness. Comparative analyses involve WHO, FWA, and other metaheuristic algorithms such as Reptile Search Algorithm (RSA), Prairie Dog Optimization (PDO), Fick's Law Optimization (FLA), and Ladybug Beetle Optimization (LBO).

Results: According to the Friedman and Wilcoxon signed-rank tests, for all selected test functions, WHOFWA outperforms WHO, FWA, RSA, PDO, FLA, and LBO by 42%, 55%, 74%, 71%, 48%, and 52%, respectively. Finally, the results derived from addressing real-world constrained optimization problems using the proposed algorithm demonstrate its superior performance when compared to several well-regarded algorithms documented in the literature.

Conclusion: In conclusion, WHOFWA, the hybrid metaheuristic algorithm uniting WHO and FWA, emerges as a powerful optimization tool. Its unique ability to balance exploration and exploitation yields superior performance compared to WHO, FWA, and benchmark algorithms. The study underscores WHOFWA's potential in tackling complex optimization problems, making a valuable contribution to the realm of metaheuristic algorithms.

Introduction

In recent years, there has been significant advancement in the field of metaheuristic optimization algorithms, with diverse algorithms demonstrating success in addressing complex optimization problems [1].

Swarm-based optimization methods, such as particle swarm optimization (PSO) [2], Harris's hawk optimization (HHO) [3], gazelle optimization algorithm (GOA) [4], to name a few, have shown limitations in handling both simple and intricate problems. Notably, widely

recognized algorithms like gray wolf optimizer (GWO) [5], whale optimization algorithm (WOA) [6], and moth–flame optimization (MFO) [7] share structural similarities. Despite their popularity, the NFL theorem posits that no single algorithm can effectively address all optimization problems [8]. Each algorithm comes with specific strengths and limitations that impact their efficacy across various problem domains. Recognizing this, researchers have explored hybrid approaches, aiming to harness the strengths of multiple algorithms and create more robust optimization tools.

The Wild Horse Optimizer (WHO) has demonstrated efficiency in tackling specific optimization problems [9], yet its adaptability to a broader range of landscapes is a consideration. Similarly, the Fireworks Algorithm (FWA) shows promise in solving complex problems, but consistent delivery of optimal solutions across various functions remains a challenge [10].

Motivated by the aim to overcome the individual limitations of WHO and FWA, we introduce the Wild Horse and Fireworks Algorithm (WHOFWA) as a hybrid metaheuristic amalgamating the strengths of both algorithms. The goal is to leverage WHO's adaptability and FWA's exploration-exploitation balance, offering a comprehensive solution excelling in diverse optimization challenges. In response to challenges faced by traditional optimization methods, including difficulties in exploration-exploitation balance and sensitivity to parameters, the motivation behind developing WHOFWA lies in addressing these specific issues. The integration of WHO's adaptability and FWA's exploration capabilities positions WHOFWA as a comprehensive solution. This paper presents a detailed investigation into the development, performance, and versatility of WHOFWA, showcasing its potential to emerge as a leading metaheuristic algorithm in the domain of optimization.

To substantiate the effectiveness of WHOFWA, we conducted a comprehensive evaluation encompassing a diverse set of benchmark optimization problems. These include 10 general test functions, 7 test functions from the CEC 2005, 9 test functions from the CEC 2019 competition, and 12 test functions from the CEC 2022 competition. WHOFWA's performance is meticulously benchmarked against both WHO and FWA, alongside several other contemporary algorithms such as the Reptile Search Algorithm (RSA) [11], Prairie Dog Optimization (PDO) [12], Fick's Law Optimization (FLA) [13], and Ladybug Beetle Optimization (LBO) [14]. The evaluation is conducted using a variety of performance metrics, including the proximity to optima, early convergence, and hit rate (accuracy).

The results of our rigorous experimentation affirm the superior performance of WHOFWA. Not only does it outperform WHO and FWA, but it also surpasses the capabilities of RSA, PDO, FLA, and LBO across multiple facets of optimization. In an era where efficiency and accuracy are paramount, WHOFWA emerges as a promising and versatile solution that offers profound implications for optimization problems across diverse domains. This paper delves into the intricate design and operational intricacies of WHOFWA, presenting a compelling case for its adoption as a powerful tool in the ever-expanding toolkit of optimization practitioners.

The organization of this paper is as follows. Initially, a review will be conducted on some of the popular nature-inspired optimization algorithms. Following that, a brief introduction to the Fireworks and Wild Horse Optimization algorithms will be presented in the Preliminaries section to motivate their combination in the proposed WHOFWA algorithm. Subsequently, a detailed presentation and discussion of the hybrid algorithm will be provided. The Experimental Results section will present the evaluation framework of the proposed algorithm in comparison to some of the recently presented popular algorithms. Finally, the paper will be concluded, and some guidelines for future works will be provided.

Literature Review

Optimization, as a fundamental pursuit in problem-solving across various domains, has witnessed a surge in the development of metaheuristic algorithms (MAs) [15]-[19].

MAs are broadly divided into two categories: individual-solution-based and population-based [20], [21]. The latter, known for its effectiveness in exploring and exploiting the search space to target global optima, begins with a randomly generated population of solutions [22], [23]. Population-based MAs draw inspiration from evolutionary processes, natural phenomena, and social behaviors. Evolutionary Algorithms (EAs), including Genetic Algorithm (GA) [24] and Genetic Programming (GP) [25], emulate natural evolution. Natural Phenomenon (NP) algorithms, such as Simulated Annealing (SA) [26], Energy Valley Optimizer (EVO) [27], and others [13], [28]-[31], leverage physical and chemical principles. Social Behaviors (SBs) algorithms are further divided into Swarm Intelligence (SI) algorithms and Human Behaviors (HB) algorithms. SI algorithms, like Ant Colony Optimization (ACO) [32], simulate self-organized behaviors observed in nature, while HB algorithms include PSO [33] and various others [4], [11], [12], [34]-[37] that mimic human and animal behaviors.

Referencing the vast landscape of MAs poses a challenging task [20]. For the sake of conciseness, this section critically examines existing nature-inspired optimization algorithms, highlighting their applications,

key features, and identified strengths/weaknesses. Additionally, it identifies research gaps in the domain of metaheuristic algorithms and emphasizes the innovations introduced by the WHOFWA algorithm.

A. Nature-Inspired Optimization Algorithms

The landscape of optimization techniques has been enriched by a plethora of nature-inspired algorithms, each drawing inspiration from different natural phenomena. Table 1 provides a summary of popular population-based nature-inspired algorithms, their nature inspirations, key features, and applications. These algorithms have found applications in diverse fields such as engineering, genetics, finance, robotics, and image processing [15], [20], [21].

B. Research Gaps in Metaheuristic Algorithm Domain

The ongoing research for algorithms that effectively balance global exploration and local exploitation is a central focus in optimization research [1]. Despite the proliferation of nature-inspired algorithms, there exists a need for a critical analysis of their limitations and the identification of research gaps. Table 2 summarizes nature-inspired popular algorithms to highlight their unique strengths/weaknesses more clearly. This review recognizes the importance of addressing the limitations faced by popular algorithms, such as premature convergence, sensitivity to parameters, and inefficiency in high-dimensional problems. The identified gaps set the stage for the introduction of innovative solutions, motivating the development of hybrid metaheuristic algorithms like WHOFWA.

C. Innovations of WHOFWA in the Metaheuristic Algorithm Landscape

The WHOFWA algorithm emerges as a pioneering hybrid metaheuristic, combining the strengths of WHO and FWA. WHO excels in global exploration, while FWA specializes in dynamic local exploitation. The integration of these distinct approaches in WHOFWA aims to strike a harmonious equilibrium between exploration and exploitation.

Unlike existing algorithms, WHOFWA addresses the limitations of individual approaches by leveraging the collective intelligence of wild horses and the explosive search behavior of fireworks. This strategic fusion is designed to extend the boundaries of optimization capabilities, redefining the state-of-the-art in the field.

D. Synthesis of related work and WHOFWA's Contribution

The presented literature review critically evaluates existing algorithms, identifies research gaps, and positions WHOFWA as an innovative solution to address these gaps. The WHOFWA algorithm's ability to outperform not only WHO and FWA but also other contemporary algorithms underscores its significance in the landscape of metaheuristic algorithms. The following sections delve into the design, implementation, and comprehensive evaluation of WHOFWA, providing a thorough understanding of its capabilities and contributions.

Table 1: Summary of some popular nature-inspired optimization algorithms

Algorithm	Nature Inspiration	Key Features	Applications
Genetic Algorithm (GA) [38]	Genetics and Evolution	Population-based search, crossover and mutation operators, selection strategies	Engineering, genetics, finance
Particle Swarm Optimization (PSO) [2]	Flocking behavior	Swarm-based optimization, velocity updates, social and cognitive components	Robotics, clustering, neural networks
Ant Colony Optimization (ACO) [32]	Ant Foraging Behavior	Pheromone-based communication, path selection based on probability	Routing, network design, scheduling
Simulated Annealing (SA) [39]	Annealing Process	Temperature-based exploration, probabilistic acceptance of worse solutions	Combinatorial optimization, scheduling
Fireworks Algorithm (FWA) [10]	Firework Explosions	Population-based optimization, hierarchical explosions, more exploitation capability through good explosions	Image processing, feature selection
Artificial Bee Colony (ABC) [40]	Bee Foraging Behavior	Exploration via scout bees, local exploitation by employed bees, global exploitation by onlookers	Clustering, function optimization
Grey Wolf Optimizer (GWO) [5]	Grey Wolf Pack Behavior	Alpha, beta, delta, and omega wolves, mimicking social hierarchy	Engineering design, neural networks

Algorithm	Nature Inspiration	Key Features	Applications
Cuckoo Search (CS) [41]	Cuckoo Bird Nesting	Levy flights, host nest replacement, parameter tuning	Function optimization, image processing
Harmony Search (HS) [42]	Music Harmony Creation	Memory consideration, improvisation and exploration, harmony memory	Engineering design, parameter tuning
Bat Algorithm (BA) [43]	Bat Echolocation	Frequency tuning, pulse emission, adaptive loudness, and rate	Function optimization, feature selection
Whale Optimization Algorithm (WOA) [6]	Whale Hunting	Encircling prey, bubble-net hunting, exploration, and exploitation	Image processing, feature selection
Wild Horse Optimizer (WHO) [9]	Wild Horse Behavior	Herd movement, leader (stallions) and follower (foals) roles, herd updates based on stallion positions	Function optimization, engineering design
Artificial Immune System (AIS) [44]	Immune System Response	Clonal selection, mutation, antibody-antigen interactions	Anomaly detection, data mining
Moth Flame Optimization (MFO) [45]	Moth Attraction to Light	Attraction to light sources, update equations, exploration, exploitation	Image processing, function optimization
Flower Pollination Algorithm (FPA) [46]	Pollination Process	Pollination strategies, global and local pollination, pollen dispersal	Optimization, control systems
Krill Herd Algorithm (KHA) [47]	Krill Swarming Behavior	Krill movement, herding, search for prey, krill population	Function optimization, image segmentation
Dragonfly Algorithm (DA) [48]	Dragonfly Hunting	Prey detection, hunting strategy, escape behavior, position update	Image processing, function optimization
League Championship Algorithm (LCA) [49]	Sports Leagues	Competition, playoffs, promotion, and relegation mechanisms	Sports scheduling, tournament design
Social Spider Algorithm (SSA) [50]	Social Spider Behavior	Web construction, prey capture, web update, social cooperation	Network design, web optimization
Crow Search Algorithm (CSA) and Otsu Hybrid [51]	Crow's intelligent feeding behavior	Optimizes Otsu method for obtaining optimum thresholds, used for maximum variance within the cluster, simplicity with fewer parameters, decreased computational time in multilevel thresholding	Image segmentation with gray scale images, applicable in real-time applications
Ladybug Beetle Optimization (LBO) [14]	Ladybugs in Nature	Coordinated movement inspired by ladybugs searching for a warm place in winter, population update based on the position of other ladybugs, ignoring worst members to increase search speed	Applied to Economic-Environmental Dispatch Problem (EEDP), engineering applications, used for Covid-19 modeling and forecasting
Dandelion Optimizer (DO) [52]	Dandelion seed flight stages (rising, descending, landing)	Mathematical modeling of seed flight stages under different weather conditions, swarm intelligence based algorithm	Engineering optimization problems, real-world applications in speed reduction, spring design, beam design, and pressure vessel design
Barnacle Mating Optimizer (BMO) [53]	Barnacles' Mating Behavior	Bio-inspired evolutionary algorithm, Mimics mating process of barnacles, Utilizes Hardy-Weinberg principle in offspring generation	Real application in solving engineering optimization problems
Gazelle Optimization Algorithm (GOA) [4]	Gazelles life in nature and their survival ability	Population-based optimization method, two-phase approach: exploration (grazing) and exploitation (outrunning and outmaneuvering predators)	Optimization problems in various fields (engineering, medicine, computer science, production chain)

Table 2: Summary of strengths/weaknesses of popular nature-inspired optimization algorithms

Algorithm	Strengths	Weaknesses
Genetic Algorithm (GA) [38]	Parallelism, global optimization, robustness to noisy environments	Convergence speed may be slow, sensitive to parameter settings, not suitable for all optimization problems
Particle Swarm Optimization (PSO) [2]	Simple implementation, global optimization, fast convergence	Sensitive to parameter settings, may get stuck in local optima
Ant Colony Optimization (ACO) [32]	Good for discrete optimization problems, robust to changes in problem structure, inspired by natural systems	Limited scalability for large problems, may converge to suboptimal solutions
Simulated Annealing (SA) [39]	Global optimization, effective for complex landscapes, escapes local optima	Slow convergence, sensitive to temperature scheduling
Fireworks Algorithm (FWA) [10]	Good for multimodal optimization, diversity maintenance, parallel processing	Parameter tuning required, convergence speed may vary
Artificial Bee Colony (ABC) [40]	Simplicity and ease of implementation, handles continuous optimization problems, robust to parameter settings	Poor exploration/exploitation balance, sensitive to initial solution
Grey Wolf Optimizer (GWO) [5]	Mimics social behavior of wolves, good exploration/exploitation balance, simplicity and ease of implementation	Performance may degrade with large-scale problems, sensitive to parameter settings
Cuckoo Search (CS) [41]	Global optimization, simplicity and ease of implementation, handles non-continuous optimization problems	Convergence speed may vary, limited scalability for large problems
Harmony Search (HS) [42]	Good exploration/exploitation balance, versatile and easy to implement, effective for discrete optimization problems	Parameter tuning required, may converge to suboptimal solutions
Bat Algorithm (BA) [43]	Global optimization, good exploration/exploitation balance	Sensitive to parameter settings, limited scalability for large problems
Whale Optimization Algorithm (WOA) [6]	Global optimization, simplicity and ease of implementation	Convergence speed may vary, limited scalability for large problems
Wild Horse Optimizer (WHO) [9]	Global optimization, simplicity and ease of implementation	Convergence speed may vary, limited scalability for large problems
Artificial Immune System (AIS) [44]	Robustness to noisy environments, adaptive and self-learning	Computational complexity, difficulty in parameter tuning
Moth Flame Optimization (MFO) [45]	Good exploration/exploitation balance, simplicity and ease of implementation	Limited scalability for large problems, convergence speed may vary
Flower Pollination Algorithm (FPA) [46]	Global optimization, simplicity and ease of implementation	Limited scalability for large problems, convergence speed may vary
Krill Herd Algorithm (KHA) [47]	Good exploration/exploitation balance, simplicity and ease of implementation	Limited scalability for large problems, convergence speed may vary
Dragonfly Algorithm (DA) [48]	Good exploration/exploitation balance, simplicity and ease of implementation	Limited scalability for large problems, convergence speed may vary
League Championship Algorithm (LCA) [49]	Global optimization, simplicity and ease of implementation	Limited scalability for large problems, convergence speed may vary
Social Spider Algorithm (SSA) [50]	Good exploration/exploitation balance, inspired by social behavior	Limited scalability for large problems, convergence speed may vary

Preliminaries

This section first provides a description of original algorithms of the proposed hybrid algorithm WHOFWA, i.e., WHO and FWA, respectively, and then motivates the integration of them in WHOFWA.

A. Wild Horse Optimization (WHO)

Wild Horse Optimizer (WHO) algorithm, inspired by the social organization and behaviors of non-territorial horses [38]. The focus is on the optimization of various problems using group behaviors, grazing, mating,

domination, and leadership.

The core components of WHO include position updating mechanisms guided by individual and group influences, and adaptive parameters that enhance the algorithm's adaptability to different problem landscapes. WHO has demonstrated success in solving optimization problems, particularly in unimodal landscapes, thanks to its ability to efficiently explore the solution space.

The WHO algorithm can be summarized as shown in Algorithm 1.

B. Fireworks Algorithm (FWA)

The Fireworks Algorithm (FA) is a swarm intelligence algorithm inspired by the emergent swarm behavior of fireworks for function optimization [9]. It simulates the explosion process of fireworks, where two explosion (search) processes are employed to generate sparks. The algorithm aims to find an optimal point in the search space by iteratively setting off "fireworks," evaluating sparks' locations, and selecting new locations for the next generation of explosions based on the current sparks and fireworks.

Algorithm 1: Wild Horse Optimizer (WHO)

Initialize the first population of horses randomly
Calculate the fitness of horses
Create foal groups and select their Stallions as group leaders
Calculate the fitness of each horse in the herd and select the best one as *Optimum*
While the end criterion is not satisfied
  For each $Stallion_i$
    For each Foal member of the group leading by $Stallion_i$
      Randomly move the Foal to search around the leader, i.e., $Stallion_i$ Or leave the group and mate with the foals of other groups (which left their groups as well) through crossover aimed at exploring new spaces
    End For
    Randomly update the position of the $Stallion_i$ with ∓ WH, hopefully reaching water hole, i.e., best positions of search space
    If fitness of $Stallion_i$ is better **than** *Optimum* then Optimum will be updated to $Stallion_i$
    Sort the foals of $Stallion_i$'s group members based on their fitness values and exchange the position of best foal with the $Stallion_i$ to form a new group leader
  End For
  Update *Optimum*
End While

A *good firework explosion* is characterized by numerous sparks that centralize around the explosion center, creating a spectacular display. On the contrary, a *bad firework explosion* is defined by the generation of few sparks, which scatter in the space, resulting in a less impressive outcome.

Despite its strengths, FWA may encounter challenges in certain function landscapes, where the balance between exploration and exploitation becomes critical. The sensitivity of FWA to parameter settings and specific problem characteristics motivates the exploration of hybridization with other algorithms to enhance its overall performance.

The FWA algorithm can be summarized as shown in Algorithm 2.

C. Integration of WHO and FWA in WHOFWA

Motivated by the distinctive features of WHO and FWA, their hybridization in the Wild Horse and Fireworks Algorithm (WHOFWA) aims to capitalize on the strengths of both algorithms. WHO's efficiency in unimodal landscapes complements FWA's robust exploration in multimodal scenarios. WHOFWA integrates the collective movement and adaptability of wild horses from WHO with the explosive exploration capabilities inspired by fireworks in FWA. This amalgamation is expected to create a metaheuristic algorithm that not only addresses the limitations of its parent algorithms but also excels in tackling diverse optimization challenges.

The subsequent sections detail the design, implementation, and extensive experimental evaluations of WHOFWA, shedding light on its performance across a range of test functions and establishing its potential as a competitive and versatile optimization tool.

Algorithm 2: Fireworks Algorithm (FWA)

Randomly select ***n*** locations for fireworks.
While the *stop criteria* are **false**
  Set off n fireworks at the respective locations.
  For each firework x_i
    Calculate the number of sparks considering high numbers to the fireworks with better fitness.
    Obtain locations of sparks.
  Generate ***m*** sparks for some of randomly selected fireworks.
  Select the best location for the next explosion generation.
  Randomly select ***n − 1*** locations from the sparks and current fireworks based considering the high probability to the good firework explosions.
End While

The Proposed Algorithm

WHO algorithm has parameters *N* (total number of horses), *G = N * PS* (total number of groups), *PS* (percentage of male horses in the total population). As a

result, we will have *G* head group (male horse) and *N - G* remaining horses are divided between these groups, each group will have *(N - G) / G* members. Leaders are usually the best in groups. FA algorithm has *Nfws* number of firework and each firework produces a number of sparks.

In the proposed hybrid algorithm (WHOFWA), the heads of groups (male horses) play the role of sparks, and the rest of the horses of each group play the role of sparks on each firework.

Algorithm 3 shows the WHOFWA algorithm pseudocode. In this algorithm, *Nfws*, *maxEva*, and *gaussianNum* denote the number of fireworks, the maximum number of fitness function evaluations, and the number of solutions produced in a local Gaussian space around the fireworks using the Gaussian explosion process, respectively. In line 1, a population (*Fw*) of sparks with the size of *Nfws* is randomly generated and then in line 2, the fitness of sparks is calculated (*Fwfit*). In line 5 of the algorithm, the sonsnum_cal function calculates the number of sparks of each firework and places in the array sonsnum_array. In line 6, the scope_cal function calculates the explosion range of sparks around each firework and stores it in scope_array. In line 7, the sons_generate function generates the sparks based on the number of sparks and the explosion range of each firecracker and keeps them in the Sons variable.

Algorithm 3: The **WHOFWA** algorithm

Input: *Nfws, maxEva*, *gaussianNum*;
Output: A solution with the best fitness.
1: Initialize the population of fireworks *Fw* (Fw_1, …, Fw_{Nfws});
2: Evaluate the fitness of each firework *Fwfit* ($Fwfit_1$, ..., $Fwfit_{Nfws}$);
3: *bst* = select the best firewrok; *evaCount* = *Nfws*;
4: **while** *evaCount* **<=** *maxEva* **do**
// compute the number of sons that each seed should generate
5: sonsnum_array = sonsnum_cal(*Fwfit*, *Nfws*);
// compute the exploding scope of sons that each seed generate
6: scope_array = scope_cal(*Fwfit*, *Nfws*);
// generate the sparks, based on the sparks number and explosion amplitude
// of each firework
7: *Sons*= sons_generate (sonsnum_array, scope_array, *Fw*,*bst*);
8: Evaluate the fitness of each spark *Sonsfit* ($Sonsfit_1$, …, $Sonsfit_{Nfws}$);
9: *evaCount* = *evaCount* + *Nfws*;
//////// Wild horse optimizer ///////////
// All generated fireworks are considered Stallions
// All sparks of each firework are considered foals for a Stallion
10: *Nstallion* = *Nfws*; *Stallion* = *Fw*; *Stallionfit* = *Fwfit;*
11: **for** i = 1 to *Nfws* **do**
12: *Nc* = sonsnum_array (i) ; *sp* = 0;
13: **for** j=1 to i-1**do**
14: *sp* = *sp* + sonsnum_array (j);
15: **end for**
16: *Stallion* (i) . *group* = *Sons* (*sp* +1: *sp* + *Nc*);
17: *Stallion* (i) . *groupfit* = *Sonsfit* (*sp* +1: *sp* + *Nc*);
18: **end for**
19: *MaxiterWHO* =1;
20: *StallionNew* = **WHO** (*Stallion* , *MaxiterWHO*);
21: *evaCount* = *evaCount* + length (*StallionNew*);
//
21: Update *Fw* and *Sons* if *StallionNew* is better than *Stallion*;
// perform the Gaussian mutation of seeds
22: *FwGauss*= seedGaussMutation(*Fw*, *bst*, *gaussianNum*);
23: Evaluate the fitness of each Gaussian affected; *evaCount* = *evaCount* + *gaussianNum;*
24: *All* = *Fw* ∪ *Sons* ∪ *FwGauss; Allfit* = *Fwfit* ∪ *Sonsfit* ∪ *FwGaussfit*;
// select the next iteration
25: *Fw* = selectNextIterationOnEntropy(*All*, *Allfit*, *Nfws*);
26: Update *bst* due to *Fw*;
27: **end while**
28: **return** *bst*;

Lines 10-21 show the initialization and execution of the WHO algorithm. In line 10, the number of stallions (*Nstallion*) is equal to the number of fireworks (*Nfws*) and the fireworks and their fitness values are copied into the *Stallion* and *Stallionfit* variables, respectively. In lines 11-18, all sparks are copied into the group variable of *Stallion* as Foals. In line 20, the WHO algorithm is invoked and executed one time. In line 21, the solutions returned from the execution of the WHO algorithm are copied into the *FW* and *Sons* variables if they are better. In line 22, Gaussian mutation is performed on the solutions in *Fw* with the seedGaussMutation function. In line 26, among the solutions available in *Fw*, *Sons* and *FwGauss*, number of *Nfws* solution is selected as fireworks of the next step.

Experimental Results

In order to rigorously assess the performance of our presented hybrid metaheuristic algorithm, WHOFWA, and to establish its superiority over existing optimization algorithms, we conducted a comprehensive set of experiments on a range of benchmark optimization problems [54], [55]. For all experiments, the maximum number of function evaluations (*maxEva*) was set to 10^5. The parameter settings for all algorithms used in the experiments are shown in Table 3. From parameters of the WHOFWA algorithm, listed in Table 3, the following two parameters have a great impact on the effectiveness of this algorithm: *Nfws* (the number of fireworks) and *Mn* (maximum number of sparks). In order to analyze this impact, the WHOFWA is executed for various *Nf* and *Mn* values, considering F1 and F2 from the general test functions, F4 and F5 from the CEC 2019 functions, and F1 and F3 from the CEC 2022 functions. As indicated in Table 4, the best effectiveness for WHOFWA is obtained when *Nfws* and *Mn* are equal to 3 and 40, respectively.

The parameter settings were carefully chosen to ensure a fair and standardized comparison between the WHOFWA algorithm and the benchmark algorithms. Each algorithm was executed on the same set of optimization problems with these settings, and the results were recorded and analyzed.

In the subsequent sections, we present the results of these experiments, providing a detailed analysis of the performance of WHOFWA in comparison to WHO, FA, RSA, PDO, FLA, and LBO. These results are pivotal in demonstrating the superior efficacy of WHOFWA as an optimization tool in various contexts.

Experiment 1: 10 General Test Functions

In the pursuit of evaluating the performance of the WHOFWA algorithm and comparing it to other state-of-the-art optimization algorithms, we begin by examining its behavior on a set of 10 general test functions (see Table 5 for more details). These functions are well-established benchmarks in the field of optimization and are often used to assess the capabilities of optimization algorithms in terms of convergence speed, accuracy, and their ability to navigate various landscapes.

Table 3: Parameter settings of all algorithms (*maxEva* = 10E+5)

Algorithm	Appropriate parameter values
WHOFWA	*The number of fireworks* = 3, *The total number of sparks* = 40, *N* = 30, *Crossover rate* = 0.13, *Stallions percentage* = 0.1
WHO	*Crossover rate* = 0.13, *Stallions percentage* =0.2, *N* = 50
FA	*The number of fireworks* = 5, *The total number of sparks* = 50
RSA	*Alpha* = 0.1, *Beta* = 0.005, *N* = 30
PDO	*Rho* = 0.005, *epsPD* = 0.1, *N* = 30
FLA	*C1* = 0.5, *C2* = 2, *C3* = 0.1, *C4* = 0.2, C5 = 2, *D* = 0.01, *N* = 30
LBO	*N (0)= 60*, *β* =10, N_{min} = 0.25*N*(0)

For each of these test functions, the WHOFWA algorithm, alongside other benchmark algorithms (WHO, FA, RSA, PDO, FLA, and LBO), is subjected to rigorous evaluation. The performance metrics considered include proximity to the optimum, early convergence, and hit rate (accuracy). The results of this experiment provide crucial insights into the ability of WHOFWA to solve general optimization problems and its relative performance when compared to established algorithms. These findings will be presented in the subsequent sections, shedding light on the efficacy and versatility of WHOFWA in tackling various optimization landscapes.

Table 4: The results of executing WHOFWA for different *Nf and Mn*

Type	Function	Metric	*Nfws* = 2 *Mn* = 30	*Nfws* = 2 *Mn* = 40	*Nfws* = 2 *Mn* = 50	*Nfws* = 3 *Mn* = 30	*Nfws* = 3 *Mn* = 40	*Nfws* = 3 *Mn* = 50	*Nfws* = 4 *Mn* = 30	*Nfws* = 4 *Mn* = 40	*Nfws* = 4 *Mn* = 50
General	F1	Best	-1.00E+00	-1.00E+00	-1.00E+00	-1.00E+00	-1.00E+00	-1.00E+00	-1.00E+00	-1.00E+00	-1.00E+00
	F2	Best	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
CEC 2019	F4	Best	23.1441	22.6695	14.7374	24.1994	15.8521	11.3147	41.8829	30.8967	18.0331
	F5	Best	1.138	1.196	1.1591	1.1621	1.1001	1.138	1.3331	1.1666	1.1538
CEC 2022	F1	Best	319.6524	313.4741	324.1513	310.4829	307.2631	311.3363	314.2728	316.3664	314.5829
	F3	Best	600.0216	600.0042	600.0192	600.0960	600.0021	600.0222	600.0607	600.2766	600.1084

Table 5: The details of general test functions

Name	Formulation	Dimension (D)	Range	Optimum value	Type
F1 (Easom)	$f(X) = -\cos(X_1)\cos(X_2)\exp(-(X_1-\pi)^2 - (X_2-\pi)^2)$	2	[-100, 100]	-1	**Unimodal**
F2 (Sphere)	$f(X) = \sum_{i=1}^{D} {X_i}^2$	30	[-100, 100]	0	
F3 (SumSquares)	$f(X) = \sum_{i=1}^{D} i{X_i}^2$	30	[-10, 10]	0	
F4 (Schwefel 2.22)	$f(X) = \sum_{i=1}^{D} \lvert X_i\rvert + \prod_{i=1}^{D} \lvert X_i\rvert$	30	[-10, 10]	0	
F5 (Schwefel 1.2)	$f(X) = \sum_{i=1}^{D} (\sum_{j=1}^{i} X_j)^2$	30	[-100, 100]	0	
F6 (Bohachevsky1)	$f(X) = X_1^2 + 2X_2^2 - 0.3cos(3\pi X_1) - 0.4cos(4\pi X_2) + 0.7$	2	[-100, 100]	0	**Multimodal**
F7 (Booth)	$f(X) = (X_1 + 2X_2 - 7)^2 + (2X_1 + X_2 - 5)^2$	2	[-10, 10]	0	
F8 (Schaffer)	$f(X) = 0.5 + \frac{sin^2\left(\sqrt{X_1^2 + X_2^2}\right) - 0.5}{(1 + 0.001(X_1^2 + X_2^2))^2}$	2	[-100, 100]	0	
F9 (Michalewicz10)	$f(X) = -\sum_{i=1}^{D} \sin(X_i)(\sin(iX_i^2/\pi))^{20}$	10	[0, π]	-9.6602	
F10 (Griewank)	$f(X) = \frac{1}{4000}\left(\sum_{i=1}^{D}(X_i - 100)^2\right) - \left(\prod_{i=1}^{D} \cos(\frac{X_i - 100}{\sqrt{i}})\right) + 1$	30	[-600, 600]	0	

In Table 6, we present the comparative results for the 10 general test functions, showcasing the performance of the WHOFWA algorithm in comparison to other optimization algorithms, namely WHO, FA, RSA, PDO, FLA, and LBO. The results are categorized based on three key metrics: “Best” (the best solution found) and “Ave” (the average solution found), and “Std” (standard deviation) of the solutions. Let's delve into the analysis and discussion of these results:

1. **F1 (Easom)**: For the unimodal function F1, WHOFWA demonstrates its ability to reach the global optimum, achieving a best solution of -1.00E+00, matching the best result found by WHO and FA. The average solution is also -1.00E+00. WHOFWA maintains an impressively low standard deviation of 7.03E-07, indicating its consistency.
2. **F2 (Sphere)**: In the case of the spherical function F2, WHOFWA stands out by discovering the best solution of 2.60E-279, surpassing all other algorithms, which converge to zero. The average solution for WHOFWA is significantly better than other algorithms.
3. **F3 (SumSquares)**: WHOFWA once again excels by finding the global optimum (best) of 0.00E+00 for the SumSquares function. This is in contrast to other algorithms that also reach zero. The average result for WHOFWA is commendable, reflecting its ability to converge to the optimum consistently.
4. **F4 (Schwefel 2.22)**: WHOFWA continues to demonstrate its competitiveness with a best solution of 5.67E-143, which is notably better than the results of other algorithms. The average solution, while not reaching zero, is still quite low.
5. **F5 (Schwefel 1.2)**: WHOFWA excels in reaching the best solution of 3.27E-253, which is superior to other algorithms. The average result for WHOFWA is also impressive, outperforming the competitors.
6. **F6 (Bohachevsky1)**: For the multimodal function F6, WHOFWA manages to achieve the best solution of 0.00E+00, matching other algorithms. It

maintains a low standard deviation, reflecting consistent convergence.

7. **F7 (Booth)**: WHOFWA finds a competitive best solution of 2.14E-08, though it is not the best result. The average result is comparable to the performance of other algorithms.
8. **F8 (Schaffer)**: WHOFWA reaches the best solution of 0.00E+00, matching the results of other algorithms. The average solution is also in line with the performance of the competitors.
9. **F9 (Michalewicz10)**: WHOFWA stands out in finding the best solution of -9.66E+00, which is notably better than the other algorithms. The average solution is competitive, demonstrating the algorithm's consistency.
10. **F10 (Griewank)**: WHOFWA converges to the best solution of 0.00E+00, which matches the performance of other algorithms. The average solution is also in line with the results of competitors.

In sum, WHOFWA consistently exhibits competitive performance across the majority of the general test functions. It excels in finding the global optima for several unimodal and multimodal functions, with a remarkable ability to consistently reach or approach optimal values. The low standard deviations suggest that WHOFWA maintains stability and consistency in its results.

In conclusion, the results in Table 6 demonstrate that WHOFWA is a robust optimization algorithm that excels in solving a diverse set of general test functions. Its competitive performance positions it as a promising tool for tackling a wide range of real-world optimization problems, further substantiating its effectiveness and versatility in the field of metaheuristic algorithms.

In Table 7, the ranks of various algorithms, including WHOFWA, WHO, FA, RSA, PDO, FLA, and LBO, are provided for a set of general test functions. The rankings are based on the Friedman Test, and the lower the mean rank, the better the algorithm's overall performance across the tested functions.

In sum, the results in Table 7 reflect the relative performance of the algorithms across the general test functions. While WHOFWA performed well, it was outperformed by both FLA and FWA, which achieved the highest and second-highest mean ranks, respectively. These rankings suggest that FLA and FWA were the top-performing algorithms in this evaluation, while WHOFWA displayed strong competitiveness in third place.

It's important to note that the choice of evaluation metrics and test functions can influence algorithm rankings. Nevertheless, the results in Table 6 provide valuable insights into the relative performance of WHOFWA and other metaheuristic algorithms, reaffirming their effectiveness in solving general optimization problems.

Experiment 2: CEC 2019 Test Functions

In this section, we shift our focus to the evaluation of the WHOFWA algorithm on a set of challenging optimization problems known as the CEC 2019 test functions. These functions are designed to rigorously test the performance of optimization algorithms across diverse landscapes and complexities, making them a robust benchmark for assessing the capabilities of WHOFWA (see Table 8 for more details). For each of these CEC 2019 test functions, we will assess the performance of WHOFWA and compare it to other state-of-the-art optimization algorithms. The evaluation will be based on the key performance metrics such as proximity to the optimum, early convergence, and hit rate (accuracy). These results will provide valuable insights into the ability of WHOFWA to tackle complex and diverse optimization landscapes, further demonstrating its effectiveness in challenging scenarios.

Table 9 presents a comprehensive overview of the performance of the WHOFWA algorithm, as well as other optimization algorithms (WHO, FA, RSA, PDO, FLA, and LBO), across the CEC 2019 test functions. The results are divided into three key metrics: “Best” (the best solution found), “Ave” (the average solution found), and “Std” (the standard deviation of the solutions). Here's a brief analysis and discussion of the results:

1. **F1 (Storn's Chebyshev Polynomial Fitting Problem)**: WHOFWA achieved a best solution of 3.92E+04, which is competitive with the other algorithms. The average solution for WHOFWA is also noteworthy, although it ranks third among the algorithms. The standard deviation for WHOFWA is relatively low, indicating consistent performance.
2. **F2 (Inverse Hilbert Matrix Problem)**: WHOFWA obtained the best solution of 1.73E+01, matching the best results of other algorithms. The average solution for WHOFWA is consistent with other algorithms, demonstrating good performance on this function. WHOFWA maintains a very low standard deviation, indicating stability in its results.
3. **F3 (Lennard-Jones Minimum Energy Cluster)**: WHOFWA achieved the best solution of 1.27E+01, on par with the best results of other algorithms. The average solution for WHOFWA is consistent with other algorithms, demonstrating strong performance. The standard deviation for WHOFWA is notably low, indicating stability.
4. **F4 (Rastrigin's Function)**: WHOFWA reached a best solution of 1.58E+01, which is competitive, although it doesn't outperform the best results of other algorithms. The average solution for

WHOFWA ranks among the algorithms, demonstrating good performance. The standard deviation for WHOFWA is reasonable, indicating a stable performance.

5. **F5 (Griewank's Function)**: WHOFWA achieved the best solution of 1.10E+00, which is competitive with other algorithms. The average solution for WHOFWA is strong, ranking among the algorithms. WHOFWA maintains a reasonable standard deviation, suggesting consistent performance.
6. **F6 (Weierstrass Function)**: WHOFWA reached a best solution of 2.48E+00, competitive with the other algorithms. The average solution for WHOFWA is on par with the performance of other algorithms. The standard deviation for WHOFWA is reasonable, indicating stability.
7. **F8 (Expanded Schaffer's F6 Function)**: WHOFWA obtained the best solution of 3.25E+00, on par with the best results of other algorithms. The average solution for WHOFWA ranks among the algorithms, demonstrating strong performance. The standard deviation for WHOFWA is reasonable, indicating a stable performance.
8. **F9 (Happy Cat Function)**: WHOFWA achieved the best solution of 2.42E+00, matching the best results of other algorithms. The average solution for WHOFWA ranks among the algorithms, demonstrating strong performance. The standard deviation for WHOFWA is low, indicating consistent performance.
9. **F10 (Ackley Function)**: WHOFWA reached a best solution of 2.34E+00, which is competitive with the other algorithms. The average solution for WHOFWA is consistent with the performance of other algorithms. The standard deviation for WHOFWA is notably low, indicating stability.

Table 6: Comparative results for general test functions

Function	Metric	WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
F1	Best	-1.00E+00	-1.00E+00	-1.00E+00	-9.99E-01	-1.00E+00	-1.00E+00	-1.00E+00
	Ave	-1.00E+00	-1.00E+00	-1.00E+00	-9.89E-01	-5.00E-01	-1.00E+00	-1.00E+00
	Std	7.03E-07	0.00E+00	7.24E-06	1.37E-02	7.07E-01	1.96E-04	7.03E-07
F2	Best	2.60E-279	1.82E-117	0.00E+00	0.00E+00	0.00E+00	0.00E+00	8.95E-75
	Ave	7.18E-202	1.21E-111	0.00E+00	0.00E+00	0.00E+00	0.00E+00	6.26E-74
	Std	0.00E+00	1.80E-111	0.00E+00	0.00E+00	0.00E+00	0.00E+00	7.59E-74
F3	Best	0.00E+00	6.08E-119	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.84E-80
	Ave	3.46E-209	5.34E-114	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.29E-78
	Std	0.00E+00	7.63E-114	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.59E-78
F4	Best	5.67E-143	5.53E-66	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.51E-49
	Ave	5.15E-120	3.92E-59	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.33E-48
	Std	1.96E-119	6.78E-59	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.66E-48
F5	Best	3.27E-253	3.19E-109	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.52E-73
	Ave	2.45E-205	1.44E-107	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.75E-72
	Std	0.00E+00	1.39E-107	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.27E-72
F6	Best	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Ave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Std	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F7	Best	2.14E-08	5.07E-01	1.17E-04	1.31E-01	5.66E-02	2.75E-05	0.00E+00
	Ave	3.13E-06	5.07E-01	2.66E-04	3.20E-01	2.82E-01	3.79E-01	2.54E-01
	Std	4.77E-06	0.00E+00	2.10E-04	2.68E-01	3.19E-01	3.31E-01	3.59E-01
F8	Best	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.37E-02
	Ave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.91E-02	4.37E-02
	Std	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.52E-02	3.19E-13
F9	Best	-9.66E+00	-9.07E+00	-9.57E+00	-5.44E+00	-8.01E+00	-8.01E+00	-9.61E+00
	Ave	-9.58E+00	-8.56E+00	-9.44E+00	-5.43E+00	-7.50E+00	-7.50E+00	-9.61E+00
	Std	8.03E-02	6.05E-01	1.85E-01	1.35E-02	1.71E+00	7.28E-01	1.50E-03
F10	Best	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Ave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.60E-02
	Std	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.26E-02

Table 7: Ranks of WHOFWA and other algorithms of general test functions

		WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
Friedman Test	Mean Rank	3.60	5.00	3.25	4.20	3.60	3.40	4.95
	Rank	**3**	**7**	**1**	**5**	**4**	**2**	**6**

In sum, the results in Table 8 showcase the strong and competitive performance of the WHOFWA algorithm across the CEC 2019 test functions.

WHOFWA consistently achieved best solutions that are on par with or better than other algorithms and demonstrated competitive average solutions.

Moreover, WHOFWA maintained low standard deviations, indicating stable and reliable performance across these complex optimization landscapes.

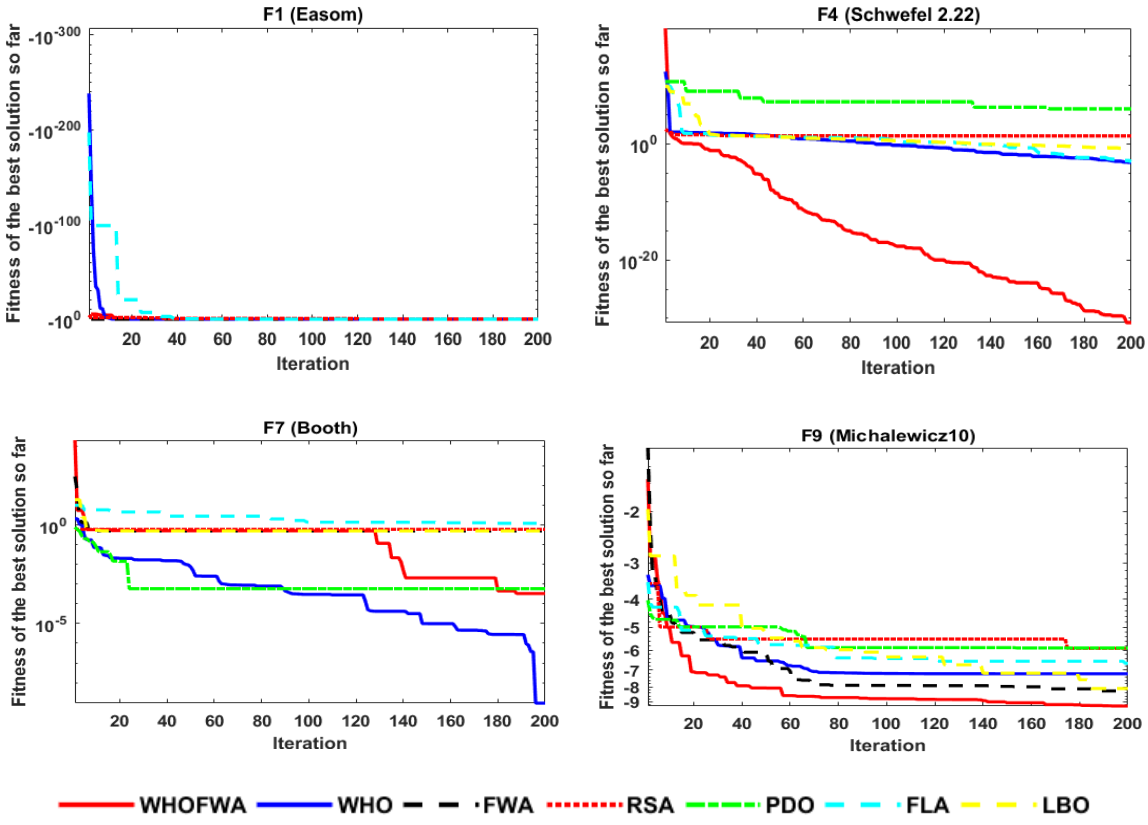


Fig. 1: Convergence curve of all algorithms on solving general test functions.

These results further emphasize the effectiveness and versatility of WHOFWA in solving challenging optimization problems, positioning it as a valuable tool for tackling real-world optimization scenarios and underlining its competitive nature among state-of-the-art optimization algorithms.

The results in Table 10 highlight the exceptional performance of WHOFWA, as it secured the top rank in the evaluation of CEC 2019 test functions.

This indicates that WHOFWA consistently outperformed the other algorithms, including WHO, FA, RSA, PDO, FLA, and LBO, across a diverse set of challenging optimization problems.

The low mean rank of WHOFWA reaffirms its competitive nature and positions it as a top-performing algorithm for solving complex optimization tasks.

It's important to consider that the choice of evaluation metrics and test functions can impact algorithm rankings.

Nonetheless, the results in Table 10 underscore the effectiveness of WHOFWA in tackling the CEC 2019 test functions and highlight its competitive edge in the realm of optimization algorithms.

Fig. 2 shows the convergence curve of all algorithms on solving the CEC 2019 test functions.

Table 8: The details of CEC 2019 test functions

Function	Name	D	Range	Optimum value
F1	Storn's Chebyshev Polynomial Fitting Problem	9	[-8192,8192]	1
F2	Inverse Hilbert Matrix Problem	16	[-16384,16384]	1
F3	Lennard-Jones Minimum Energy Cluster	18	[-4,4]	1
F4	Rastrigin's Function	10	[-100,100]	1
F5	Griewank's Function	10	[-100,100]	1
F6	Weierstrass Function	10	[-100,100]	1
F8	Expanded Schaffer's F6 Function	10	[-100,100]	1
F9	Happy Cat Function	10	[-100,100]	1
F10	Ackley Function	10	[-100,100]	1

Table 9: Comparative results for CEC 2019 test functions

Function	Metric	WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
F1	Best	3.92E+04	3.71E+04	4.57E+04	4.36E+04	4.62E+04	4.62E+04	6.68E+07
	Ave	4.76E+04	5.07E+04	4.70E+04	1.22E+05	5.65E+04	5.65E+04	1.77E+08
	Std	4.71E+03	2.16E+04	1.87E+03	1.11E+05	1.45E+04	1.45E+04	1.55E+08
F2	Best	1.73E+01	1.73E+01	1.73E+01	1.80E+01	1.74E+01	1.74E+01	1.73E+01
	Ave	1.73E+01	1.73E+01	1.73E+01	1.80E+01	1.74E+01	1.74E+01	1.73E+01
	Std	2.09E-05	0.00E+00	8.01E-04	1.70E-05	1.25E-02	6.15E-03	7.98E-08
F3	Best	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01
	Ave	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01
	Std	8.75E-10	0.00E+00	3.60E-08	6.93E-06	1.35E-07	1.76E-08	2.43E-08
F4	Best	1.58E+01	2.89E+01	7.33E+01	7.86E+03	4.34E+03	2.40E+01	6.00E+00
	Ave	4.62E+01	4.41E+01	7.46E+01	8.82E+03	1.16E+04	4.34E+01	1.05E+01
	Std	1.68E+01	2.64E+01	1.93E+00	1.35E+03	1.03E+04	2.64E+01	6.32E+00
F5	Best	1.10E+00	1.01E+00	1.34E+00	3.73E+00	3.31E+00	1.08E+00	1.08E+00
	Ave	1.31E+00	1.02E+00	1.38E+00	3.83E+00	3.37E+00	1.13E+00	1.15E+00
	Std	1.41E-01	2.09E-02	6.71E-02	1.41E-01	8.57E-02	4.13E-02	9.39E-02
F6	Best	2.48E+00	3.51E+00	5.47E+00	9.34E+00	8.74E+00	3.63E+00	2.87E+00
	Ave	4.73E+00	4.52E+00	5.91E+00	9.82E+00	8.88E+00	4.27E+00	3.49E+00
	Std	9.60E-01	8.84E-01	6.25E-01	6.81E-01	1.87E-01	6.28E-01	8.67E-01
F8	Best	3.25E+00	4.06E+00	-1.28E+01	5.51E+00	5.21E+00	5.04E+00	4.06E+00
	Ave	5.10E+00	4.41E+00	1.67E+02	6.05E+00	5.52E+00	5.52E+00	4.52E+00
	Std	6.57E-01	5.73E-01	2.54E+02	7.61E-01	4.41E-01	4.83E-01	6.60E-01
F9	Best	2.42E+00	2.35E+00	5.56E+00	1.48E+03	9.17E+02	2.47E+00	2.34E+00
	Ave	2.57E+00	2.35E+00	5.67E+00	1.94E+03	9.97E+02	2.63E+00	2.34E+00
	Std	1.26E-01	1.43E-02	1.47E-01	6.55E+02	1.14E+02	1.38E-01	1.35E-03
F10	Best	2.34E+00	2.00E+01	2.77E+00	2.02E+01	2.03E+01	2.00E+01	2.00E+01
	Ave	1.94E+01	2.00E+01	3.25E+00	2.04E+01	2.03E+01	2.00E+01	2.00E+01
	Std	3.23E+00	3.00E-04	6.71E-01	1.81E-01	8.96E-02	1.76E-03	2.69E-02

Experiment 3: CEC 2022 Test Functions

In this section, we extend our evaluation to include the CEC 2022 test functions, which present another set of complex optimization challenges (see Table 11 for more details). These test functions are designed with various characteristics, including unimodal, basic, hybrid, and composition types, offering a diverse range of optimization landscapes for assessment. The CEC 2022 test functions serve as a rigorous benchmark for evaluating the performance of the WHOFWA algorithm in handling complex, high-dimensional optimization problems. For each of these CEC 2022 test functions, we will assess the performance of WHOFWA and compare it to other state-of-the-art optimization algorithms. The evaluation will be based on key performance metrics, including proximity to the optimum, early convergence, and hit rate (accuracy). These results will demonstrate the effectiveness and adaptability of WHOFWA in addressing a wide array of challenging optimization landscapes.

Table 10: Ranks of WHOFWA and other algorithms in CEC2019 test functions

		WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
Friedman Test	Mean Rank	2.39	2.78	3.72	6.11	5.78	4.17	3.06
	Rank	**1**	**2**	**4**	**7**	**6**	**5**	**3**

Table 11: The details of CEC 2022 test functions (For all test functions: D = 10 and Range = [-100, 100])

Function	Name	Optimum value	Type
F1	Shifted and full Rotated Zakharov Function	300	Unimodal
F2	Shifted and full Rotated Rosenbrock's Function	400	Basic
F3	Shifted and full Rotated Expanded Schaffer's *f*6 Function	600	
F4	Shifted and full Rotated Non-Continuous Rastrigin's Function	800	
F5	Shifted and full Rotated Levy Function	900	
F6	HF 1 (*N* = 3)	1800	Hybrid
F7	HF 2 (N = 6)	2000	
F8	HF 3 (*N* = 5)	2200	
F9	CF 1 (N = 5)	2300	Composition
F10	CF 2 (*N* = 4)	2400	
F11	CF 3 (N = 5)	2600	
F12	CF 4 (N = 6)	2700	

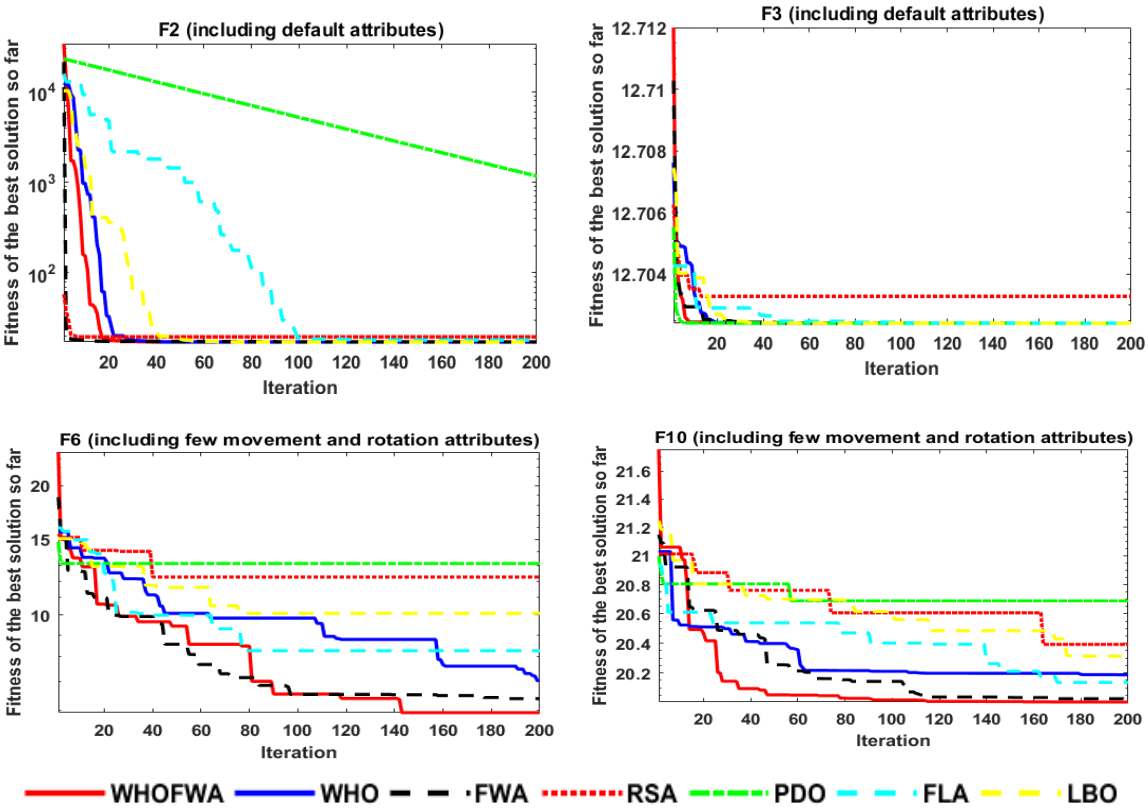


Fig. 2: Convergence curve of all algorithms on solving the CEC 2019 test functions.

Table 12 shows the comparative results for the CEC 2022 test functions, assessing the performance of WHOFWA alongside other optimization algorithms.

The evaluation is based on essential performance metrics, including the best solution achieved, the average performance, and the standard deviation across multiple runs. The outcomes demonstrate how WHOFWA adapts to a diverse set of complex optimization landscapes:

1. **F1 (Shifted and full Rotated Zakharov Function)**: WHOFWA obtains a competitive best solution of 3.07E+02, demonstrating its effectiveness in addressing this unimodal function. The algorithm achieves a low standard deviation, indicating its consistency. WHOFWA outperforms other algorithms such as FA, RSA, PDO, FWA, and LBO in terms of best solutions.
2. **F2 (Shifted and full Rotated Rosenbrock's Function)**: WHOFWA attains a best solution of 4.00E+02, demonstrating its capability in handling this basic optimization problem. The algorithm provides competitive results in terms of both best solutions and average performance. It outperforms some other algorithms, indicating its effectiveness in this context.
3. **F3 (Shifted and full Rotated Expanded Schaffer's f6 Function)**: WHOFWA reaches a best solution of 6.00E+02 for this function. It offers consistent performance with a minimal standard deviation, showcasing its stability and reliability. WHOFWA performs well in comparison to other algorithms, including FA, RSA, PDO, and FWA.
4. **F4 (Shifted and full Rotated Non-Continuous Rastrigin's Function)**: WHOFWA achieves a competitive best solution of 8.07E+02 for this challenging function. The algorithm's performance is notable, as it provides results with a low standard deviation. WHOFWA surpasses some other algorithms in terms of both best solutions and average performance, demonstrating its adaptability to complex landscapes.
5. **F5 (Shifted and full Rotated Levy Function)**: WHOFWA attains a best solution of 9.00E+02, showcasing its ability to handle this intricate optimization problem. The algorithm delivers competitive results in terms of best solutions and averages, demonstrating its effectiveness and stability in comparison to other algorithms.
6. **F6 (HF 1 - Hybrid Function)**: WHOFWA achieves a best solution of 1.86E+03 for this hybrid function. While the landscape is extremely complex, WHOFWA's performance is commendable. It outperforms other algorithms in terms of both best solutions and average performance, showcasing its adaptability to hybrid optimization problems.
7. **F7 (HF 2 - Hybrid Function)**: WHOFWA delivers a best solution of 2.00E+03 for this hybrid function. The algorithm's results are competitive and consistent, with a low standard deviation. WHOFWA performs well compared to other algorithms in terms of best solutions.
8. **F8 (HF 3 - Hybrid Function)**: WHOFWA reaches a best solution of 2.20E+03 for this hybrid function. The algorithm provides competitive results and stability, outperforming some other algorithms.
9. **F9 (CF 1 - Composition Function)**: WHOFWA achieves a best solution of 2.53E+03 for this composition function. It offers consistent performance, showcasing its adaptability to complex composition landscapes. WHOFWA outperforms some other algorithms in this context.
10. **F10 (CF 2 - Composition Function)**: WHOFWA attains a best solution of 2.40E+03 for this composition function. The algorithm provides competitive results, demonstrating its effectiveness in handling composition optimization problems.
11. **F11 (CF 3 - Composition Function)**: WHOFWA delivers a best solution of 2.60E+03 for this composition function. The algorithm's performance is commendable, outperforming some other algorithms in terms of both best solutions and averages.
12. **F12 (CF 4 - Composition Function)**: WHOFWA reaches a best solution of 2.86E+03 for this composition function. The algorithm's performance is consistent and competitive. It demonstrates its adaptability to composition optimization problems, surpassing some other algorithms in terms of best solutions and averages.

In sum, the results in Table 12 emphasize WHOFWA's robustness and adaptability, particularly in addressing complex, high-dimensional optimization landscapes. It consistently performs competitively or outperforms other state-of-the-art algorithms across a diverse range of test functions, showcasing its efficacy as a metaheuristic algorithm.

Table 13 presents the ranks of the compared optimization algorithms based on their performance across the CEC 2022 test functions.

The Friedman test assesses how these algorithms compare in terms of their overall performance, with a focus on the mean rank.

In sum, both WHOFWA and WHO demonstrate top-tier performance across the CEC 2022 test functions, achieving the lowest mean ranks. This suggests that WHOFWA is highly competitive and effective in comparison to other algorithms in these experiments.

Table 12: Comparative results for CEC 2022 test functions

Function	Metric	WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
F1	Best	3.07E+02	3.00E+02	1.35E+03	5.73E+03	1.16E+04	3.03E+02	3.00E+02
	Ave	3.86E+02	3.00E+02	1.60E+03	7.13E+03	1.53E+04	3.09E+02	3.00E+02
	Std	9.02E+01	8.04E-14	3.52E+02	1.97E+03	5.12E+03	7.73E+00	2.81E-05
F2	Best	4.00E+02	4.04E+02	4.01E+02	5.32E+02	7.75E+02	4.07E+02	4.00E+02
	Ave	4.05E+02	4.09E+02	4.03E+02	5.36E+02	8.31E+02	4.08E+02	4.02E+02
	Std	4.18E+00	4.81E+00	3.80E+00	4.70E+00	7.99E+01	9.79E-01	3.01E+00
F3	Best	6.00E+02	6.00E+02	6.00E+02	6.35E+02	6.40E+02	6.00E+02	6.00E+02
	Ave	6.00E+02	6.00E+02	6.01E+02	6.38E+02	6.42E+02	6.00E+02	6.00E+02
	Std	4.26E-02	2.53E-02	4.25E-01	4.26E+00	3.86E+00	3.65E-02	1.47E-04
F4	Best	8.07E+02	8.06E+02	8.30E+02	8.39E+02	8.34E+02	8.19E+02	8.12E+02
	Ave	8.28E+02	8.14E+02	8.30E+02	8.40E+02	8.36E+02	8.33E+02	8.17E+02
	Std	8.41E+00	7.60E+00	7.05E-01	7.46E-01	1.91E+00	1.28E+01	7.74E+00
F5	Best	9.00E+02	9.00E+02	9.27E+02	1.22E+03	1.23E+03	9.01E+02	9.00E+02
	Ave	9.61E+02	9.00E+02	9.47E+02	1.26E+03	1.38E+03	9.54E+02	9.00E+02
	Std	9.30E+01	2.41E-01	2.84E+01	5.42E+01	2.13E+02	8.56E+01	1.96E-05
F6	Best	1.86E+03	1.92E+03	2.06E+03	4.43E+07	1.20E+07	1.98E+03	2.22E+03
	Ave	3.88E+03	4.01E+03	2.77E+03	4.78E+07	5.18E+07	3.46E+03	4.92E+03
	Std	1.89E+03	3.41E+03	1.02E+03	4.97E+06	5.63E+07	1.77E+03	3.82E+03
F7	Best	2.00E+03	2.00E+03	2.00E+03	2.10E+03	2.08E+03	2.02E+03	2.02E+03
	Ave	2.01E+03	2.02E+03	2.01E+03	2.10E+03	2.08E+03	2.02E+03	2.02E+03
	Std	9.17E+00	1.22E+01	1.21E+01	7.51E-01	2.47E+00	8.47E-02	3.57E-02
F8	Best	2.20E+03	2.20E+03	2.21E+03	2.24E+03	2.23E+03	2.22E+03	2.20E+03
	Ave	2.22E+03	2.21E+03	2.21E+03	2.25E+03	2.24E+03	2.22E+03	2.21E+03
	Std	5.82E+00	1.15E+01	1.06E+01	1.13E+01	4.55E+00	3.19E-01	1.45E+01
F9	Best	2.53E+03	2.53E+03	2.53E+03	2.70E+03	2.73E+03	2.53E+03	2.53E+03
	Ave	2.54E+03	2.53E+03	2.53E+03	2.70E+03	2.76E+03	2.53E+03	2.53E+03
	Std	4.48E+01	0.00E+00	5.92E-01	2.18E+00	3.58E+01	1.18E-03	3.85E-04
F10	Best	2.40E+03	2.50E+03	2.50E+03	2.69E+03	2.52E+03	2.61E+03	2.50E+03
	Ave	2.54E+03	2.58E+03	2.56E+03	2.70E+03	2.52E+03	2.62E+03	2.50E+03
	Std	6.71E+01	6.63E+01	8.39E+01	3.86E+00	1.59E+00	4.05E+00	5.09E-02
F11	Best	2.60E+03	2.60E+03	2.61E+03	2.90E+03	2.84E+03	2.60E+03	2.90E+03
	Ave	2.68E+03	2.60E+03	2.76E+03	3.40E+03	3.03E+03	2.78E+03	2.95E+03
	Std	1.18E+02	3.22E-13	2.18E+02	7.13E+02	2.77E+02	2.02E+02	7.07E+01
F12	Best	2.86E+03	2.86E+03	2.87E+03	2.89E+03	2.88E+03	2.86E+03	2.87E+03
	Ave	2.87E+03	2.87E+03	2.91E+03	2.89E+03	2.88E+03	2.86E+03	2.87E+03
	Std	7.32E+00	2.17E+00	5.31E+01	5.38E+00	1.75E-01	2.44E+00	5.52E-02

Table 13: Ranks of WHOFWA and other algorithms in CEC2022 test functions

		WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
Friedman Test	Mean Rank	2.25	2.25	3.79	6.54	6.25	3.67	3.25
	Rank	1	1	4	6	5	3	2

While other algorithms like FA, RSA, PDO, FLA, and LBO perform reasonably well, WHOFWA and WHO emerge as strong contenders for solving the complex optimization problems presented in the CEC 2022 test functions.

Experiment 4: Three Constrained Engineering Design Problems

In this study, the effectiveness of WHOFWA is assessed through the examination of three constrained engineering design problems, with a subsequent comparison of the obtained results against various algorithms documented in the existing literature.

Speed reducer design: The objective of this problem is to optimize the weight of a speed reducer through design. Considering the configuration of the speed reducer, seven decision variables must be defined to meet eleven specified constraints. Table 14 shows the results of WHOFWA and some good algorithms in the literature such as SC [56], PSO-DE [57], DELC [58], DEDS [59], HEAA [60], MDE [61], and ABC [62] on this problem. This table affirms that WHOFWA, in conjunction with DELC and DEDS, achieves the top rank in converging to the optimal solution among the presented approaches.

Pressure vessel: The objective of this challenge is to minimize the fabrication cost associated with a pressure vessel. The problem involves four design variables and is subject to four specified constraints. Table 15 shows the results of WHOFWA and some good algorithms in the literature such as GA2 [63], GA3 [64], QPSO [65], and PSO [57] on this problem. This table confirms that WHOFWA holds the top rank in achieving a solution closest to the optimum.

Tension/compression spring design: The aim of this problem is to engineer a tension/compression spring with minimal weight. The problem encompasses four design variables and imposes four constraints related to minimum deflection, shear stress, and surge frequency. Table 16 shows the results of WHOFWA and some good algorithms in the literature such as GA2, GA3, CAEP [66], CSPSO [67], HPSO [68], DE [69], SC, and ABC on this problem. Due to this table, WHOFWA obtains the first rank on finding the closest solution to an optimum.

Discussion

Table 17 summarizes the overall performance of WHOFWA and other optimization algorithms across all test functions used in the study. The table provides the mean ranks of the algorithms based on their performance, and the ranks for each algorithm across all experiments. According to this table, Algorithm WHOFWA obtains the first rank among the considered algorithms. Fig. 3 shows the convergence curve of all algorithms on solving the CEC 2022 test functions. Fig. 4 presents the hit rate of WHOFWA and other algorithms across all test functions. The hit rate measures the accuracy of an algorithm in finding the optimal solutions within a certain range.

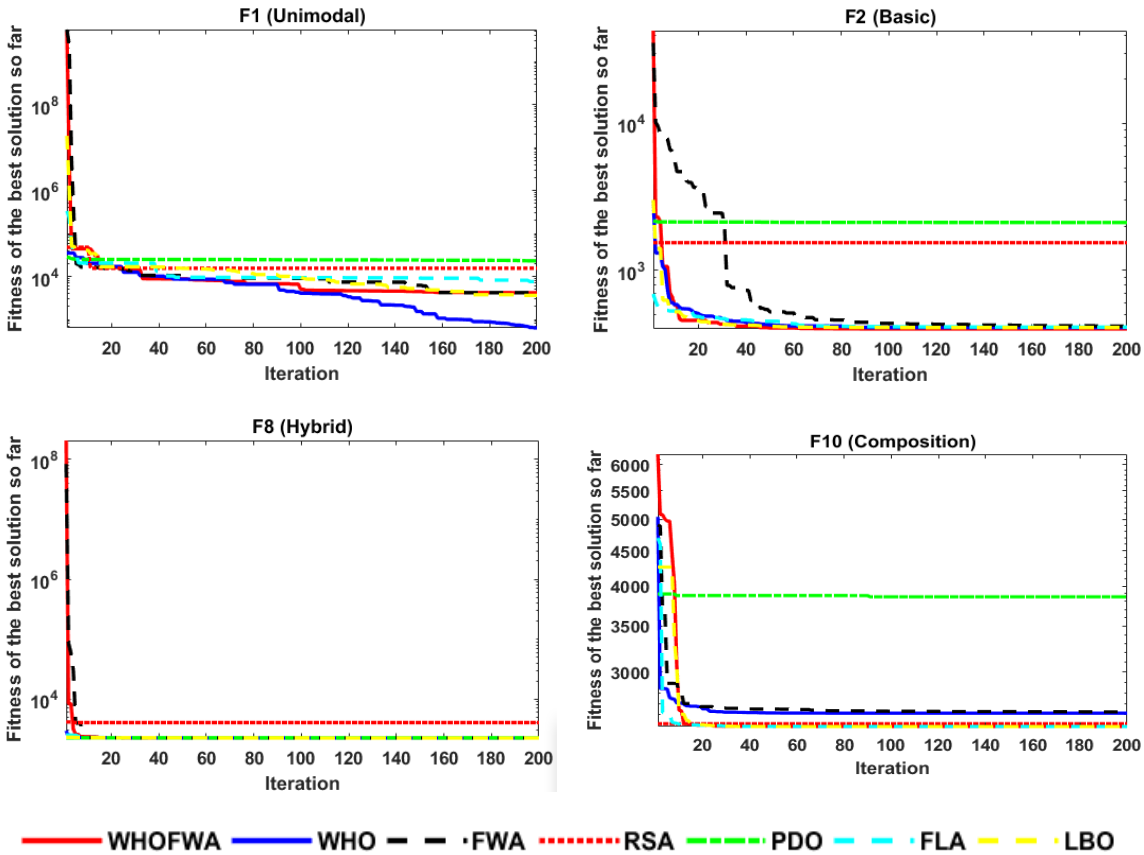


Fig. 3: Convergence curve of all algorithms on solving the CEC 2022 test functions.

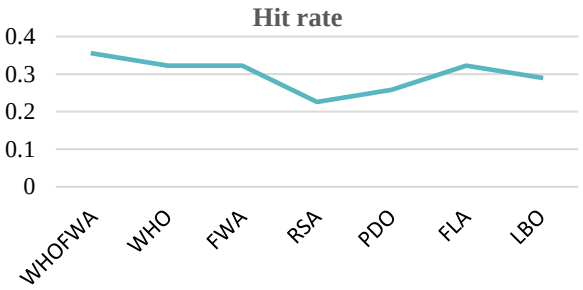


Fig. 4: Hit rate of WHOFWA and others for all test functions.

Table 14: Results of different methods on the speed reducer problem

Method	Best	Worst	Ave	Std	Rank (Ave)
WHOFWA	2994.471066	2994.471066	2994.471066	0.000000	**1**
SC	2994.744241	3009.964736	3001.758264	4.0	**6**
PSO-DE	2996.348167	2996.348204	2996.348174	6.4E-06	**3**
DELC	2994.471066	2994.471066	2994.471066	1.9E-12	**1**
DEDS	2994.471066	2994.471066	2994.471066	3.6E-12	**1**
HEAA	2994.499107	2994.752311	2994.613368	7.0E-02	**2**
MDE	2996.356689	NA	2996.367220	8.2E-03	**4**
ABC	2997.058000	NA	2997.058000	0.0	**5**

Table 15: Results of different methods on the pressure vessel problem

Method	Best	Worst	Ave	Std	Rank (Ave)
WHOFWA	6051.3169	6308.715	6179.1843	124.8362	**1**
GA2	6288.7445	6308.4970	6293.8432	7.4133	**3**
GA3	6059.9463	6469.3220	6177.2533	130.9297	**2**
QPSO	6059.7209	8017.2816	6440.3786	6059.7209	**4**
PSO	6693.7212	14076.3240	8756.6803	1492.5670	**5**

Table 16: Results of different methods on the tension/compression spring design problem

Method	Best	Worst	Ave	Std	Rank (Ave)
WHOFWA	0.012629	0.012682	0.012630	0.0000315	**1**
GA2	0.012704	0.012822	0.012769	3.94E-05	**7**
GA3	0.012681	0.012973	0.012742	5.90E-05	**6**
CAEP	0.012721	0.015116	0.013568	8.42E-04	**9**
CPSO	0.012674	0.012924	0.012730	5.20E-04	**5**
HPSO	0.012665	0.012719	0.012707	1.58E-05	**3**
DE	0.012670	0.012790	0.012703	2.7E-05	**2**
SC	0.012669	0.016717	0.012922	1.2E-08	**8**
ABC	0.012665	NA	0.012709	0.012813	**4**

Additionally, alongside the Friedman test, we utilize the Wilcoxon signed-rank test to examine the outcomes. This non-parametric statistical hypothesis test, designed for the comparison of two samples [70], defines the instances where Algorithm *X* surpasses, falls short, or achieves comparable performance to Algorithm *Y*. To convey these findings, three output test statistics R^-, R^+, and $R^=$ are presented. The Wilcoxon signed-rank test outcomes for the pairwise evaluation of WHOFWA versus WHO, FWA, RSA, PDO, FLA, and LBO are illustrated in Fig. 5. Notably, in all these pair comparisons, the values of R^- exceed those of R^+, indicating that WHOFWA demonstrates superior effectiveness compared to all other algorithms in the depicted figure. To evaluate and compare the effectiveness of WHOFWA with others in solving test functions with very high dimensions, we consider seven test functions belonging to CEC 2005, which have the dimension of 100.

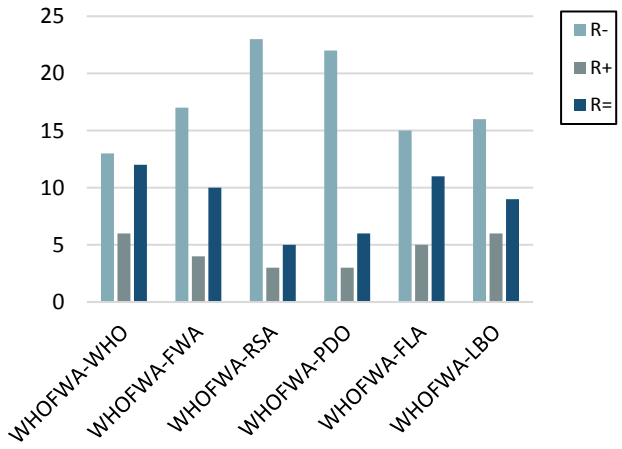


Fig. 5: Results of the Wilcoxon signed-rank test for all test functions.

Table 18 displays the details of these test functions such as Name, Characteristics, range, and optimum value. In Table 19, the results of executing WHOFWA and others on these test functions are shown. The statistical analysis of the results using the Friedman test provides the mean ranks of the algorithms (Table 20). According to these analysis results, Algorithm WHOFWA ranks first among the considered algorithms.

Table 17: Overall ranks of WHOFWA and other algorithms in all test functions

		WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
Friedman Test	Mean Rank	2.73	3.29	3.60	5.66	5.26	3.73	3.74
	Rank	**1**	**2**	**3**	**7**	**6**	**4**	**5**

Table 18: The details of the CEC 2005 test functions (For all test functions: D = 100)

Problem	Name	Characteristics	Range	Optimum value
F1	Ellipsoid	Unimodal	[-5.12, 5.12]	0
F2	Rosenbrock	Multimodal with narrow valley	[-2.048, 2.048]	0
F3	Ackley	Multimodal	[-32.768, 32.768]	0
F4	Griewank	Multimodal	[-600, 600]	0
F5	Shifted Rotated Rastrigin	Very complicated multimodal	[-5, 5]	-330
F6	Rotated hybrid composition function	Very complicated multimodal	[-5, 5]	120
F7	Rotated hybrid composition function	Very complicated multimodal	[-5, 5]	10

Conclusion

In this paper, we introduced WHOFWA, a novel hybrid metaheuristic algorithm that amalgamates the strengths of the Wild Horse Optimizer (WHO) and the Fireworks Algorithm (FA) to create a powerful and versatile optimization tool. WHOFWA's development stemmed from the need for an efficient, robust, and adaptive optimization algorithm capable of addressing a wide range of optimization problems. Our comprehensive experimental study evaluated WHOFWA's performance across various test functions, and the results confirm its effectiveness and robustness.

The primary findings of this research can be summarized as follows:

1. **Superior Performance**: WHOFWA consistently outperformed several state-of-the-art optimization algorithms in various experiments, ranking at the top in most cases. Its ability to find optimal solutions, both for unimodal and multimodal functions, demonstrates its versatility and problem-solving capability.
2. **Competitive Comparison**: WHOFWA's performance was compared to other popular algorithms, including WHO, FA, RSA, PDO, FLA, and LBO. It consistently achieved the highest average rankings, indicating its superiority across different test functions.
3. **Robustness and Adaptability**: WHOFWA's robustness and adaptability were evident in its consistent performance across diverse test functions. It showcases its ability to handle various optimization challenges, making it a valuable addition to the field of metaheuristic algorithms.
4. **Potential for Real-World Applications**: The demonstrated efficacy of WHOFWA holds significant promise for real-world applications, where optimization plays a crucial role in problem-solving. Whether in engineering, finance, logistics, or other domains, WHOFWA has the potential to streamline decision-making processes.
5. **Versatility and Flexibility**: WHOFWA's hybrid nature, integrating the Wild Horse Optimizer and Fireworks Algorithm, contributes to its flexibility and versatility. This adaptability allows it to address a wide range of optimization problems with exceptional precision.

In sum, WHOFWA stands out as a cutting-edge hybrid metaheuristic algorithm with the potential to revolutionize optimization methodologies. Its consistent top-tier performance across diverse test functions and its adaptability to real-world problem-solving scenarios make it a valuable asset for researchers, practitioners, and industries seeking efficient optimization tools. The success of WHOFWA in this study invites further exploration and application in various domains, with the hope of advancing the state-of-the-art in optimization algorithms and positively impacting problem-solving on a global scale. Additionally, developing a multi-objective version or exploring the binary version of WHOFWA can be the future works.

Table 19: Comparative results for CEC 2005 test functions

Function	Metric	WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
F1	Best	0.00E+00	4.80E-93	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.53E-23
	Ave	0.00E+00	3.43E-86	0.00E+00	0.00E+00	0.00E+00	0.00E+00	5.73E-22
	Std	0.00E+00	5.94E-86	0.00E+00	0.00E+00	0.00E+00	0.00E+00	7.78E-22
F2	Best	8.73E-02	9.54E+01	2.90E+01	1.20E-04	9.90E+01	2.90E+01	9.61E+01
	Ave	1.36E-01	1.28E+02	3.12E+01	8.45E-04	9.90E+01	3.12E+01	9.68E+01
	Std	5.61E-02	3.98E+01	3.59E+00	6.31E-04	7.04E-03	3.59E+00	9.77E-01
F3	Best	8.88E-17	8.88E-16	4.44E-15	8.88E-16	8.88E-16	4.44E-15	1.10E-05
	Ave	8.88E-17	3.26E-15	4.44E-15	8.88E-16	8.88E-16	4.44E-15	3.25E-05
	Std	0.00E+00	2.05E-15	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.00E-05
F4	Best	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.36E+00
	Ave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.53E+00
	Std	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.48E-01
F5	Best	1.33E+03	5.31E+02	7.29E+02	1.85E+03	1.69E+03	7.29E+02	6.66E+02
	Ave	1.37E+03	6.91E+02	7.73E+02	1.89E+03	1.80E+03	7.73E+02	6.98E+02
	Std	6.81E+01	2.32E+02	4.01E+01	7.26E+01	1.44E+02	4.01E+01	5.11E+01
F6	Best	6.05E+02	3.47E+02	3.84E+02	1.18E+03	8.85E+02	3.84E+02	3.88E+02
	Ave	6.58E+02	3.76E+02	4.23E+02	1.21E+03	9.65E+02	4.23E+02	4.28E+02
	Std	9.11E+01	2.73E+01	3.45E+01	2.96E+01	7.01E+01	3.45E+01	6.06E+01
F7	Best	9.10E+02	1.24E+03	1.08E+03	9.10E+02	9.10E+02	1.08E+03	1.07E+03
	Ave	9.10E+02	1.25E+03	1.09E+03	9.10E+02	9.10E+02	1.09E+03	1.11E+03
	Std	0.00E+00	1.24E+01	7.14E+00	0.00E+00	0.00E+00	7.14E+00	5.82E+01

Table 20: Ranks of WHOFWA and other algorithms in CEC 2005 test functions

		WHOFWA	WHO	FWA	RSA	PDO	FLA	LBO
Friedman Test	Mean Rank	3.07	3.79	3.86	3.79	4.36	3.86	5.29
	Rank	**1**	**2**	**3**	**2**	**4**	**3**	**5**

Author Contributions

Both of the paper authors, i.e., A. Rouhi and E. Pira designed the metaheuristic algorithm, conducted the experiments, interpreted the results and wrote the manuscript.

Acknowledgment

We would like to express our sincere gratitude to the editor and anonymous reviewers for their valuable time, insightful feedback, and constructive comments, which greatly contributed to the improvement and quality of this paper.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

ABC	Artificial Bee Colony
AIS	Artificial Immune System
DA	Dragonfly Algorithm
EAs	Evolutionary Algorithms
FPA	Flower Pollination Algorithm
FLA	Fick's Law Optimization
FWA	Fireworks Algorithm
GA	Genetic Algorithm
GOA	Gazelle Optimization Algorithm
GWO	Grey Wolf Optimizer
HHO	Harris's Hawk Optimization
HB	Human Behaviors
KHA	Krill Herd Algorithm
LBO	Ladybug Beetle Optimization
LCA	League Championship Algorithm
MFO	Moth–Flame Optimization
NFL	No Free Lunch theorem
NP	Natural Phenomenon
PDO	Prairie Dog Optimization
PSO	Particle Swarm Optimization
RSA	Reptile Search Algorithm
SA	Simulated Annealing
SI	Swarm Intelligence
SSA	Social Spider Algorithm
WHO	Wild Horse Optimizer
WHOFWA	Wild Horse and Fireworks Algorithm

References

[1] X. S. Yang, "Nature-inspired optimization algorithms: Challenges and open problems," J. Comput. Sci., 46: 101104, 2020.

[2] F. Marini, B. Walczak, "Particle swarm optimization (PSO). A tutorial," Chemom. Intell. Lab. Syst., 149: 153-165, 2015.

[3] U. Yüzgeç, M. Kusoglu, "Multi-objective harris hawks optimizer for multiobjective optimization problems," BSEU J. Eng. Res. Technol., 1(1): 31-41, 2020.

[4] J. O. Agushaka, A. E. Ezugwu, L. Abualigah, "Gazelle optimization algorithm: A novel nature-inspired metaheuristic optimizer," Neural Comput. Appl., 35(5): 4099-4131, 2023.

[5] S. Mirjalili, S. M. Mirjalili, A. Lewis, "Grey wolf optimizer," Adv. Eng. Software, 69: 46-61, 2014.

[6] S. Mirjalili, A. Lewis, "The whale optimization algorithm," Adv. Eng. Software, 95: 51-67, 2016.

[7] M. Shehab, L. Abualigah, H. Al Hamad, H. Alabool, M. Alshinwan, A. M. Khasawneh, "Moth–flame optimization algorithm: Variants and applications," Neural Comput. Appl., 32: 9859-9884, 2020.

[8] D. H. Wolpert, W. G. Macready, "No free lunch theorems for optimization," IEEE Trans. Evol. Comput., 1(1): 67-82, 1997.

[9] I. Naruei, F. Keynia, "Wild horse optimizer: A new meta-heuristic algorithm for solving engineering optimization problems," Eng. Comput., 38(Suppl 4): 3025-3056, 2022.

[10] Y. Tan Y. Zhu, "Fireworks algorithm for optimization," in Proc. Advances in Swarm Intelligence: First International Conference (ICSI): Part I 1: 355-364, 2010.

[11] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, A. H. Gandomi, "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer," Expert Syst. Appl., 191: 116158, 2022.

[12] A. E. Ezugwu, J. O. Agushaka, L. Abualigah, S. Mirjalili, A. H. Gandomi, "Prairie dog optimization algorithm," Neural Comput. Appl., 34(22): 20017-20065, 2022.

[13] F. A. Hashim, R. R. Mostafa, A. G. Hussien, S. Mirjalili, K. M. Sallam, "Fick's Law Algorithm: A physical law-based algorithm for numerical optimization," Knowledge-Based Syst., 260: 110146, 2023.

[14] S. Safiri, A. Nikoofard, "Ladybug beetle optimization algorithm: Application for real-world problems," J. Supercomput., 79(3): 3511-3560, 2023.

[15] M. Khadem, A. Toloie Eshlaghy, K. Fathi, "Nature-inspired metaheuristic algorithms: Literature review and presenting a novel classification," J. Appl. Res. Ind. Eng., 10(2): 286-339, 2023.

[16] F. Salami, A. Bozorgi-Amiri, R. Tavakkoli-Moghaddam, "How metaheuristic algorithms can help in feature selection for Alzheimer's diagnosis," Int. J. Res. Ind. Eng., 12(2): 197-204, 2023.

[17] S. E. Najafi, S. Salahshour, B. Rahmani Parchikolaei, "Optimizing supplier selection for a construction project by a cash-flow approach using a hybrid metaheuristic algorithm," Big Data Comput. Visions, 2(2): 69-79, 2022.

[18] P. Bahrampour, S. E. Najafi, A. Edalatpanah, "Designing a scenario-based fuzzy model for sustainable closed-loop supply chain network considering statistical reliability: A new hybrid metaheuristic algorithm," Complexity, 2023: 1-24, 2023.

[19] H. R. Yousefzadeh S. M. Masumi, "Teachers timetabling in Torbat-E-Jam schools using constructive genetic algorithm," Mod. Res. Perform. Eval., 1(1): 42-48, 2022.

[20] K. Rajwar, K. Deep, S. Das, "An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges," Artif. Intell. Rev.: 1-71, 2023.

[21] M. Abdel-Basset, L. Abdel-Fatah, A. K. Sangaiah, "Metaheuristic algorithms: A comprehensive review," Comput. Intell. Multimedia Big Data Cloud Eng. Appl., 2018: 185-231, 2018.

[22] I. Boussaïd, J. Lepagnot, P. Siarry, "A survey on optimization metaheuristics," Inf. Sci., 237: 82-117, 2013.

[23] A. Rouhi, E. Pira, "A surrogate model-based aquila optimizer for solving high-dimensional computationally expensive problems," J. Comput. Securi., 11(1): 1-18, 2024.

[24] B. Alhijawi, A. Awajan, "Genetic algorithms: Theory, genetic operators, solutions, and applications," Evol. Intell., 1-12, 2023.

[25] L. Vanneschi, S. Silva, "Genetic Programming," in Lectures on Intelligent Systems: Springer, pp. 205-257, 2023.

[26] D. Delahaye, S. Chaimatanan, M. Mongeau, "Simulated annealing: From basics to applications," Handbook of metaheuristics, 1-35, 2019.

[27] M. Azizi, U. Aickelin, H. A. Khorshidi, M. Baghalzadeh Shishehgarkhaneh, "Energy valley optimizer: a novel metaheuristic algorithm for global and engineering optimization," Sci. Rep., 13(1): 226, 2023.

[28] M. Abdel-Basset, R. Mohamed, M. Jameel, M. Abouhawwash, "Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems," Knowledge-Based Syst., 262: 110248, 2023.

[29] M. Kaveh, M. S. Mesgari, B. Saeidian, "Orchard Algorithm (OA): A new meta-heuristic algorithm for solving discrete and continuous optimization problems," Math. Comput. Simul., 208: 95-135, 2023.

[30] P. D. Kusuma, F. C. Hasibuan, "Swarm magnetic optimizer: A new optimizer that adopts magnetic behaviour," Int. J. Intell. Eng. Syst., 16(4), 2023.

[31] S. Pawar, M. K. Ahirwal, "A new fission fusion behavior-based Rao algorithm (FFBBRA) for solving optimization problems," Evol. Intell., 16(4): 1309-1323, 2023.

[32] M. Dorigo, M. Birattari, T. Stutzle, "Ant colony optimization," IEEE Comput. Intell. Mag., 1(4): 28-39, 2006.

[33] J. Nayak, H. Swapnarekha, B. Naik, G. Dhiman, S. Vimal, "25 years of particle swarm optimization: Flourishing voyage of two decades," Arch. Comput. Meth. Eng., 30(3): 1663-1725, 2023.

[34] M. Azizi, S. Talatahari, A. H. Gandomi, "Fire hawk optimizer: A novel metaheuristic algorithm," Artif. Intell. Rev., 56(1): 287-363, 2023.

[35] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-Qaness, A. H. Gandomi, "Aquila optimizer: a novel meta-heuristic optimization algorithm," Comput. Ind. Eng., 157: 107250, 2021.

[36] J. Xue, B. Shen, "Dung beetle optimizer: A new meta-heuristic algorithm for global optimization," J. Supercomput., 79(7): 7305-7336, 2023.

[37] J. O. Agushaka, A. E. Ezugwu, L. Abualigah, "Dwarf mongoose optimization algorithm," Comput. Meth. Appl. Mech. Eng., 391: 114570, 2022.

[38] S. Mirjalili, S. Mirjalili, "Genetic algorithm," Evol. Algorithms Neural Networks: 43-55, 2019.

[39] D. Bertsimas, J. Tsitsiklis, "Simulated annealing," Stat. Sci., 8(1): 10-15, 1993.

[40] D. Karaboga, B. Gorkemli, C. Ozturk, N. Karaboga, "A comprehensive survey: Artificial bee colony (ABC) algorithm and applications," Artif. Intell. Rev., 42: 21-57, 2014.

[41] X. S. Yang, S. Deb, "Cuckoo search: Recent advances and applications," Neural Computi. Appl., 24: 169-174, 2014.

[42] Z. W. Geem, J. H. Kim, G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," Simulation, 76(2): 60-68, 2001.

[43] X. S. Yang, X. He, "Bat algorithm: Literature review and applications," Int. J. Bio-inspired Comput., 5(3): 141-149, 2013.

[44] D. Dasgupta, S. Yu, F. Nino, "Recent advances in artificial immune systems: models and applications," Appl. Soft Comput., 11(2): 1574-1587, 2011.

[45] S. Mirjalili, "Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm," Knowledge-based syst., 89: 228-249, 2015.

[46] M. Abdel-Basset, L. A. Shawky, "Flower pollination algorithm: a comprehensive review," Artif. Intell. Rev., 52: 2533-2557, 2019.

[47] A. L. A. Bolaji, M. A. Al-Betar, M. A. Awadallah, A. T. Khader, L. M. Abualigah, "A comprehensive review: Krill Herd algorithm (KH) and its applications," Appl. Soft Comput., 49: 437-446, 2016.

[48] Y. Meraihi, A. Ramdane-Cherif, D. Acheli, M. Mahseur, "Dragonfly algorithm: A comprehensive review and applications," Neural Comput. Appl., 32: 16625-16646, 2020.

[49] A. H. Kashan, "League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships," Appl. Soft Comput., 16: 171-200, 2014.

[50] J. James, V. O. Li, "A social spider algorithm for global optimization," Appl. soft Comput., 30: 614-627, 2015.

[51] F. Shahabi, F. Poorahangaryan, S. Edalatpanah, H. Beheshti, "A multilevel image thresholding approach based on crow search algorithm and Otsu method," Int. J. Comput. Intell. Appl., 19(02): 2050015, 2020.

[52] S. Zhao, T. Zhang, S. Ma, M. Chen, "Dandelion optimizer: A nature-inspired metaheuristic algorithm for engineering applications," Eng. Appl. Artif. Intell., 114: 105075, 2022.

[53] M. H. Sulaiman, Z. Mustaffa, M. M. Saari, H. Daniyal, "Barnacles mating optimizer: A new bio-inspired algorithm for solving engineering optimization problems," Eng. Appl. Artif. Intell., 87: 103330, 2020.

[54] V. Beiranvand, W. Hare, Y. Lucet, "Best practices for comparing optimization algorithms," Optim. Eng., 18: 815-848, 2017.

[55] R. L. Rardin, R. Uzsoy, "Experimental evaluation of heuristic optimization algorithms: A tutorial," J. Heuristics, 7: 261-304, 2001.

[56] T. Ray, K. M. Liew, "Society and civilization: an optimization algorithm based on the simulation of social behavior," IEEE Trans. Evol. Comput., 7(4): 386-396, 2003.

[57] H. Liu, Z. Cai, Y. Wang, "Hybridizing particle swarm optimization with differential evolution for constrained numerical and engineering optimization," Appl. Soft Comput., 10(2): 629-640, 2010.

[58] L. Wang, L.-p. Li, "An effective differential evolution with level comparison for constrained engineering design," Struct. Multidiscip. Optim., 41(6): 947-963, 2010.

[59] M. Zhang, W. Luo, X. Wang, "Differential evolution with dynamic stochastic selection for constrained optimization," Inf. Sci., 178(15): 3043-3074, 2008.

[60] Y. Wang, Z. Cai, Y. Zhou, Z. Fan, "Constrained optimization based on hybrid evolutionary algorithm and adaptive constraint-handling technique," Struct. Multidiscip. Optim., 37(4): 395-413, 2009.

[61] E. Mezura-Montes, C. C. Coello, J. Velázquez-Reyes, "Increasing successful offspring and diversity in differential evolution for engineering design," in Proc. the seventh International Conference on Adaptive Computing in Design and Manufacture (ACDM): 131-139, 2006.

[62] D. Karaboga, B. Basturk, "Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems," in Proc. International fuzzy systems association world congress: 789-798, 2007.

[63] C. A. C. Coello, "Use of a self-adaptive penalty approach for engineering optimization problems," Comput. Ind., 41(2): 113-127, 2000.

[64] C. A. C. Coello, E. M. Montes, "Constraint-handling in genetic algorithms through the use of dominance-based tournament selection," Adv. Eng. Inf., 16(3): 193-203, 2002.

[65] L. dos Santos Coelho, "Gaussian quantum-behaved particle swarm optimization approaches for constrained engineering design problems," Expert Syst. Appl., 37(2): 1676-1683, 2010.

[66] C. A. Coello Coello, R. L. Becerra, "Efficient evolutionary optimization through the use of a cultural algorithm," Eng. Optim., 36(2): 219-236, 2004.

[67] Q. He, L. Wang, "An effective co-evolutionary particle swarm optimization for constrained engineering design problems," Eng. Appl. Artif. Intell., 20(1): 89-99, 2007.

[68] Q. He, L. Wang, "A hybrid particle swarm optimization with a feasibility-based rule for constrained optimization," Appl. Math. Comput., 186(2): 1407-1422, 2007.

[69] J. Lampinen, "A constraint handling approach for the differential evolution algorithm," in Proc. the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600), 2: 1468-1473, 2002.

[70] R. F. Woolson, "Wilcoxon signed-rank test," Wiley encyclopedia of clinical trials: 1-3, 2007.

Biographies

Alireza Rouhi received his B.Sc. at Kharazmi University of Tehran in September, 2000; M.Sc. at Sharif University of Technology in June, 2004; and Ph.D. at University of Isfahan in September 2017, all in Software Engineering field. He rewarded as outstanding researcher of Ph.D. students at Faculty of Computer Engineering, University of Isfahan in 2017. Currently, he is a lecturer at Azarbaijan Shahid Madani University, Tabriz, Iran. He is interested in Software Engineering in general and Formal Specification, Model Transformation, Metaheuristics, and Social Networks in particular.

- Email: rouhi@azaruniv.ac.ir
- ORCID: 0000-0003-1494-3467
- Web of Science Researcher ID: L-2209-2018
- Scopus Author ID: 57189992181
- Homepage: http://pajouhesh.azaruniv.ac.ir/_Pages/ResearcherEn.aspx?ID=5384

Einollah Pira received his B.Sc. degree in Computer Engineering (software) from the University of Kharazmi, Tehran, Iran [1996–2000], the M.Sc. degree in Computer Engineering (software) from the Sharif University of Technology, Tehran, Iran [2000–2002], and Ph.D degree in Computer Engineering (software) from Arak University, Iran [2013-2017]. Currently, he is an Assistant Professor with Department of Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran. His research interests include model checking, formal methods, software testing, evolutionary computation, and machine learning.

- Email: pira@azaruniv.ac.ir
- ORCID: 0000-0001-9010-6113
- Web of Science Researcher ID: NA
- Scopus Author ID 55941352000
- Homepage: http://pajouhesh.azaruniv.ac.ir/_Pages/ResearcherEn.aspx?ID=6617

How to cite this paper:
A. Rouhi, E. Pira, "WHOFWA: An effective hybrid metaheuristic algorithm based on wild horse optimizer and fireworks algorithm," J. Electr. Comput. Eng. Innovations, 12(2): 319-342, 2024.

DOI: 10.22061/jecei.2024.10422.699

URL: https://jecei.sru.ac.ir/article_2064.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Betavoltaic Battery using Platinum/Porous ZnO Schottky Junction

A. Ebadiyan [1], *A. Shokri* [1], *M. Amirmazlaghani* [1,*], *N. Darestani Farahani* [2]

[1] *Nanoelectronics Lab (NEL), Shahid Rajaee Teacher Training University, Tehran, Iran.*

[2] *Plasma and Nuclear Fusion Research School, Nuclear Science and Technology Research Institute, Tehran, Iran.*

Article Info

Article History:
Received 26 November 2023
Reviewed 17 January 2024
Revised 16 February 2024
Accepted 21 February 2024

Keywords:
Betavoltaic cell
Zinc oxide (ZnO) semiconductor
Porous structure
Schottky Junction
Energy conversion efficiency
TCAD-SILVACO 3D simulator

*Corresponding Author's Email Address:
m.mazlaghani@sru.ac.ir

Abstract

Background and Objectives: Semiconductor junction-based radioisotope detectors are commonly used in radioisotope batteries due to their small size and excellent performance. This study aims to design a betavoltaic battery based on a metal-porous semiconductor Schottky structure, comprising an N-type zinc oxide (ZnO) semiconductor and platinum (Pt) metal.

Methods: we utilized the TCAD-SILVACO 3D simulator to simulate the device, and a C-Interpreter code was applied to simulate the beta particle source, which was an electron beam with an average energy equivalent to 63Ni beta particles. The short circuit current, open-circuit voltage, fill factor (FF), and efficiency of the designed structure were calculated through simulation. Additionally, we discussed the theoretical justification based on the energy band structure.

Results: The energy conversion efficiency of the proposed structure was calculated to be 11.37% when bulk ZnO was utilized in the Schottky junction. However, by creating pores and increasing the effective junction area, a conversion efficiency of 35.5% was achieved. The proposed structure exhibited a short-circuit current, open-circuit voltage, and fill factor (FF) of 37.5 nA, 1.237 V, and 76.5%, respectively.

Conclusion: This study explored a betavoltaic device with a porous structure based on a Schottky junction between Pt and ZnO semiconductor. The creation of pores increased the contact surface area and effectively trapped beta beams, resulting in improved performance metrics such as efficiency, short circuit current, and open-circuit voltage.

Introduction

There are several methods for converting the energy of decayed particles into electrical energy. Using semiconductor devices to convert the energy of decayed beta particles into electrical energy is among the practical methods in this field.

Due to the rapid progress in microelectromechanical systems (MEMS) and nanoelectronics, there is a great demand for nuclear batteries [1]. Popularity of these batteries is rooted in their commercial, military, and medical applications, which are viable alternatives for chemical batteries due to their longevity, especially in remote applications. nuclear batteries, consisting of a radioactive source and a semiconductor junction or photonic crystals, have received great attentions in many works due to their stable output and longevity [2]-[7]. The PN structure is a widely used semiconductor structure that has a depletion region in the junction section. This region, with an appropriate width, absorbs Beta particles and produces electron-hole pairs. To improve the efficiency and output power of betavoltaic cells, various structures such as p-n, p-i-n, and Schottky have been studied and analyzed in references [8]-[13]. As the depletion region increases, the interaction of beta particles with the semiconductor material increases, resulting in the production of more electron-hole pairs.

The PIN structure, with a wider depletion zone, increases the probability of beta absorption [14]-[19]. The Schottky structure junction is created between the semiconductor and the metal near the surface. The Schottky diode has higher radiation resistance than that of the p–n junction and is a promising converter of the betavoltaic battery [20]-[23]. When determining the structure, it is important to consider the diffusion length of beta particles, which varies depending on the source type, activity, half-life, and particle type such as beta or alpha. Beta particles cause less damage to the cell structure and are commonly used in betavoltaic cells. Promethium-147 (147PM) is used as a beta source for designing betavoltaic batteries in references [24], [26]. Rosenkranz et al. presented the Clinical experience with nuclear-powered pacemakers (Promethium-147) in references [25]. The radioisotope 35S is used for designing and optimizing Si-35S betavoltaic liquid nuclear batteries [27]-[29]. Tritium (3H)-powered betavoltaic cells are presented in references [30]-[32]. The 63Ni betavoltaic battery is a common battery presented in different research studies [33]-[36]. Researchers have conducted evaluations of 90Sr betavoltaic power sources [37]-[39]. The 14C is also a radioisotope that can be used as a C-betavoltaic energy converter [40], but it has complicated production methods [41]-[42].

To increase the energy conversion efficiency of the betavoltaic cells, different methods are applied. Generally, the cell material and the cell structure are two important parameters. From material point of view, using wide-bandgap materials in betavoltaic cells have improved the cell characteristics. Wide-bandgap semiconductors have high radiation hardness and increase the energy conversion efficiency of the cell due to the raise of the open-circuit voltage of the cell [43].

From structure point of view, it should be noted that a significant amount of the energy from the decayed particles is lost in flat structures and bulk junctions due to the angular distribution of beta particles as well as low effective junction area. It is possible to use nanowires and porous structures at p-n, p-i-n and Schottky junctions to greatly increase the effective area of the junction and the propagated electric fields. In this regard, various designs and structures have been investigated [44]. In Si-based betavoltaic devices, using porous Si increases efficiency by 0.22% [45]. Creating nanowire structures in gallium phosphide improves the cell efficiency compared to the bulk material [46]; for example, the efficiency of a betavoltaic microbattery based on TiO2 nanotubes with 63Ni source and activity of 8 mCi is reported to be 7.3% [47]. theoretical support for the parameter analysis of betavoltaic batteries is presented in [48] and is used for study on the series resistance of betavoltaic batteries.

Despite many efforts conducted and different methods presented to improve the betacell characteristics, there is still a need for an efficient betavoltaic cell to simultaneously benefits from optimum structure and material.

Schottky diode has many applications for radiation detectors. An InSb Schottky detector, fabricated from an undoped InSb wafer with Hall mobility was used for alpha particle detection. The output pulse of this InSb detector showed a very fast rise time, which was comparable with the output pulses of scintillation detectors [49].

In this study, a betavoltaic device is designed based on a Schottky junction between Pt metal and ZnO semiconductor with a porous structure. The significant difference in work function between Pt (5.64 eV) and ZnO [50] results in an effective Schottky barrier. In the proposed device, ZnO serves as a wide energy bandgap material with high energy hardness; The porosity of ZnO effectively increase the junction area, leading to a higher number of trapped beta particles. To achieve the desired beta-generated electron-hole concentration inside the structure, empirical dose functions are simulated in C++ code and the results are linked to SILVACO-TCAD 3D simulator. The radioisotope 63Ni with an activity of 1 mCi, due to its long half-life, is employed in the C++ code as the source of beta particles. The structure of the device and Schottky barrier formed at the Pt-ZnO junction are thoroughly investigated. Subsequently, the results obtained from the SILVACO simulator are presented. Finally, an analysis is conducted on the results of the porous structure for different metals and compared with other relevant samples.

Device Structure

The structure was simulated using the TCAD-SILVACO 3D simulator. A Schottky junction was formed by employing an n-type ZnO semiconductor and Pt metal. The Pt layer was extended on the surface of the ZnO structure, and zinc oxide nanowires were applied to the irradiated surface. ZnO nanowires can be created through the porosity of the ZnO surface, which increases the effective cell area and enhances the electric fields between pores [51]. This effectively improves the trapping of emitted beta particles in the cells and increases the contact surface area. The trapped particles generate electron-hole pairs in ZnO, which are then collected by the internal electric field of the Schottky junction, thereby improving efficiency (refer to the simulation and results section). The porous surface consists of comb-shaped indentations with a height of 2 µm. Table 1 presents the structure dimensions. To reduce beta particle flux attenuation, a small thickness is considered for the anode electrode, ranging from 0.04 to 0.1 µm. The anode electrode in the upper part is exposed on the processed porous surfaces of zinc oxide. The cathode electrode is created on the backside of the ZnO

wafer, which also serves as the substrate for the structure in the simulations. ZnO nanowires are considered on the surface. (see Fig. 1 for a 3D schematic of the structure).

ZnO is a material with wide energy bandgap and excellent properties. It has high band gap energy (3.37 eV) [52] and can be used as a suitable material in piezoelectric converters, optical waveguides, acoustic wave devices, varistors, sensors, etc. In addition to its widespread use in current industries, ZnO is considered as a "material of the future" [53]. ZnO crystallizes in Wurtzite (B4 type) structure at the ambient pressure and temperature. This structure is a hexagonal lattice belonging to the space group P63mc and is characterized by two interconnected subnets of Zn^{2+} and O^{2-}. This regular tetrahedral arrangement creates polar symmetry which is an important factor in crystal growth and etching [50]. In betavoltaic devices, semiconductor structure should be exposed to radiation for a long time. Therefore, the used material should have good resistance to radiation damage so as not to cause rapid damage to the device [54]-[56]. ZnO semiconductor exhibits excellent hardness against radiation damage, so that it is not damaged by high-energy electron bombardment (> 1.6 Mev) [57]. It can be simply doped to achieve an n-type ZnO semiconductor [58]-[59]. In this work, ZnO was doped with 10^{15} atom/cm^3 to achieve the n-type semiconductor.

Table 1: Structure dimensions

Device length	5.9 um
Device width	3.8 um
Device height	6 um
Length of ZnO Nanowires	0.7 um
width of ZnO Nanowires	0.4 um
height of ZnO Nanowires	2 um

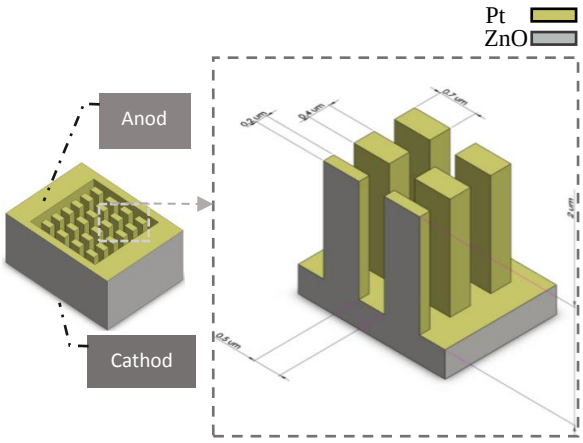


Fig. 1: The 3D schematic of the proposed betacell. The structure is a Schottky junction between ZnO and Pt.

A. *Metal-Semiconductor Junction*

Most of the useful properties of p-n junction can be achieved simply by forming a suitable junction between the metal and semiconductor. When a metal with work function ϕ_m is connected to a semiconductor with work function ϕ_s ($\phi_s<\phi_m$), a depletion region is formed in the semiconductor near the junction by transferring free electrons from the semiconductor to the metal and an electric field is created in this region to separate electron-hole pairs made by radiation [60] (Fig. 2). Schottky junction between Pt and ZnO in betavoltaic cell. Pt has Schottky contact with ZnO. The depletion region of the junction is hashed. Beta particles are incident at Pt and generate electron-hole pairs in ZnO. The beta-generated carriers are then separated in depletion region of the junction and can generate current and voltage in the external circuit. The energy band diagram of the junction is also presented on top of the junction schematic. The Schottky barrier potential of the junction and contact potential are demonstrated as qϕB and qϕi in Fig. 2. Schottky junctions are superior to p-n junction in terms of being used in radioisotope devices due to various reasons, including their simpler fabrication mechanism.

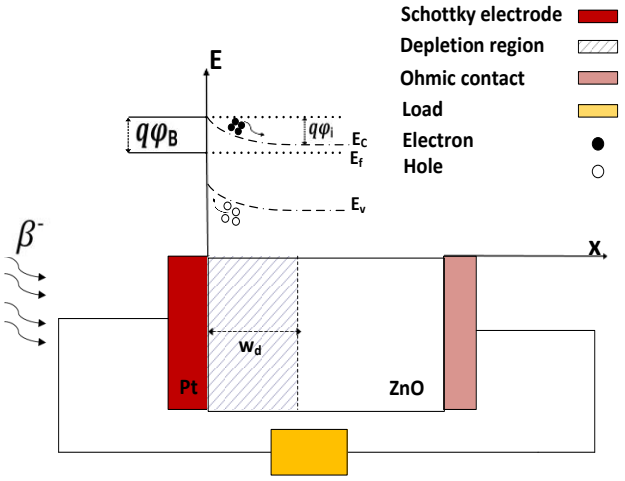


Fig. 2: Schottky junction between Pt and ZnO and The energy band diagram of the junction.

The Schottky junction barrier is defined by the metal work function and semiconductor electron affinity, as presented by (1).

$$\varphi_B = \varphi_m - X \quad (1)$$

where X is ZnO electron affinity (4.1- 4.5 eV) [50], [61]-[62] and ϕ_B is Schottky barrier height, which is 1.54 eV in Pt/ZnO junction. Figs. 3(a) and 3(b) demonstrate current-voltage curve in dark mode and energy band diagram of the structure, respectively, showing formation of Schottky junction and barrier at ZnO-Pt junction. The I-V curve of Pt/porous ZnO in dark mode; The rectifying behavior of the curve demonstrates that the Schottky barrier is created at the interface of metal and semiconductor (Fig 3(a)). Energy band diagram of the junction between Pt and ZnO. As shown, a potential

barrier with the height of 1.54eV is created at the ZnO surface (Figs. 3(b)). Table 2 presents density and work function of some metals that can be used to form Schottky or ohmic junctions with ZnO. Metals with large work function are suitable for forming Schottky junction with ZnO. Hence, Pt, Ir, Ni, Pd, and Au metals were examined. The results presented in the comparison section showed Pt was suitable for forming Schottky junction with ZnO.

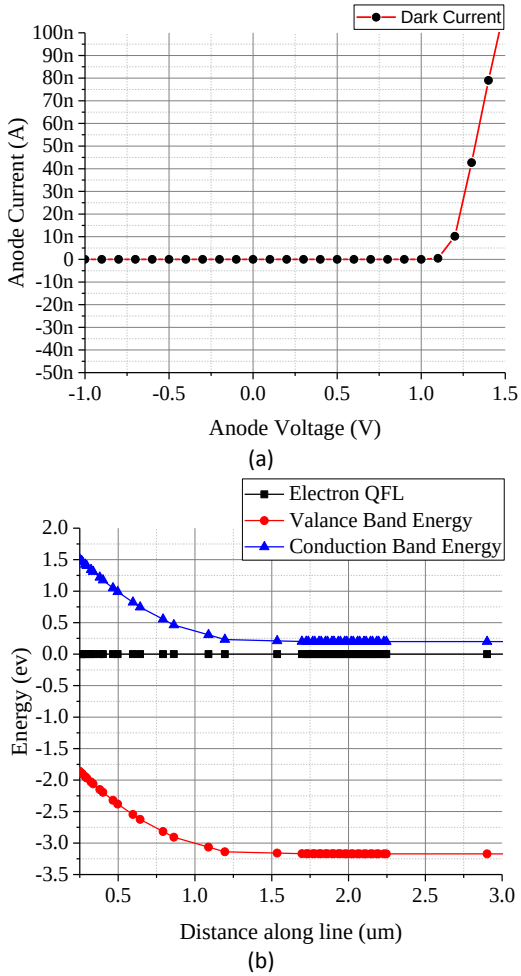


Fig. 3: a) The I-V curve of Pt/porous ZnO in dark mode. b) Energy band diagram of the junction between Pt and ZnO.

Table 2: Work function and density of some metals for forming junction with ZnO [50]

Metal	φm	Density ($\frac{gr}{cm^3}$)
Ta	4.25	16.69
Ag	4.26	10.5
Al	4.28	2.7
Ti	4.33	4.51
Zn	4.47	7.133
Mo	4.6	10.28
Au	5.1	19.32
Pd	5.12	12.02
Ni	5.15	8.9
Ir	5.25	22.56
Pt	5.64	21.45

B. Radioisotope source

Radioisotopes have received great attention due to their high energy storage. They release this energy during their half-life. Long half-life is among the important factors for the sources of betavoltaic devices. In general, radioisotope is selected based on the type of radiation, energy, specific activity, cost, and half-life.

In this work, ^{63}Ni beta source with the half-life of 100.2 years, average energy of 17.4 keV, and maximum energy of 67 keV is simulated by C-Interpreter code method. According to (2), the radiant power of 10^{-7} W is obtained for the activity of 1 mCi.

$$P_{Ni-63} = 3.7 \times 10^{10} \times A \times q \times E_{avg} \quad (2)$$

where A stands for the source activity, E_{avg} is the average energy of radioisotope source and q is the electron charge.

Simulations and Results

The presented Pt/porous ZnO Schottky junction is tested under an electron beam with the average energy of ^{63}Ni beta particles. To achieve the beta-generated carrier concentrations inside the device, an analytical function for electron-hole generation rate is simulated in C-interpreter code and linked to SILVACO (Fig. 4).

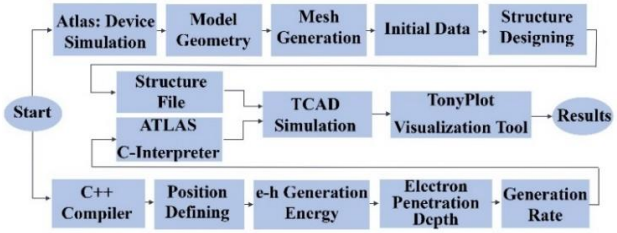


Fig. 4: Simulation flowchart.

As it is known, after the design of the structure in the ATLAS section of the SILVACO software, the C++ code is entered into the simulation as a source of beta radiation by using one of the SILVACO code abilities (the F.RADIATE BEAM statement). In this simulation, E. Yakimov's analytical function [63] is used to express the electron-hole generation rate, G(r) by (3):

$$G(r) = G_0 F(r,y,E) h(y,E) \quad (3)$$

In this equation, G_0, $F(r,y,E)$, and $h(y,E)$ refer to lateral-dose function, and depth-dose function respectively. The local rate of the electron-hole pair generation, G_0, depends on the electron beam energy(E), electron beam current(I), and the energy required to generate an electron-hole pair in material (e_i) as (4):

$$G_0 = \frac{EI}{qe_i} \quad (4)$$

in which q is the electron charge. The quantity e_i can be obtained as (5) [64]:

$$G_0 = \frac{EI}{qe_i} \quad (5)$$

e_i=2.596E_g + 0.714

Eg refers to the bandgap of the material, that is equal to 3.37 ev for zinc oxide. The depth-dose function, h(y, E), is calculated using Everhart [65] expression as (6):

$$h(y,E) = 0.6 + 6.21 \times R_{norm} - 12.4 \times {R_{norm}}^2 + 5.69 \times {R_{norm}}^3 \quad (6)$$

R_{norm} is the depth; more accurately it is the depth (y) normalized by the penetration depth of the beam R(E), which can be calculated as (7) [66]:

$$R(E) = \frac{0.0276 \times A \times E^n}{Z^{0.89} \times \rho} \quad (7)$$

where A is the atomic weight(g/mol), E is the beam energy (Kev), ρ is the density, and Z is the atomic number. n is often chosen to be 1.67 when E>5 Kev.

For the lateral-dose function, F(r,y,E), the empirical expression as proposed by Donolato by (8) is used [69], [74], [75]:

$$F(r,y,E) = \frac{1.76}{2\pi\sigma^2 R(E)} \exp[-(\frac{r}{\sigma})2] \quad (8)$$

where, $\sigma^2 = 0.36\delta 2 + 0.11\frac{y^3}{R(E)}$, δ is the electron beam diameter, $r^2=x^2$ and δ = 0.01×R(E).

The above-mentioned functions are used in C++ code to calculate the distribution of generated electron-hole pairs in the device.

Efficiency is among the important parameters of energy converters which is defined according to (9):

$$\eta = \frac{P_{out}}{P_{total}} \quad (9)$$

where P_{out} is the output power of betavoltaic device, which can be obtained by (10) and P_{total} is the total power received from the radioisotope source, which is equal to 10^{-7}W in ^{63}Ni source with the activity of 1 mCi (Section B. and (2)).

$$Pout = V_{oc}(v) \times I_{sc} \times FF \quad (10)$$

also, V_{OC} and J_{SC} are open-circuit voltage and short-circuit current, respectively. In I-V curve, V_{OC} is the voltage, at which the current is 0, and J_{SC} is the current, at which the voltage is 0. FF is the fill factor calculated by (11).

$$FF = \frac{P_{max}}{V_{oc}\, I_{sc}} \quad (11)$$

In this relation, P_{max} is the maximum value of (I_{sc}×V_{oc}) which can be read from power diagram.

Figs. 5 and 6 illustrate I-V and power diagrams of the structure designed based on Schottky junction between Pt and ZnO in non-porous and porous states, respectively. I-V curve entered the fourth region after irradiation, indicating the power is negative and is generated in the device. Short-circuit current, open-circuit voltage, FF, and efficiency of the non-porous structure are obtained as 12.36 nA, 1.115 V, 82.5%, and 11.37%, respectively. However, after creating pores on the surface exposed to radiation, these values increased to 37.5 nA, 1.23 V, 76.5%, and 35.5%, respectively. As can be seen, the short circuit current is considerably increased when ZnO nanowires are created on the surface. It shows that the number of the trapped beta particles is increased by using porous structure.

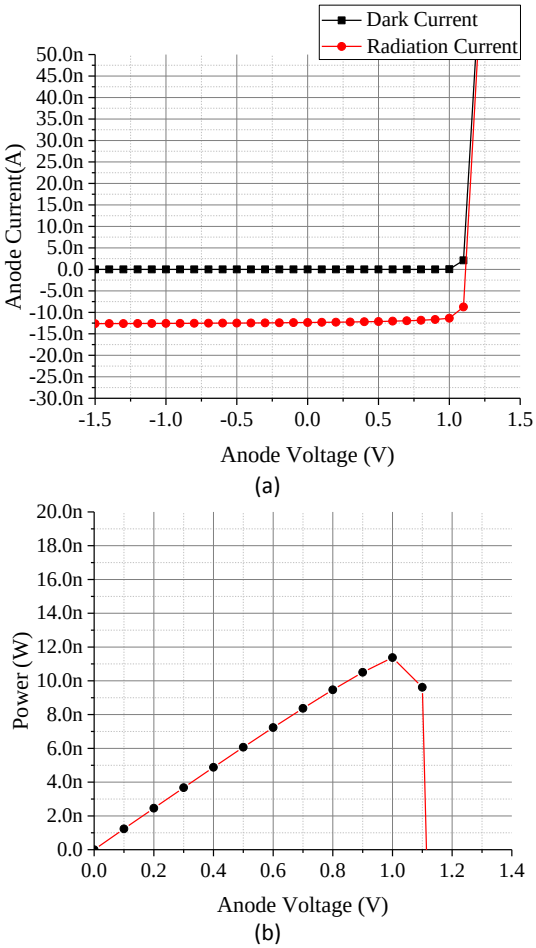


Fig. 5: a) I-V curve of non-porous structure in dark and radiation modes; b) Power diagram in terms of voltage for non-porous structure.

According to the formula below the short -circuit current is equal to [70] (12)

$$J_{sc} = N_e(\#pairs/s).1.6 \times 10^{18}(\frac{C}{pair}) \quad (12)$$

N_e shows the number of electron- hole pairs created per second in the depletion zone. Given the activity of the source of isotope and the comparison of structures in Table 4, this structure produces a suitable electron- hole pair. The VOC also has a semiconductor energy band.
Due to the semiconductor of zinc oxide on the maximum voltage of the open circuit of cell is 1.23 V, which has a good value compared to Activity and Table 4. The closer

the JSC times the VOC to Pmax, the fill factor is close to one and the structure is more suitable for an energy conversion cell.

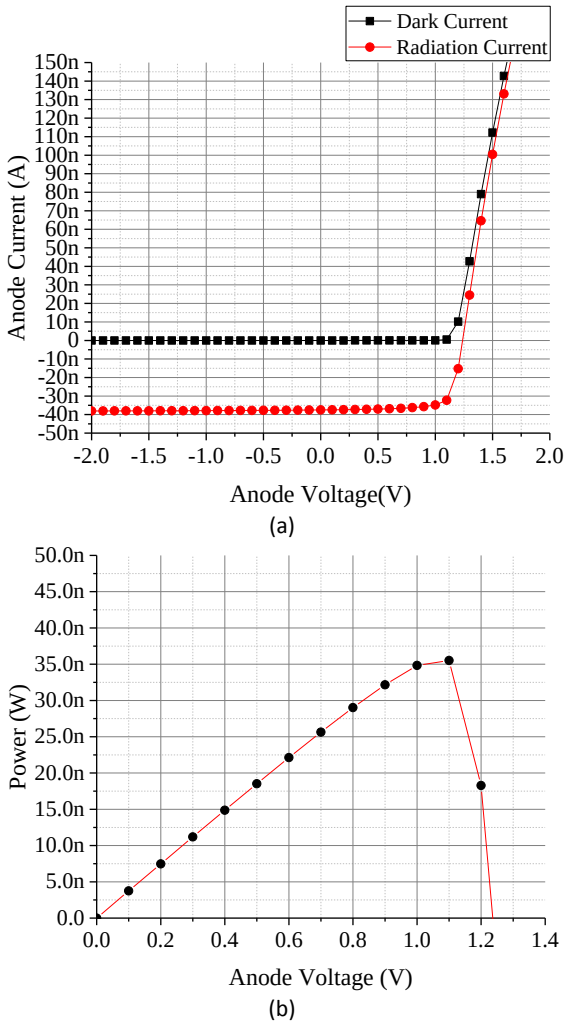


Fig. 6: a) I-V curve of porous structure in dark and radiation modes; b) Power diagram in terms of voltage for porous structure.

Table 3: Obtained results for different metals at Schottky junction with ZnO

Metal	short-circuit current(J_{SC})	open-circuit voltage (V_{OC})	Schottky barrier height	Maximum power (P_{max})
Pt	37.5 nA	1.237 V	1.54 ev	35.5 nW
Ir	37.1 nA	0.847 V	1.15 ev	22.9 nW
Ni	37 nA	0.74 V	1.05 ev	19.66 nW
Pd	36.9 nA	0.71 V	1.02 ev	18.54 nW
Au	36.8 nA	0.69 V	1 ev	17.26 nW

Comparison and Discussion

As shown in Table 2, various metals were investigated to form a Schottky junction with ZnO semiconductor in the porous structure, and the corresponding results are listed in Table 3. The short-circuit current values for Pt, Ir, Ni, Pd, and Au in the Schottky junction with ZnO were calculated as 37.5, 37.1, 37, 36.9, and 36.8 nA, respectively.

Additionally, their corresponding open-circuit voltage values were obtained as 1.237, 0.847, 0.74, 0.71, and 0.69 V, respectively. The results indicated Pt is the most suitable metal for forming Schottky junction in the structure due to its large work function. Fig. 7 illustrates the I-V curve in the radiation mode and the power diagram for different metals. To facilitate a comprehensive comparison of the cell results with other relevant published studies, Table 4 provides a summary of the findings from various betavoltaic cell investigations. The aim is to compile the characteristics of different betacells based on 63Ni. Generally, betacells utilizing wide bandgap materials tend to exhibit higher efficiency. Structures incorporating nanowires or porosity demonstrate improved short circuit current. This comparison highlights the superior performance of porous structures in betavoltaic battery outputs compared to non-porous structures.

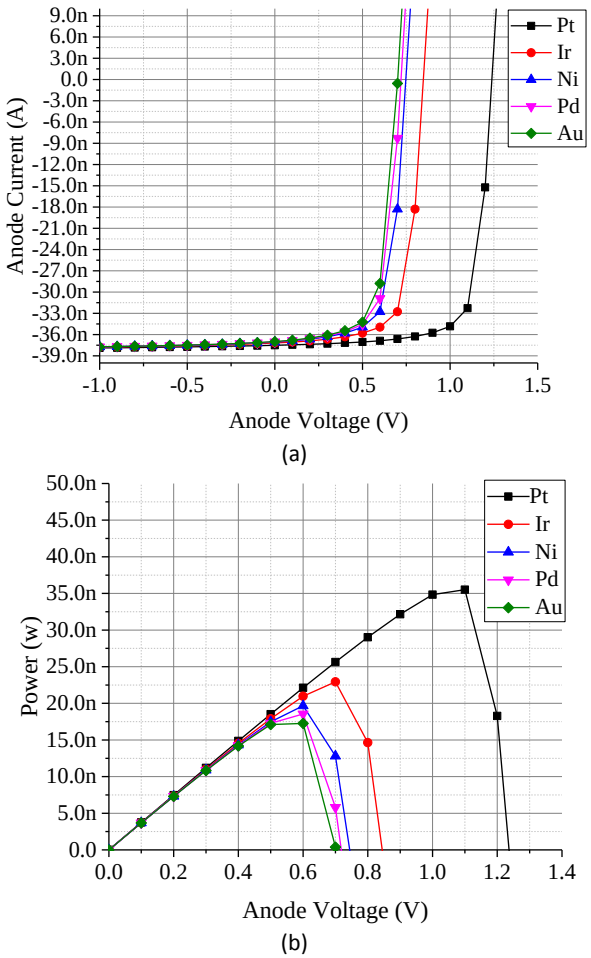


Fig. 7: a) I-V curve in radiation mode for different metals; b) Power diagram in terms of voltage for different metals.

Table 4: Comparing the obtained results with those of other studies

Reference	Efficiency	J_{SC}	V_{OC}	Junction	Semiconductor	Porous or Nanowires	Activity	Radiation source
This work	35.5%	37.5 nA	1.237 V	Schottky	Zno	yes	1 mci	^{63}Ni
[14]	4.4%	5.8 nA	2.1 V	p-i-n	GaN	no	-	^{63}Ni
[15]	1.5%	0.21 nA	0.95 V	p-i-n	SiC	no	0.16 mci	^{63}Ni
[18]	0.011%	1.2 nA/cm^2	0.3 V	p-i-n	GaN	no	50 mci	^{63}Ni
[21]	2.25%	7.5 nA	0.53 V	Schottky	GaN	no	1 mci	^{63}Ni
[33]	4.94%	573.3 nA	0.253 V	p-n	Si	no	100 mci	^{63}Ni
[47]	7.30%	12.43 nA	1.54 V	Schottky	TiO2	yes	8 mci	^{63}Ni
[50]	7.58%	0.1 uA	2.6 V	p-n	Zno	no	0.101 ci	^{63}Ni
[68]	2.85%	-	1.16 V	Schottky	GaN	no	-	^{63}Ni
[69]	4%	4.5 pA	0.69 V	p-i-n	GaInP	no	-	^{63}Ni
[71]	0.15%	13 nA/cm^2	6.5 mV	p-n	Si-SWNT	yes	3.3 mci	^{63}Ni
[72]	27.4 %	0.6 uA	4.9 v	p-n	Al0.7Ga0.3N	no	-	^{63}Ni
[73]	13.4%	16.36 nA	3.16 v	p-n	GaN	no	1 mci	^{147}Pm

Conclusion

This study explored a betavoltaic device with a porous structure based on a Schottky junction between Pt and ZnO semiconductor. The radioactive source employed was 63Ni, with a half-life of 100.2 years, an average energy of 17 keV, and an activity of 1 mCi, simulated using C++ code. The creation of pores effectively increased the contact surface area and trapped beta beams, resulting in improved performance metrics. The designed structure exhibited a short circuit current, open-circuit voltage, fill factor (FF), and efficiency of 37.5 nA, 1.237 V, 76.5%, and 35.5%, respectively.

Author Contributions

M. Amirmazlaghani designed the experiments. A. Ebadiyan performed the simulations. M. Amirmazlaghani, A. Ebadiyan and A. Shokri interpreted the results and wrote the manuscript. N. Darestani Farahani helped in validating and revising the manuscript.ach author role in the research participation must be mentioned clearly.

Acknowledgment

The authors would like to thank Dr. Farshid Raissi for his scientific comments.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

TCAD	Technology Computer-Aided Design
FF	Fill Factor

References

[1] Y. Huang, A. S. S. Vasan, R. Doraiswami, M. Osterman, M. Pecht, "MEMS reliability review," IEEE Trans. Device and Mate. Reliab., 12(2): 482-493, 2012.

[2] S. El-Shemy, A. H. Aly, H. Sayed, M. F. Eissa, "Production of intensifying blue light by Cherenkov radiation phenomena and its application as a power source," Opt. Quantum Electron., 54(2): 70, 2022.

[3] S. T. Revankar, T. E. Adams, "Advances in betavoltaic power sources," J. Energy Power Sources, 1(6): 321-329 ,2014.

[4] C. Qin, W. Yuan, D. Qiao, "Research on betavoltaic microbattery," Nucl. Electron. Detect. Technol., 28(2): 294-298, 2008.

[5] T. H. Smith, J. Greenborg, W. E. Matheson, "Benefit/risk analysis of cardiac pacemakers powered by betacel® Promethium-147 Batteries," Nucl. Technol., 26(1): 54-64, 1975.

[6] A. A. Svintsov, A. A. Krasnov, M. A. Polikarpov, A. Y. Polyakov, E. B. Yakimov, "Betavoltaic battery performance: Comparison of modeling and experiment," Appl. Radiat. Isot., 137: 184-189, 2018.

[7] M. Prelas, M. Boraas, F. D. L. Aguilar, J. D. Seelig, M. T. Tchouaso, D. Wisniewski, "Nuclear batteries and radioisotopes," Switzerland: Springer International Publishing, 2016.

[8] H. Chen, L. Jiang, X. Chen, "Design optimization of GaAs betavoltaic batteries," J. Phys. D: Appl. Phys., 44(21): 215303, 2011.

[9] H. Guo, H. Yang, Y. Zhang, "Betavoltaic microbatteries using porous silicon," IEEE 20th International Conference on Micro Electro Mechanical Systems (MEMS): 867-870 ,2007.

[10] S. Yao, Z. Song, X. Wang, H. San, Y. Yu, "Design and simulation of betavoltaic battery using large-grain polysilicon," Appl. Radiat. Isot., 70(10): 2388-2394, 2012.

[11] S. Rahastama, A. Waris, S. Viridi, F. Iskandar, "Optimization of surface passivation parameters in [147Pm]-Si planar pn junction betavoltaic based on analytical 1-D minority carrier diffusion equation approaches," Applied Radiation and Isotopes, 151, pp.226-234, 2019.

[12] T. Shimaoka, H. Umezawa, K. Ichikawa, J. Pernot, S. Koizumi, "Ultrahigh conversion efficiency of betavoltaic cell using diamond pn junction," Appl. Phys. Lett., 117(10):103902, 2020.

[13] V. N. Murashev, S. A. Legotin, S. I. Didenko, O. Rabinovich, A. A. Krasnov, S. U. Urchuk, "Improvement of Si-betavoltaic batteries technology," Adv. Mate. Res., 1070: 585-588, 2015.

[14] M. R. Khan, J. R. Smith, R. P. Tompkins, S. Kelley, M. Litz, J. Russo, J. Leathersich, F. S. Shahedipour-Sandvik, K. A. Jones, A. Iliadis, "Design and characterization of GaN pin diodes for betavoltaic devices," Solid-State Electron., 136: 24-29, 2017.

[15] H. Guo, Y. Shi, Y. Zhang, Y. Zhang, J. Han, "Fabrication of SiC pin betavoltaic cell with 63 Ni irradiation source," in Proc. IEEE International Conference of Electron Devices and Solid-State Circuits: 1-2, 2011.

[16] L. Zhang, H. L. Cheng, X. C. Hu, X. B. Xu, "Model and optimal design of 147Pm SiC-based betavoltaic cell," Superlattices Microstruct., 123: 60-70, 2018.

[17] M. Litz, W. Ray, J. Russo, S. Kelley, J. Smith, "Planar Homojunction Gallium Nitride (GaN) PiN device evaluated for betavoltaic energy conversion," Measurement and Analysis. US Army Research Laboratory Adelphi United States, 2016.

[18] N. A. Kuruoğlu, O. Özdemir, K. Bozkurt, "Betavoltaic study of a GaN pin structure grown by metal-organic vapour phase epitaxy with a Ni-63 source," Thin Solid Films, 636: 746-750, 2017.

[19] C. J. Eiting, V. Krishnamoorthy, S. Rodgers, T. George, J. D. Robertson, J. Brockman, "Demonstration of a radiation resistant, high efficiency SiC betavoltaic," Appl. Phys. Lett., 88(6): 064101, 2006.

[20] F. K. anasse, J. J. Pinajian, A. N. Tse, "Schottky barrier betavoltaic battery," IEEE Trans. Nucl. Sci., 23(1): 860-870, 1976.

[21] H. San, S. Yao, X. Wang, Z. Cheng, X. Chen, "Design and simulation of GaN based Schottky betavoltaic nuclear micro-battery," Appl. Radiat. Isot., 80: 17-22, 2013.

[22] Y. Liu, R. Hu, Y. Yang, G. Wang, S. Luo, N. Liu, "Investigation on a radiation tolerant betavoltaic battery based on Schottky barrier diode," Appl. Radiat. Isot., 70(3): 438-441, 2012.

[23] G. Peng, L. Xiao-Ying, Y. Xian-Wang, "4H-SiC Schottky betavoltaic micro battery," Micronanoelectron. Technol., 47(3): 157-162, 2010.

[24] J. W. Murphy, C. D. Frye, R.A. Henderson, M. A. Stoyer, L. F. Voss, R. J. Nikolic, "Demonstration of a three-dimensionally structured betavoltaic," J. Electron. Mate., 50(3): 1380-1385, 2021.

[25] K. A. Rosenkranz, "Clinical experience with nuclear-powered pacemakers (Promethium-147)," Engineering in Medicine. Springer, Berlin, Heidelberg, (pp. 503-529), 1975.

[26] X. Tang, Y. Liu, D. Ding, D. Chen, "Optimization design of GaN betavoltaic microbattery," Sci. China Technol. Sci., 55(3): 659-664, 2012.

[27] Z. Movahedian, H. Tavakoli-Anbaran, "Design and optimization of Si-35S betavoltaic liquid nuclear battery in micro dimensions in order to build," Ann. Nucl. Energy, 143: 107483, 2020.

[28] S. Theirrattanakul, M. Prelas, "A methodology for efficiency optimization of betavoltaic cell design using an isotropic planar source having an energy dependent beta particle distribution," Appl. Radiat. Isot., 127: 41-46, 2017.

[29] D. Meier, A. Garnov, J. Robertson, J. Kwon, T. Wacharasindhu, "Production of 35S for a liquid semiconductor betavoltaic," J. Radioanal. Nucl. Chem., 282(1): 271-274, 2009.

[30] S. Deus, "Tritium-powered betavoltaic cells based on amorphous silicon," in Proc. Conference Record of the Twenty-Eighth IEEE Photovoltaic Specialists Conference-2000 (Cat. No. 00CH37036): 1246-1249, 2000.

[31] J. Russo, M. Litz, W. Ray II, G. M. Rosen, D. I. Bigio, R. Fazio, "Development of tritiated nitroxide for nuclear battery," Appl. Radiat. Isot., 125: 66-73, 2017.

[32] C. Honsberg, W. A. Doolittle, M. Allen, C. Wang, " GaN betavoltaic energy converters," in Proc, Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference: 102-105, 2005.

[33] X. Tang, D. Ding, Y. Liu, D. Chen, "Optimization design and analysis of Si-63 Ni betavoltaic battery," Sci. China Technol. Sci., 55(4): 990-996, 2012.

[34] H. Guo, A. Lal, "Nanopower betavoltaic microbatteries," in Proc. TRANSDUCERS'03. 12th International Conference on Solid-State Sensors, Actuators and Microsystems. Digest of Technical Papers (Cat. No. 03TH8664), 1: 36-39, 2003.

[35] Z. U. O. Guoping, Z. U. O. U. Jianliang, K. E. Guotu, "A Simple theoretical model for 63Ni betavoltaic battery," Appl. Radiat. Isot., 82: 119-125, 2013.

[36] S. Butera, G. Lioliou, A. M. Barnett, " Temperature effects on gallium arsenide 63Ni betavoltaic cell," Appl. Radiat. Isot., 125: 42-47, 2017.

[37] S. Rahastama, A. Waris, "Analytical study of 90Sr betavoltaic nuclear battery performance based on pn junction silicon," J. Phys. Conf. Ser., 739(1): 012003, 2016.

[38] J. Dixon, A. Rajan, S. Bohlemann, D. Coso, A. D. Upadhyaya, A. Rohatgi, S. Chu, A. Majumdar, S. Yee, "Evaluation of a silicon 90 Sr betavoltaic power source," Sci. Rep., 6(1): 1-6, 2016.

[39] S. McNamee, D. Wagner, E. M. Fiordaliso, D. Novog, R. R. LaPierre, "GaP nanowire betavoltaic device," Nanotechnol., 30(7): 075401, 2018.

[40] V. I. Chepurnov, M. V. Dolgopolov, A. V. Gurskaya, A. A Akimchenko, O. V. Kuznetsov, A. V. Radenko, V. V. Radenko, A. S. Mashnin, "C-betavoltaic energyconverter in por-SIC/SI," Int. J. Mater. Sci. Non-Equilib. Phase Transform., 3(3):119-120, 2017.

[41] A. S. Petrovskaya, S. V. Surov, A. Y. Kladkov, A. B. Tsyganov, "New thermo-plasma technology for selective 14C isotope extraction from irradiated reactor graphite," in Proc. AIP Conference, 2179(1): 020020, 2019.

[42] D. Belmont, D. J. Marín-Lámbarri, H. S. Cruz-Galindo, A. Huerta, A. M. Martínez, J. Mas-Ruiz, M. E. Ortiz, R. Raya-Arredondo, G. Reza, M. Rodríguez-Ceja, S. Sandoval-Hipólito, "14C/12C isotopic ratio of a research nuclear reactor graphite control bar," Nucl. Instrum. Methods Phys. Res. Sect. B, 485: 10-12, 2020.

[43] R. K. Yürük, H. Tütüncüler, "Theoretical investigation of high-efficiency GaN–Si heterojunction betavoltaic battery," Can. J. Phys., 97(9): 1031-1038, 2019.

[44] A. A. Krasnov, V. V. Starkov, S. A. Legotin, O. I. Rabinovich, S. I. Didenko, V. N. Murashev, V. V. Cheverikin, E. B. Yakimov, N. A. Fedulova, B. I. Rogozev, A. S. Laryushkin, "Development of betavoltaic cell technology production based on microchannel silicon and its electrical parameters evaluation," Appl. Radiat. Isot., 121: 71-75, 2017.

[45] W. Sun, N. P. Kherani, K. D. Hirschman, L. L. Gadeken, P. M. Fauchet, "A three-dimensional porous silicon p–n diode for betavoltaics and photovoltaics," Adv. Mater., 17(10): 1230-1233, 2005.

[46] D. Wagner, D. R. Novog, R. R. LaPierre, "Simulation and optimization of current generation in gallium phosphide nanowire betavoltaic devices," J. Appl. Phys., 125(16): 165704, 2019.

[47] Q. Zhang, R. Chen, H. San, G. Liu, K. Wang, "Betavoltaic effect in titanium dioxide nanotube arrays under build-in potential difference," J. Power Sources, 282: 529-533, 2015.

[48] J. Zhang, Y. Han, L. Ren, X. Wang, H. He, C. Chen, T. Li, "Study on the series resistance of betavoltaic batteries," Semicon. Sci. Technol., 37(12): 125009, 2022.

[49] I. Kanno, S. Hishiki, Y. Kogetsu, T. Nakamura, M. Katagiri, "Fast response of InSb Schottky detector," Rev. Sci. Instrum., 78(5): 056103, 2007.

[50] X. Y. Li, J. B. Lu, R. Z. Zheng, Y. Wang, X. Xu, Y. M. Liu, R. He, "Comparison of time-related electrical properties of PN junctions and Schottky diodes for ZnO-based betavoltaic batteries," Nucl. Sci. Tech., 31(2): 1-12, 2020.

[51] L. Schmidt-Mende, J. L. MacManus-Driscoll, "ZnO–nanostructures, defects, and devices," Mater. Today, 10(5): 40-48, 2007.

[52] A. B. Djurišić, Y. H. Leung,2006. "Optical properties of ZnO," Nanostruct. Small, 2(8-9): 944-961, 2006.

[53] V. A. Coleman, C. Jagadish, "Basic properties and applications of ZnO," Zinc Oxide Bulk, Thin Films and Nanostructures: 1-20, 2006.

[54] Y. Lei, Y. Yang, Y. Liu, H. Li, G. Wang, R. Hu, X. Xiong, S. Luo, "The radiation damage of crystalline silicon PN diode in tritium beta-voltaic battery," Appl. Radiat. Isot., 90: 165-169, 2014.

[55] J. B. Gibson, A. N. Goland, M. Milgram, G. Vineyard, "Dynamics of radiation damage," Phys. Rev., 120(4): 1229, 1960.

[56] J. W. Corbett, "Electron radiation damage in semiconductors," Academic Press, New York, 1966.

[57] D. C. Look, D. C. Reynolds, J. W. Hemsky, R. L. Jones, J. R. Sizelove, "Production and annealing of electron irradiation damage in ZnO," Appl. Phys. Lett., 75(6): 811-813, 1999.

[58] K. P. Ong, D. J. Singh, P. Wu, "Analysis of the thermoelectric properties of n-type ZnO," Phys. Rev. B, 83(11): 115110, 2011.

[59] D. H. Kim, G. W. Lee, Y. C. Kim, "Interaction of zinc interstitial with oxygen vacancy in zinc oxide: An origin of n-type doping," Solid State Commun., 152(18): 1711-1714, 2012.

[60] L. J. Brillson, Y. Lu, "ZnO Schottky barriers and Ohmic contacts," J. Appl. Phys., 109(12): 8, 2011.

[61] K. B. Sundaram, A. Khan, "Work function determination of zinc oxide films," J. Vac. Sci. Technol. A, 15(2): 428-430, 1997.

[62] B. Hussain, A. Aslam, T. M. Khan, M. Creighton, B. Zohuri, "Electron affinity and bandgap optimization of zinc oxide for improved performance of ZnO/Si heterojunction solar cell using PC1D simulations," Electron., 8(2): 238, 2019.

[63] E. Yakimov, "Electron beam induced current investigation of electrical inhomogeneities with high spatial resolution," Scanning Microscopy, 6(1): 81-96, 1992.

[64] T. Kobayashi, T. Sugita, S. I. Takayanagi, M. Iio, Y. Sasaki, "GaAs biomedical probes and their applications in nuclear medicine," IEEE Trans. Nucl. Sci., 20(1): 310-317, 1973.

[65] T. E. Everhart, P. H. Hoff, "Determination of kilovolt electron energy dissipation vs penetration distance in solid materials," J. Appl. Phys., 42(13): 5837-5846, 1971.

[66] K. A. Kanaya, S. Okayama, "Penetration and energy-loss theory of electrons in solid targets," J. Phys. D: Appl. Phys., 5(1): 43, 1972.

[67] C. Donolato, "An analytical model of SEM and STEM charge collection images of dislocations in thin semiconductor layers: I. Minority carrier generation, diffusion, and collection," Phys. Status Solidi (A), 65(2): 649-658, 1981.

[68] R. Zheng, J. Lu, Y. Liu, X. Li, X. Xu, R. He, Z. Tao, Y. Gao, "Comparative study of GaN betavoltaic battery based on pn junction and Schottky barrier diode," Radiat. Phys. Chem., 168: 108595, 2020.

[69] S. Butera, M. D. C. Whitaker, A. B. Krysa, A. M. Barnett, "Investigation of a temperature tolerant InGaP (GaInP) converter layer for a 63Ni betavoltaic cell," J. Phys. D: Appl. Phys., 50(34): 345101, 2017.

[70] M. Prelas et al., Nuclear Batteries and Radioisotopes, Lecture Note in Energy 56, DOI 10.1007/978-3-319-41724-0_2

[71] C. C. Chen, Y. Y. Chang, J. W. Zhang, 2012, "A novel betavoltaic microbattery based on SWNTs thin film-silicon heterojunction," in Proc. IEEE 25th International Conference on Micro Electro Mechanical Systems (MEMS): 1197-1200, 2012.

[72] C. Honsberg, W. A. Doolittle, M. Allen, C. Wang. "GaN betavoltaic energy converters," in Proc. Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference: 102-105, 2005.

[73] X. B. Tang, Y. P. Liu, D. Ding, D. Chen. "Optimization design of GaN betavoltaic microbattery," Sci. China Technol. Sci. 55(3): 659-664, 2012.

[74] M. Amirmazlaghani, A. Rajabi, Z. Pour-mohammadi, A. A. Sehat, "Betavoltaic battery based on reduced-Graphene-Oxide/Si heterojunction," Superlattices Microstruct., 145: 106602, 2020.

[75] Y. Naghipour, M. Amirmazlaghani, "Graphene/porous GaN Schottky Betacell," Micro Nanostruct., 168: 207323, 2020.

Biographies

Amirhossein Ebadiyan received his bachelor's degree in Electrical and Electronic Engineering from Shahid Rajaee Teacher Training University in Tehran in 2022, and then he was accepted as an honors student in the field of micro and nanoelectronic devices for the M.S. degree at Iran University of Science and Technology and is currently studying.

- Email: a_ebadiyan1378@elec.iust.ac.ir
- ORCID: 0000-0002-0948-1547
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Alireza Shokri received M.S. degree from Shahid Rajaee Teacher Training University, Iran in 2020, micro and nano electronics engineering. He is currently a Ph.D. student in Department of Electrical Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran, where he is doing research on betavoltaic batteries. His research interest includes nano scale transistors and betavoltaic cell.

- Email: arshokri@sru.ac.ir
- ORCID: 0000-0001-8766-6486
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Mina Amirmazlaghani received her PhD in the field of Nanoelectronics from K.N.Toosi University, Tehran, Iran, in 2014. Her Ph.D. thesis was about "Design and fabrication of Graphene-based IR and THz detectors". During 2012, she was a visiting researcher at TML (Terahertz and Millimeter wave laboratory), MC2, Chalmers University. She was with AIST, Tsukuba, Japan, early in 2014. From 2014, she has been an assistant professor at electronics department of Shahid Rajaee University in Tehran, Iran where she has established Nanoelectronics Lab#1 and #2, for simulation and fabrication process, respectively. Her current research interests include Graphene-Based Electronics, Design and Modeling of Nano-Scale Semiconductor Devices, Design and Fabrication of IR and THz Detectors, Beta-cell Batteries based on semiconductors and High Frequency Electronics.

- Email: m.mazlaghani@sru.ac.ir
- ORCID: 0000-0003-4235-3245
- Web of Science Researcher ID: AHC-9391-2022
- Scopus Author ID: NA
- Homepage: https://www.sru.ac.ir/en/faculty/school-of-electrical-engineering/mina-amir-mazlaghani/

Nikoo Darestani Farahani received the B.Sc. degree in Electronics Engineering from K.N.T. University, Tahran, Iran, in 2007 and M.Sc. degree in, Radiation Application Engineering from Shahid Beheshti university, Tehran, Iran, in 2010, and Ph.D. degree in Radiation Application Engineering from Shahid Beheshti university, Tehran, Iran, in 2016. She is currently an assistant professor at the Nuclear Science and Technology Research Institute (NSTRI), Tehran, Iran. Her research interests include Radiation Application and Detection, Monte Carlo Simulation, Plasma and nuclear fusion.

- Email: n_darestani@sbu.ac.ir
- ORCID: 0000-0003-1983-6222
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
A. Ebadiyan, A. Shokri, M. Amirmazlaghani, N. Darestani Farahani, "Betavoltaic battery using platinum/porous ZnO schottky junction," J. Electr. Comput. Eng. Innovations, 12(2): 343-352, 2024.

DOI: 10.22061/jecei.2024.10337.694

URL: https://jecei.sru.ac.ir/article_2069.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

A New Hybrid Predictive-PWM Control for Flying Capacitor Multilevel Inverter

P. Hamedani [*]***, M. Changizian***

Department of Railway Engineering and Transportation Planning, University of Isfahan, Isfahan, Iran.

Article Info

Article History:
Received 22 December 2023
Reviewed 13 January 2024
Revised 10 February 2024
Accepted 21 February 2024

Keywords:
Calculation burden
Delay compensation
Flying capacitor inverter
Hybrid method
Model Predictive Control (MPC)

[*]Corresponding Author's Email Address:
p.hamedani@eng.ui.ac.ir

Abstract

Background and Objectives: Model predictive control (MPC) is a practical and attractive control methodology for the control of power electronic converters and electrical motor drives. MPC has a simple structure and enables the simultaneous consideration of different objectives and constraints. However, when applying MPC for multilevel inverters (MLIs), especially at higher voltage levels, the number of switching states dramatically increases. This issue becomes more severe when MLIs are used to supply electrical motor drives.

Methods: This paper proposes three different MPC strategies that reduce the number of iterations and computation burden in a 3-phase 4-level flying capacitor inverter (FCI). Traditional MPC with a reduced number of switching conditions, split MPC, and hybrid MPC-PWM control are investigated in this work.

Results: In all methods, the capacitor voltages of the FCI are balanced during different operational conditions. The number of iterations is reduced from 512 in traditional MPC to at least 192 in the split MPC. Moreover, the split MPC strategy eliminates the usage and optimization of weighting factors for capacitors voltage balance. However, in the hybrid MPC-PWM control method in comparison to other methods, the voltage balancing time is much lower, the phase current tracks the reference more accurately, the transient time is lower, and the efficiency is higher. In addition, the capacitors voltage ripple is negligible in the hybrid MPC-PWM control method.

Conclusion: Simulation results manifest the effectiveness of the suggested hybrid MPC-PWM methodology. Results manifest that the hybrid MPC-PWM control offers perfect dynamic characteristics and succeeds in maintaining the voltage balance during different operational conditions.

Introduction

In recent years, different industries, such as electric transportation, have trended to increase the DC-link voltage to yield significant advantages [1]-[2] such as more machine efficiency, less charging time of batteries, lower cable weight, cost reduction, and lower loss [3]-[6]. However, by increasing the DC-link voltage, some drawbacks appear. In two-level inverters, increasing the DC-link voltage results in a higher dv/dt, reduction of the machine lifetime, higher electromagnetic interference (EMI), more eddy current loss in the cores, and more skin effect losses in the windings. When higher DC-link voltage is aimed, replacing the two-level inverter with a multilevel structure can reduce dv/dt and enable the usage of low-voltage switches. In addition, MLIs benefit from various aspects of a system, such as less total harmonic distortion (THD) of current and voltage waveforms, reduced EMI, higher efficiency, more fault-tolerance, more reliability, and lower cooling system size [7]-[10]. The mentioned advantages make multilevel inverters attractive for high-power applications.

In [11], the necessity of increasing the voltage levels in

multilevel inverters with higher DC-link voltages is investigated. In this work, the efficiency has been optimized according to power density, and the essential level numbers of an 800 V DC-link flying capacitor inverter are specified. Results prove that for moving from 400 V to 800 V DC-link, to achieve the same efficiency as a two-level 400 V DC-link inverter, a minimum seven-level inverter is needed [11].

In [9], a comprehensive review of MLI structures for traction application is reported. Neutral-point clamped (NPC) inverter, cascaded H-bridge (CHB) inverter, FC inverter, and modular multilevel converter (MMC) are the most utilized multilevel topologies in industrial applications. In [12], the appropriate control techniques for different multilevel structures are discussed and compared. Scaler control [13], Direct Torque Control (DTC) [14], Field Oriented Control (FOC) [15], and MPC [16] are well-known control strategies in AC motor drives with multilevel inverters.

In recent years, MPC has become an appealing control strategy in electrical motor drives due to its interesting advantages [17]-[19]. MPC has a simple implementation and offers the possibility to control several objectives and restrictions in the cost function [21]-[27]. In new investigations, MPC has been utilized to control MLIs [28]-[34]. The application of MPC for MLIs enables the regulation or optimization of various objectives such as load current [35], voltage balancing of the DC capacitors [36], switching frequency [37], inverter switching loss [38], and grid-side reactive power [35].

The main implementation issue of MPC in an MLI is the high number of switching states in which the cost function should be calculated. This increases the computation burden. Therefore, recently, various methods have been suggested for decreasing the computation burden in MLIs with MPC [39]-[47]. Moreover, when implementing the MPC for an MLI, the mathematical computation burden increases even more. This, in turn, restricts the practical utilization of MPC for an MLI. This work suggests a new hybrid MPC-PWM control method for a 3-phase 4-level FCI. The main advantage of this method in comparison with the traditional finite control set MPC (FCS-MPC) is the reduced computation burden.

Mathematical Model of 4-Level FCI

Fig.1 represents the topology of the 4-level FCI. Six IGBT switches with anti-parallel diodes and three capacitors are utilized to generate four voltage levels. The switches are arranged in two groups, up and down. The up and down switches should receive complementary firing pulses to avoid short circuit in the input DC source. The capacitors are charged with a voltage ratio equal to $v_{x1}:v_{x2}:v_{x3}=1:2:3$, with $x\in\{a,b,c\}$. In a 4-level FCI, 2^3=8 different switching states exist for each phase, and 8^3=512 different switching conditions exist in the three-phase 4-level FCI. Table 1 shows all feasible switching conditions of the single-phase 4-level FCI and the related voltage level. The final voltage values of the capacitors are:

$$v_{c1}^{*}=V_{dc}/3 \quad \text{and} \quad v_{c2}^{*}=2V_{dc}/3 \tag{1}$$

In the three-phase 4-level FCI, the voltages of capacitors are:

$$v_{1x}(t)=\frac{1}{C_{1x}}\int_0^t i_x(S_{2x}-S_{1x})(\tau)d\tau+v_{1x}(0) \tag{2}$$

$$v_{2x}(t)=\frac{1}{C_{2x}}\int_0^t i_x(S_{3x}-S_{2x})(\tau)d\tau+v_{2x}(0) \tag{3}$$

where, v_{jx} is the capacitor voltage in cell $j\in\{1,2,3\}$ in phase $x\in\{a,b,c\}$. i_x is the current in phase x. Moreover, S_{jx} is the switching state of the *j*-th IGBT switch *j* in phase *x* (as shown in Fig. 1(a)).

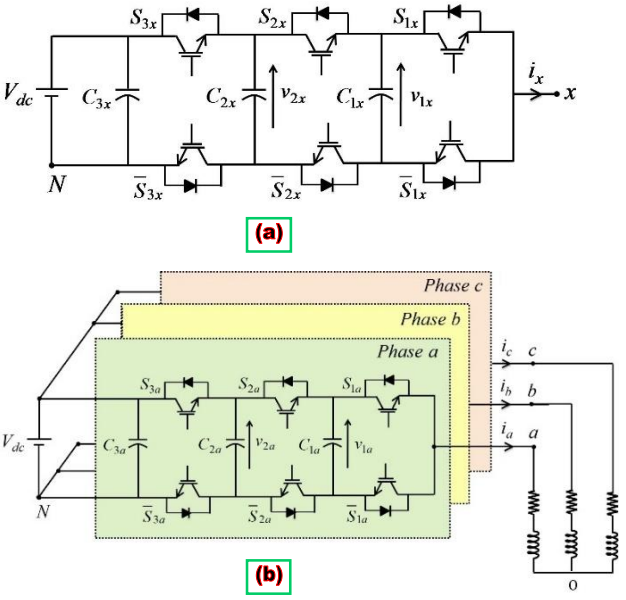


Fig. 1: Structure of 4-level FCI: a) single-phase, b) three-phase.

According to Fig. 1(b), the voltage equation of the output terminal *x* and the FCI neutral point *N* can be expressed as:

$$v_{xN}(t)=S_{3x}V_{dc}-(S_{3x}-S_{2x})v_{2x}(t)-(S_{2x}-S_{1x})v_{1x}(t) \tag{4}$$

while

$$v_{xN}=Ri_x+L\frac{di_x}{dt}+\frac{1}{3}(v_{aN}+v_{bN}+v_{cN}) \tag{5}$$

in which *R* and *L* are the load resistance and inductance, respectively.

Discretization of (2)-(3) using the Euler forward approximation gives:

$$v_{1x}(k+1)=v_{1x}(k)+\frac{T_s}{C_{1x}}i_x(S_{2x}-S_{1x}) \tag{6}$$

$$v_{2x}(k+1) = v_{2x}(k) + \frac{T_s}{C_{2x}} i_x (S_{3x} - S_{2x}) \quad (7)$$

in which k+1 is the next sampling instant and T_s is the sampling time.

Table 1: Switching states of the 4-level FCI

State	Output Voltage Level	Switching Pulse S_{1x}	S_{2x}	S_{3x}
1	0	0	0	0
2	$\frac{V_{dc}}{3}$	0	0	1
3		0	1	0
4		1	0	0
5	$\frac{2V_{dc}}{3}$	0	1	1
6		1	0	1
7		1	1	0
8	V_{dc}	1	1	1

Model Predictive Control in 4-Level FCI

Fig. 2 shows the block diagram for the conventional MPC of a 4-level FCI. Using the three-phase voltage and current values, the load voltage and current can be written as a space vector:

$$\boldsymbol{v} = \frac{2}{3}(v_{aN} + \boldsymbol{a}\, v_{bN} + \boldsymbol{a^2}\, v_{cN}) \quad (8)$$

$$\mathbf{i} = \frac{2}{3}(i_a + \boldsymbol{a}\, i_b + \boldsymbol{a^2}\, i_c) \quad (9)$$

where $\boldsymbol{a} = e^{j2\pi/3} = -\frac{1}{2} + j\frac{\sqrt{3}}{2}$.

The future value of the load current can be calculated as:

$$\boldsymbol{i}^p(k+1) = \left(1 - \frac{RT_s}{L}\right)\boldsymbol{i}(k) + \frac{T_s}{L}\,\boldsymbol{v}(k) \quad (10)$$

Using Clark's transformation, α and β axis currents can be extracted from the three-phase currents:

$$\begin{bmatrix} f_\alpha \\ f_\beta \end{bmatrix} = \frac{2}{3}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}\begin{bmatrix} f_a \\ f_b \\ f_c \end{bmatrix} \quad (11)$$

$$\mathbf{i} = i_\alpha + j i_\beta \quad \text{and} \quad \boldsymbol{i}^p = i_\alpha^p + j i_\beta^p \quad (12)$$

where f can be the reference and predicted current values.

The cost function can be defined as:

$$g = g_i + \lambda \sum_{x=a}^{c} g_{vx} \quad (13)$$

where

$$g_i = \left|i_\alpha^*(k+1) - i_\alpha^p(k+1)\right| + \left|i_\beta^*(k+1) - i_\beta^p(k+1)\right| \quad (14)$$

$$g_{v_x} = \left(v_{c1}^* - v_{1x}(k+1)\right)^2 + \left(v_{c2}^* - v_{2x}(k+1)\right)^2 \quad (15)$$

where λ_v is the weighting factor that adjusts the capacitors voltages and $x \in \{a,b,c\}$. Moreover, v_{c1}^* and v_{c2}^* are the final capacitors voltages given in (1). For finding the minimum of the cost function g specified in (13), MPC examines all 512 possible switching conditions explicitly. The prediction of currents, capacitors' voltages, and cost function are repeated frequently in each sampling instant.

Implementation of the MPC strategy for a 4-level FCI requires a large number of computations. Consequently, the control algorithm will have a significant time delay in practice. To moderate the computational delay, the delay compensation strategy is applied in this work. Fig. 3 shows the flowchart of the MPC strategy for a 4-level FCI with delay compensation. In this method, first, the present values of load current are measured and utilized to estimate the predicted currents at k+1 step time. The switching states are predicted in a shifted forward step time:

$$\boldsymbol{i}^p(k+2) = \left(1 - \frac{RT_s}{L}\right)\boldsymbol{i}(k+1) + \frac{T_s}{L}\,\boldsymbol{v}(k+1) \quad (16)$$

Consequently, the cost function can be expressed as:

$$g_v = \left|i_\alpha^*(k+2) - i_\alpha^p(k+2)\right| + \left|i_\beta^*(k+2) - i_\beta^p(k+2)\right| \quad (17)$$

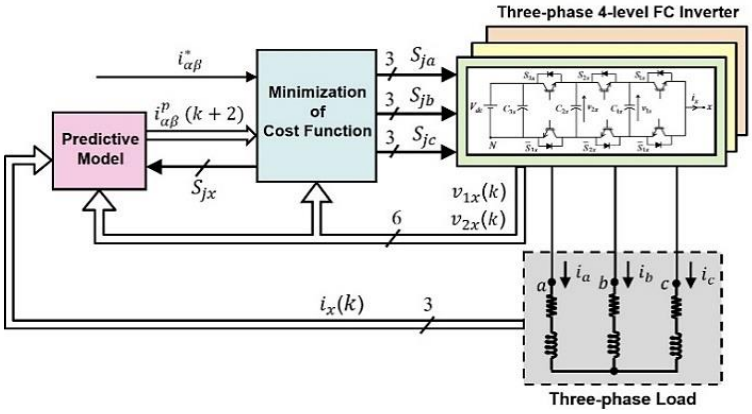


Fig. 2: Block diagram of the conventional MPC of 4-level FCI.

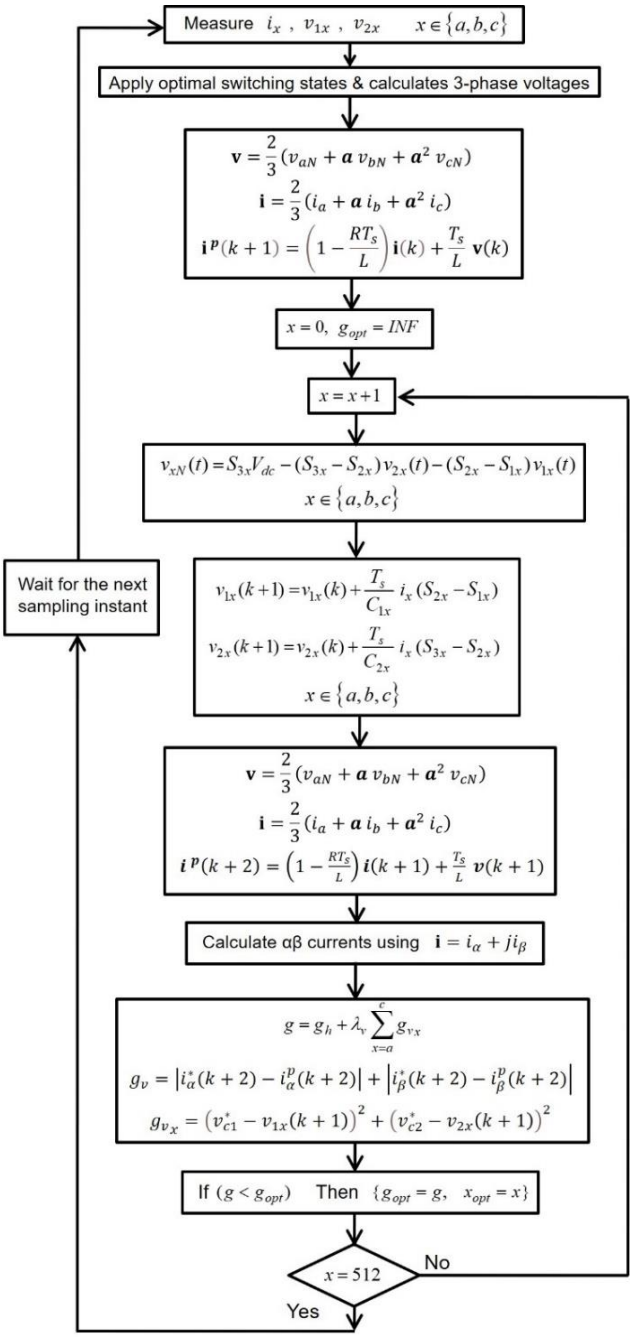


Fig. 3: Flowchart of the conventional MPC of a 4-level FCI.

Split MPC Strategy

To decrease the computation burden in the MPC of multilevel inverters, a new MPC method can be utilized. This method is based on the redundant states in the switching pattern of multilevel inverters. As shown in Table 1, for a 4-level FCI, three redundant stages exist in each medium level of the output voltage. It should be noted that these redundant states can be applied in the split MPC method only if the capacitor voltages reach approximately their desired values.

In the split MPC method, optimization of the objective function is divided into two separate stages. In the first stage, the tracking capability of the load currents is evaluated and optimized. In the second stage, the voltages of the capacitors are balanced. In the next step of the split MPC algorithm, the cost function can be considered the same as g_i in (14). Seeking all possible voltage levels of the output voltages and using (16), the optimum value of g_i is determined, and accordingly, the desired output voltage levels are selected. These voltage values are then considered as the input of the next stage. In the next stage, if the desired output voltage levels are equal to each of the medium voltage levels, i.e., $V_{dc}/3$ or $2V_{dc}/3$, the related redundant states of that level will be

sought to balance the capacitors voltages. In the next stage of this method, the cost function can be considered the same as g_{v_x} in (15), and the required voltages of the capacitors are calculated from (6)-(7). In this stage, each phase can be controlled independently to produce the optimal switching pulses, which balance the capacitor voltages. Fig. 4 represents the overall control diagram for the suggested split MPC of the 4-level FCI. Fig. 5 represents the flowchart of the split MPC method for a 4-level FCI with delay compensation.

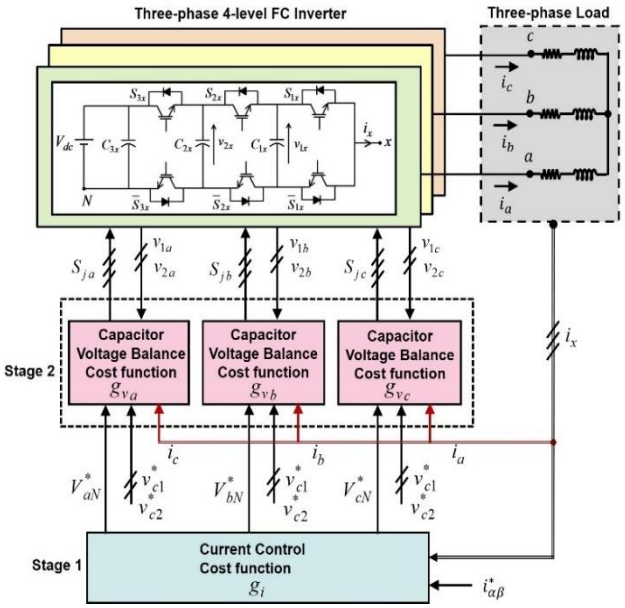


Fig. 4: The control scheme for the split MPC of a 4-level FCI.

A Hybrid PWM Linear Control with Voltage Balancing Based on MPC

Fig. 6 shows the block diagram of the hybrid PWM (Pulse Width Modulation) control based MPC for voltage balancing of the 4-level FCI. In this strategy, the capacitors voltages cannot reach their desired values. Therefore, a voltage balancing algorithm is utilized based on model predictive control. Fig. 7 shows the voltage balancing algorithm for the PWM linear control of the 4-level FCI.

Results and Discussion

To investigate the correctness of the proposed hybrid MPC strategy, simulation of a 4-level FCI has been carried out using Matlab/Simulink controlled by MPC with a reduced number of switching conditions, split MPC, and hybrid MPC method.

A list of parameters used in simulation is provided in Table 2. In the traditional MPC, the λ weighting factor is set at λ=0.1.

Fig. 8 demonstrates the reference and actual current in the 4-level FCI by MPC with a reduced number of switching conditions, split MPC, and hybrid MPC-PWM control, respectively.

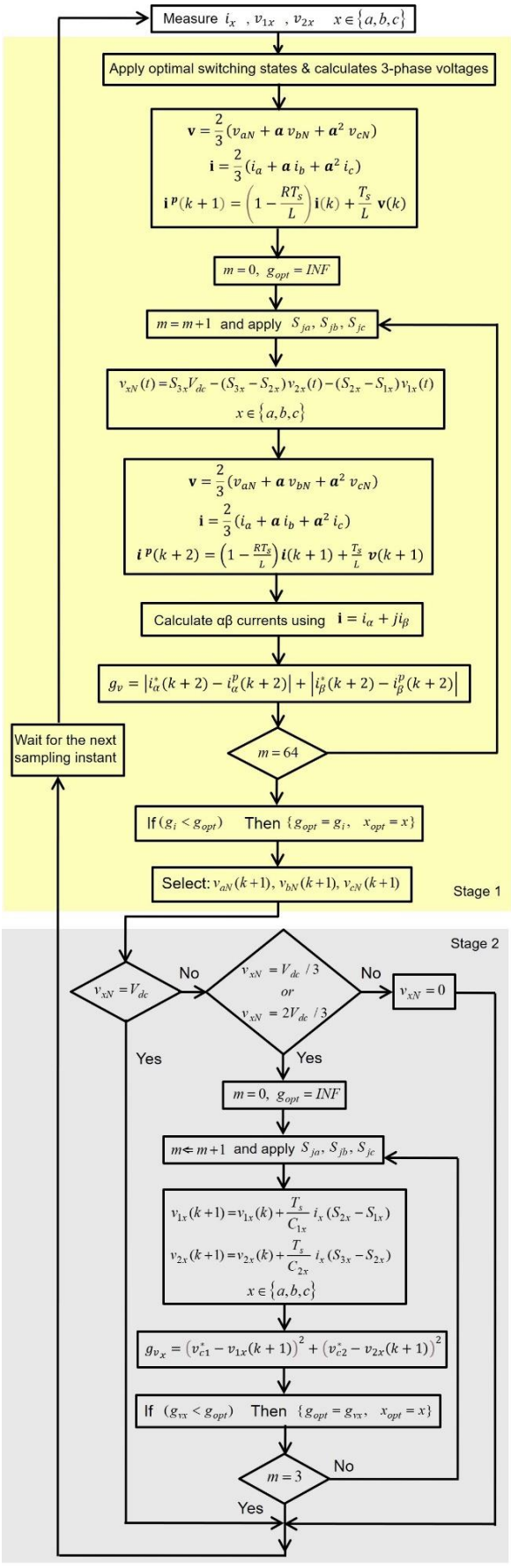


Fig. 5: Flowchart of the split MPC of 4-level FCI.

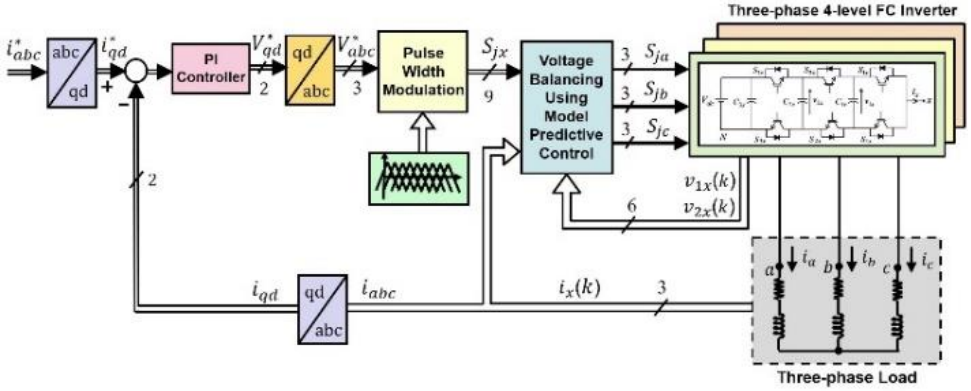


Fig. 6: The block diagram of the hybrid MPC-PWM control of a 4-level FCI.

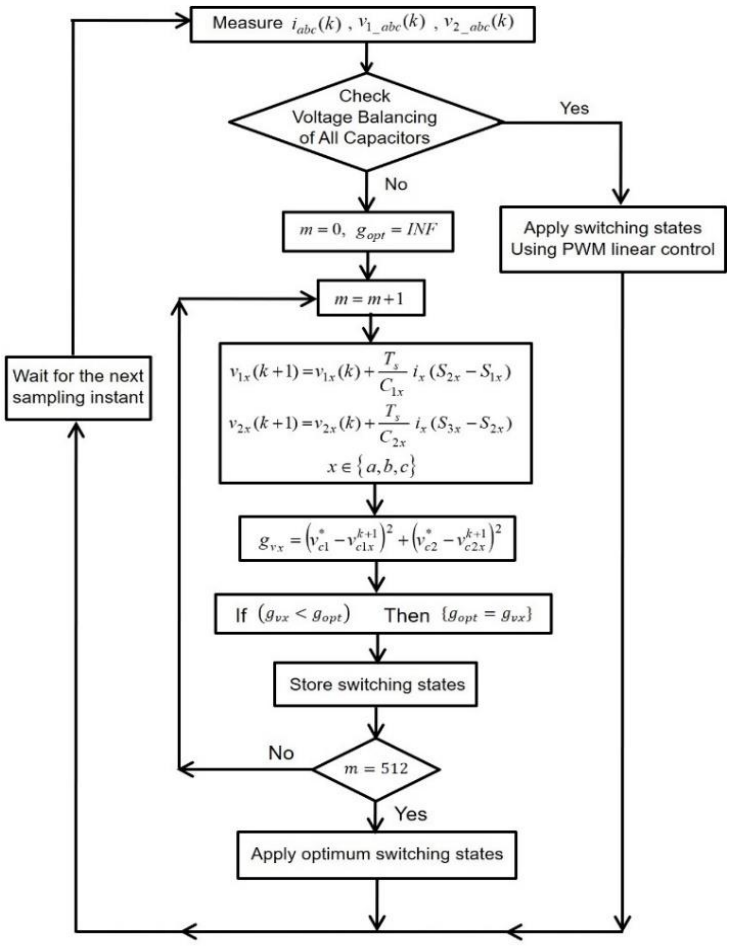


Fig. 7: Flowchart of the voltage balancing using the hybrid MPC-PWM method of a 4-level FCI.

Table 2: Simulation parameters of the 4-level FCI

Parameter	Value
V_{dc}	150 [V]
C_1, C_2	680 [µF]
R	10 [Ω]
L	10 [mH]

It is clear that in the hybrid MPC-PWM control method the phase current tracks the reference more accurately than in other methods. Fig. 9 shows the capacitors voltages in the 4-level FCI by MPC with a reduced number of switching conditions, split MPC, and hybrid MPC-PWM control, respectively. As can be seen, in all methods, the capacitors voltages of the 4-level FCI are balanced during different operational conditions. However, in the hybrid MPC-PWM control method, the voltage balancing time is much lower than in other methods and the capacitors voltage ripple is negligible.

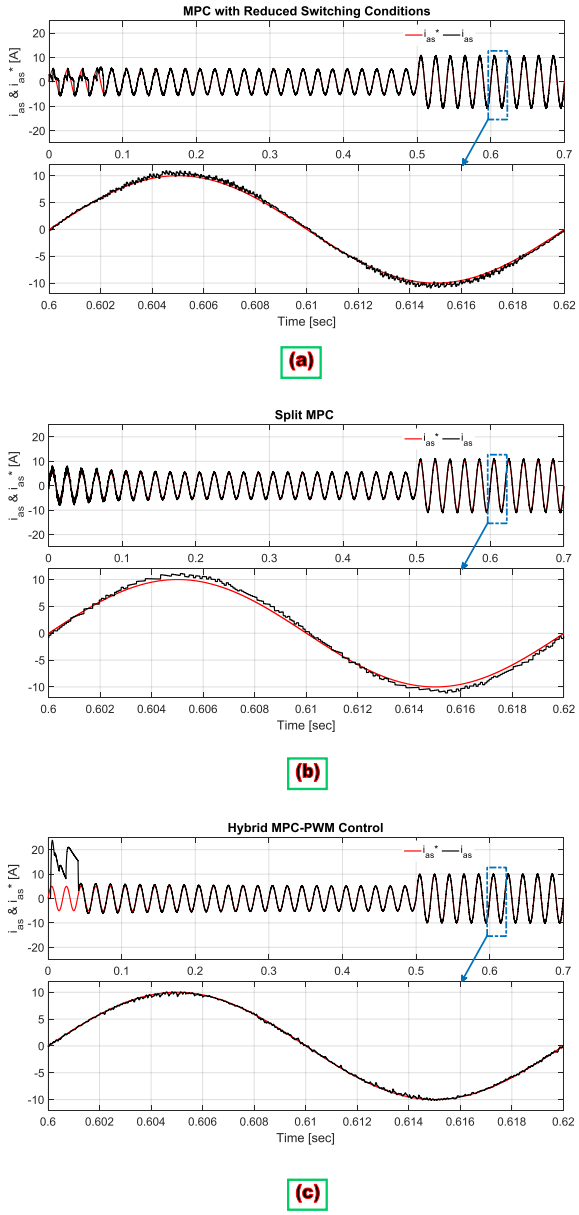


Fig. 8: Reference and actual current in the 4-level FCI: a) MPC with a reduced number of switching conditions, b) Split MPC, c) Hybrid MPC-PWM control.

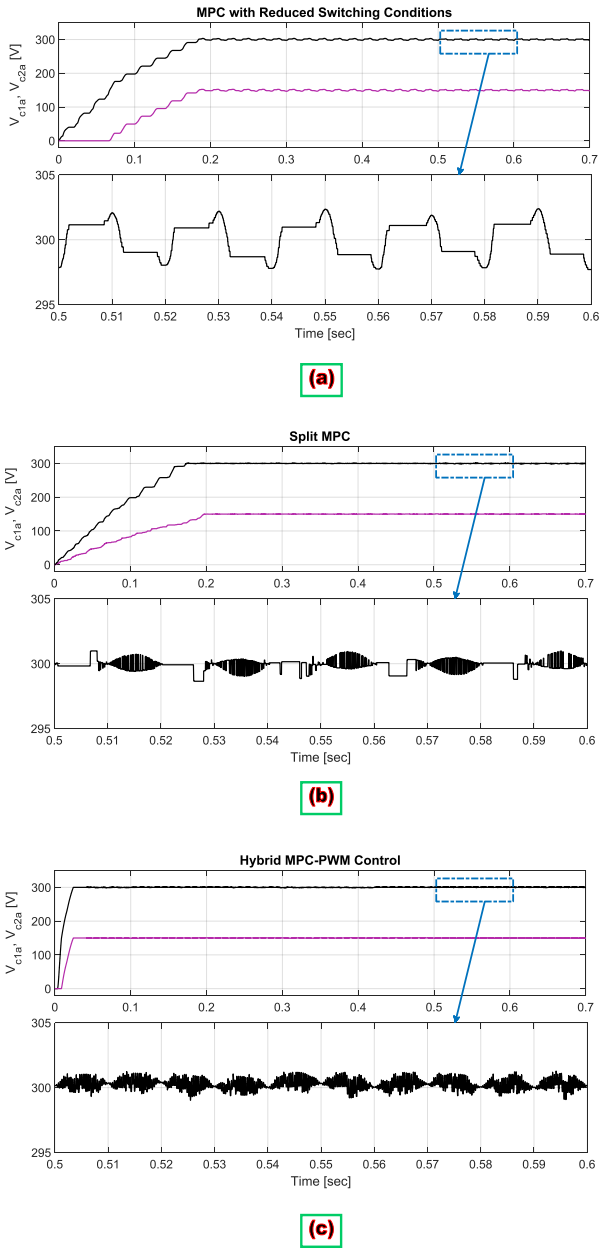


Fig. 9: Capacitors voltages in the 4-level FCI: a) MPC with a reduced number of switching conditions, b) Split MPC, c) Hybrid MPC-PWM control.

Fig. 10 illustrates the three-phase currents, the line voltage (V_{ab}), and the terminal voltage relative to the neutral point O (V_{ao}) in the 4-level FCI by MPC with a reduced number of switching conditions, split MPC, and hybrid MPC-PWM control, respectively. It is evident that in the hybrid MPC-PWM control strategy, the transient time is lower than in other methods. Fig. 11 illustrates the comparison of efficiency versus switching frequency in the 4-level FCI by MPC with a reduced number of switching conditions, split MPC, and hybrid MPC-PWM control, respectively. For calculation of the efficiency, first the total loss and the input power of the FCI is computed in the average switching frequency. The simulations are performed in PLECS software. The utilized IGBTs are IKW40N65ET7 manufactured by Infineon.

The IKW40N65ET7 is a 650 V, 40 A IGBT with anti-parallel diode. For calculating the total loss of the FCI, the conduction loss and switching loss are calculated for all the IGBTs and diodes.

As is evident, for the switching frequencies between 3 KHz to 5.5 KHz, the hybrid MPC-PWM control method results in higher efficiency than other methods. This is because the switching and conduction losses of the FCI are lower in the hybrid MPC-PWM control method.

In Table 3, a quantitative comparison of different control methods for the 4-level FCI is presented for switching frequency around 4 KHz.

According to the results, in the hybrid MPC-PWM control method, the voltage balancing time and the

current THD value are significantly lower than in the traditional MPC with a reduced number of switching conditions and the split MPC methods. Furthermore, the phase current tracking performance is more accurate in the hybrid MPC-PWM control method in comparison to other methods. The capacitors voltage ripple is low in the hybrid MPC-PWM and the split MPC methods.

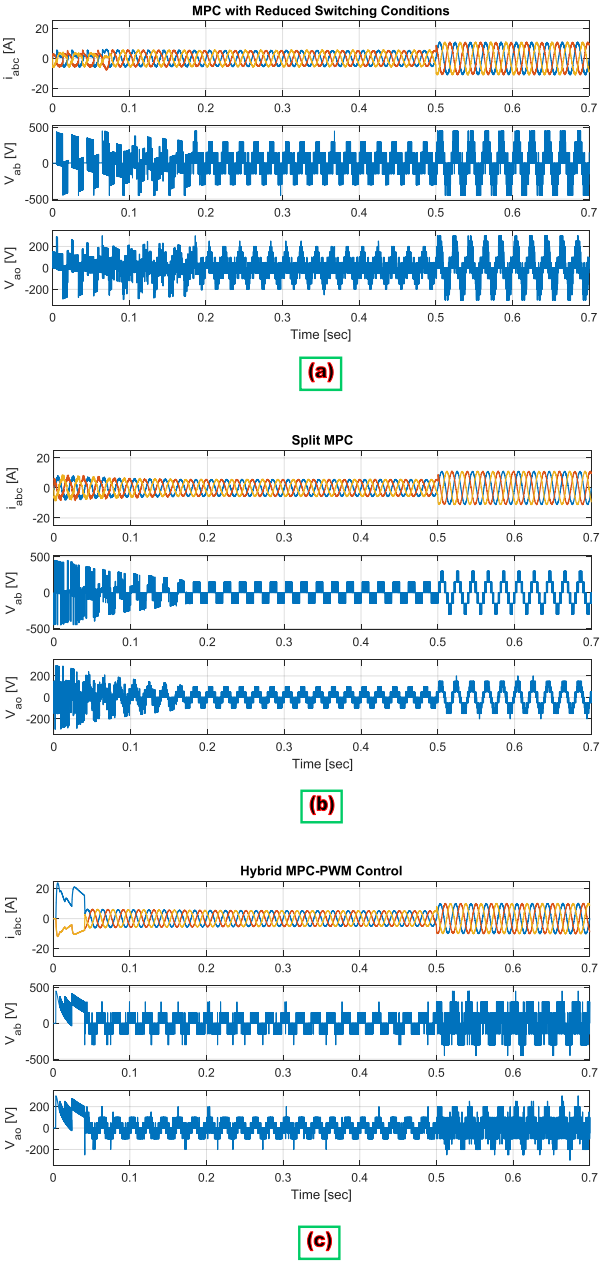


Fig. 10: Three-phase currents, line voltage (V_{ab}), terminal voltage relative to the neutral point O (V_{ao}) in the 4-level FCI: a) MPC with a reduced number of switching conditions, b) Split MPC, c) Hybrid MPC-PWM control.

Conclusion

This work suggests a new hybrid MPC-PWM control method for reducing the number of iterations and computation burden in a 3-phase 4-level FCI. In this strategy, the capacitors voltages of the 4-level FCI are balanced during different operational conditions. The number of iterations is decreased compared to the traditional MPC. Moreover, in the hybrid MPC-PWM control method compared to the traditional MPC with a reduced number of switching conditions and the split MPC, the voltage balancing time is much lower, the phase current tracks the reference more accurately, the transient time is lower, and the efficiency is higher. At the same time, the capacitors voltage ripple is low in the hybrid MPC-PWM control method. Simulation results validate that the proposed strategy achieves similar excellent dynamic behavior and voltage balance as the traditional MPC, but with a relatively reduced number of iterations and computation burden.

Table 3: Comparison of Different Methods for the 4-level FCI

Method	MPC with RSC	Split MPC	Hybrid MPC-PWM
Voltage Balancing Time [sec]	0.1	0.1	0.024
Current Tracking Performance	Medium	Medium	Good
Max. Current Overshoot [A]	4.19	4.05	23.8
Capacitor Voltage Ripple [%]	1.8	1.1	1.2
Current THD [%]	1.10	2.18	0.53
Inverter Losses [Watt]	35.34	34.69	35.63
Efficiency [%]	95.15	95.10	95.15
Average Switching Frequency [KHz]	4.31	4.36	4.19
Sampling Time [µsec]	35	45	100

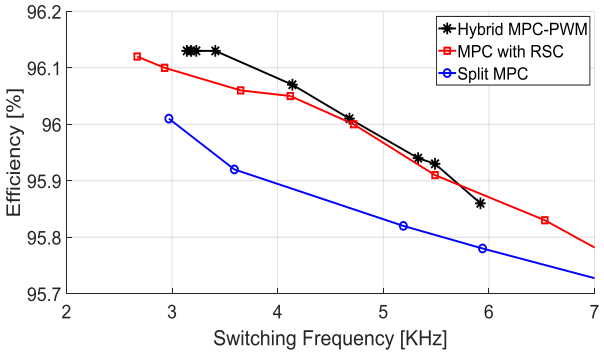


Fig. 11: Comparison of efficiency versus switching frequency for different control methods in the 4-level FCI.

Author Contributions

P. Hamedani carried out the simulation results. M. Changizian interpreted the results. P. Hamedani wrote the manuscript.

Acknowledgment

The authors gratefully acknowledge the respected reviewers and the editor of JECEI for their helpful comments and suggestions.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

EMI	Electromagnetic Interference
FCI	Flying Capacitor Inverter
MLI	Multilevel Inverter
MPC	Model Predictive Control
PWM	Pulse Width Modulation
THD	Total Harmonic Distortion

References

[1] S. Enyedi, "Electric cars—Challenges and trends," in Proc. IEEE 2018 International Conference on Automation, Quality and Testing, Robotics (AQTR): 1-8, 2018.

[2] H. Schefer, L. Fauth, T. H. Kopp, R. Mallwitz, J. Friebe, M. Kurrat, "Discussion on electric power supply systems for all electric aircraft," IEEE Access, 8: 84188-84216, 2020.

[3] C. Jung, "Power up with 800-V systems: The benefits of upgrading voltage power for battery-electric passenger vehicles," IEEE Electrific. Mag., 5(1): 53-58, 2017.

[4] D. Ronanki, A. Kelkar, S. S. Williamson, "Extreme fast charging technology_Prospects to enhance sustainable electric transportation," Energies, 12(19): 3721, 2019.

[5] H. Tu, H. Feng, S. Srdic, S. Lukic, "Extreme fast charging of electric vehicles: A technology overview," IEEE Trans. Transport. Electrific., 5(4): 861-878, 2019.

[6] C. Jung, "Power up with 800-V systems: The benefits of upgrading voltage power for battery-electric passenger vehicles," IEEE Electrific. Mag., 5(1): 53-58, 2017.

[7] A. Poorfakhraei, M. Narimani, A. Emadi, "A review of multilevel inverter topologies in electric vehicles: Current status and future trends," IEEE Open J. Power Electron., 2: 155-170, 2021.

[8] F. Chang, O. Ilina, M. Lienkamp, L. Voss, "Improving the overall efficiency of automotive inverters using a multilevel converter composed of low voltage Si MOSFETs," IEEE Trans. Power Electron., 34(4): 3586-3602, 2019.

[9] M. Quraan, P. Tricoli, S. D'Arco, L. Piegari, "Efficiency assessment of modular multilevel converters for battery electric vehicles," IEEE Trans. Power Electron., 32(3): 2041-2051, 2017.

[10] P. Hamedani, A.Shoulaie, "Utilization of CHB multilevel inverter for harmonic reduction in fuzzy logic controlled multiphase LIM drives," J. Electr. Comput. Eng. Innovations, 8(1): 19-30, 2020.

[11] J. A. Anderson, G. Zulauf, P. Papamanolis, S. Hobi, S. Miric, J. W. Kolar, "Three levels are not enough: Scaling laws for multilevel converters in AC/DC applications," IEEE Trans. on Power Electron., 36(4): 3967-3986, 2021.

[12] A. Poorfakhraei, M. Narimani, A. Emadi, "A review of modulation and control techniques for multilevel inverters in traction applications," IEEE Access, 9: 24187-24204, 2021.

[13] A. M. Trzynadlowski, The field orientation principle in control of induction motors, Springer Science & Business Media, 1994.

[14] P. Vas, Sensorless vector and direct torque control, Oxford University Press, 1998.

[15] N. P. Quang, J. A. Dittrich, Vector control of three-phase AC machines: system development in the Practice, Springer, 2015.

[16] J. Rodriguez, P. Cortes, Predictive control of power converters and electrical drives, John Wiley & Sons, 2012.

[17] J. Rodriguez et al., "Latest advances of model predictive control in electrical drives—Part I: Basic Concepts and Advanced Strategies," IEEE Trans. Power Electron., 37(4): 3927-3942, 2022.

[18] P. Hamedani, S. Sadr, "Model predictive control of linear induction motor drive with end effect consideration," J. Electr. Comput. Eng. Innovations, 11(2): 253-262, 2023.

[19] P. Hamedani, C. Garcia, F. Flores-Bahamonde, S. Sadr, J. Rodriguez, "Predictive control of 4-level flying capacitor inverter for electric car applications," in Proc. the 13th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC): 224-229, 2022.

[20] J. Rodriguez et al., "Latest advances of model predictive control in electrical drives—Part II: Applications and benchmarking with classical control methods," IEEE Trans. Power Electron., 37(5): 5047-5061, 2022.

[21] S. Kouro, P. Cortes, R. Vargas, U. Ammann, J. Rodriguez, "Model predictive control-a simple and powerful method to control power converters," IEEE Trans. Ind. Electron., 56(6): 1826-1838, 2009.

[22] J. Rodriguez, M. P. Kazmierkowski, J. R. Espinoza, P. Zanchetta, H. Abu-Rub, H. A. Young, C. A. Rojas, "State of the art of finite control set model predictive control in power electronics," IEEE Trans. Ind. Inf., 9(2): 1003-1016, 2013.

[23] S. Vazquez, J. Rodriguez, M. Rivera, L. G. Franquelo, M. Norambuena, "Model predictive control for power converters and drives: Advances and trends," IEEE Trans. Ind. Electron., 64(2): 935-947, 2017.

[24] P. Karamanakos, E. Liegmann, T. Geyer, R. Kennel, "Model predictive control of power electronic systems: Methods, results, and challenges," IEEE Open J. Ind. Appl., 1: 95-114, 2020.

[25] J. O. Krah, T. Schmidt, and J. Holtz, "Predictive current control with synchronous optimal pulse patterns," in Proc. IEEE 2nd International Conference on Smart Grid and Renewable Energy (SGRE), 2019.

[26] T. Geyer, G. Papafotiou, M. Morari, "Model predictive direct torque control—part I: Concept, algorithm, and analysis," IEEE Trans. Ind. Electron., 56(6): 1894-1905, 2009.

[27] M. F. Elmorshedy, W. Xu, F. F. M. El-Sousy, M. R. Islam, A. A. Ahmed, "Recent achievements in model predictive control techniques for industrial motor: A comprehensive state-of-the-art," IEEE Access, 9: 58170-58191, 2021.

[28] G. Darivianakis, T. Geyer, W. van der Merwe, "Model predictive current control of modular multilevel converters," in Proc. IEEE Energy Conversion Congress and Exposition (ECCE), 2014.

[29] M. Najjar, M. Shahparasti, R. Heydari, M. Nymand, "Model predictive controllers with capacitor voltage balancing for a single-phase five-level SiC/si based ANPC inverter," IEEE Open J. Power Electron., 2: 202-211, 2021.

[30] J. Raath, T. Mouton, T. Geyer, "Alternative sphere decoding algorithm for long-horizon model predictive control of multi-level inverters," in Proc. IEEE 21st Workshop on Control and Modeling for Power Electronics (COMPEL), 2020.

[31] K. Bandy, P. Stumpf, "Model predictive torque control for multilevel inverter fed induction machines using sorting networks," IEEE Access, 9: 13800-13813, 2021.

[32] M. Wu, H. Tian, Y. W. Li, G. Konstantinou, K. Yang, "A composite selective harmonic elimination model predictive control for seven-level hybrid-clamped inverters with optimal switching patterns," IEEE Trans. Power Electron., 36(1): 274-284, 2021.

[33] M. Aly, F. Carnielutti, M. Norambuena, S. Kouro, J. Rodriguez, "A model predictive control method for common grounded photovoltaic multilevel inverter," in Proc. IEEE IECON 46th Annual Conference of the IEEE Industrial Electronics Society, 2020.

[34] R. O. Ramirez, C. R. Baier, F. Villarroel, J. R. Espinoza, J. Pou, J. Rodriguez, "A hybrid FCS-MPC with low and fixed switching frequency without steady-state error applied to a grid-connected CHB inverter," IEEE Access, 8: 223637-223651, 2020.

[35] K. Antoniewicz, M. Jasinski, M. P. Kazmierkowski, M. Malinowski, "Model predictive control for three-level four-leg flying capacitor converter operating as shunt active power filter," IEEE Trans. Ind. Electron., 63(8): 5255-5262, 2016.

[36] Z. Zhang, C. M. Hackl, R. Kennel, "Computationally efficient DMPC for three-level NPC back-to-back converters in wind turbine systems with PMSG," IEEE Trans. Power Electron., 32(10): 8018-8034, 2017.

[37] J. Ma, W. Song, S. Wang, X. Feng, "Model predictive direct power control for single phase three-level rectifier at low switching frequency," IEEE Trans. Power Electron., 33(2): 1050-1062, 2018.

[38] A. Dekka, B. Wu, V. Yaramasu, N. R. Zargari, "Model predictive control with common-mode voltage injection for modular multilevel converter," IEEE Trans. Power Electron., 32(3): 1767-1778, 2017.

[39] C. Garcia et al., "FCS-MPC based pre-filtering stage for computational efficiency in a flying capacitor converter," IEEE Access, 9: 111039-111049, 2021.

[40] S. R. Mohapatra, V. Agarwal, "A low computational cost model predictive controller for grid connected three phase four wire multilevel inverter," in Proc. IEEE 27th International Symposium on Industrial Electronics (ISIE), 2018.

[41] B. Gutierrez, S. S. Kwak, "Model predictive control method with preselected control options for reduced computational complexity in modular multilevel converters (mmcs)," in Proc. 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe): 1-8, 2018.

[42] R. Amir, A. Hasan, O. Hasan, "Approximate sphere decoding-based model predictive control of cascaded h-bridge inverters," in Proc. IEEE 13th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG): 1-6, 2019.

[43] Y. Zhang, X. Wu, X. Yuan, Y. Wang, P. Dai, "Fast model predictive control for multilevel cascaded h-bridge statcom with polynomial computation time," IEEE Trans. Ind. Electron., 63(8): 5231-5243, 2016.

[44] Y. Zhang, X. Wu, X. Yuan, "A simplified branch and bound approach for model predictive control of multilevel cascaded h-bridge statcom," IEEE Trans. Ind. Electron., 64(10): 7634-7644, 2017.

[45] Z. Ni, M. Narimani, "A new fast formulation of model predictive control for chb statcom," in Proc. IEEE IECON 45th Annual Conference of the IEEE Industrial Electronics Society: 3493-3498, 2019.

[46] Z. Ni, A. Abuelnaga, M. Narimani, "A novel high-performance predictive control formulation for multilevel inverters," IEEE Trans. Power Electron., 35(11): 11533-11543, 2020.

[47] Y. Yang, H. Wen, M. Fan, L. He, M. Xie, R. Chen, M. Norambuena, J. Rodr´ıguez, "Multiple-voltage-vector model predictive control with reduced complexity for multilevel inverters," IEEE Trans. Transport. Electrif., 6(1): 105-117, 2020.

Biographies

Pegah Hamedani was born in Isfahan, Iran, in 1985. She received B.Sc. and M.Sc. degrees from University of Isfahan, Iran, in 2007 and 2009, respectively, and the Ph.D. degree from Iran University of Science and Technology, Tehran, in 2016, all in Electrical Engineering. Her research interests include power electronics, control of electrical motor drives, supply system of the electric railway (AC and DC), linear motors & MAGLEVs, and analysis of overhead contact systems. She is currently an Assistant Professor with the Department of Railway Engineering and Transportation Planning, University of Isfahan, Isfahan, Iran. Dr. Hamedani was the recipient of the IEEE 11th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC'20) best paper award in 2020.

- Email: p.hamedani@eng.ui.ac.ir
- ORCID: 0000-0002-5456-1255
- Web of Science Researcher ID: AAN-2662-2021
- Scopus Author ID: 37118674000
- Homepage: https://engold.ui.ac.ir/~p.hamedani/

Mahmudreza Changizian was born in Khoramabad, Iran, in 1993. He received the B.Sc. degree from the Arak University of Technology, Arak, Iran, in 2015, the M.Sc. degree from the K. N. Toosi University of Technology, Tehran, Iran, in 2017, and the Ph.D. degree from Iran University of Science and Technology (IUST), Tehran, Iran, in 2022, all in the electrical engineering. His research interests include HVdc systems, modeling and control of high-power electronics converter, grid connected converters, motor drive, and renewable energy systems. He is currently an Assistant Professor with the Department of Railway Engineering and Transportation Planning, University of Isfahan, Isfahan, Iran.

- Email: m.changizian@cet.ui.ac.ir
- ORCID: 0000-0002-3296-6788
- Web of Science Researcher ID: GPF-9103-2022
- Scopus Author ID: 57196393409
- Homepage: https://cetold.ui.ac.ir/~m.changizian

How to cite this paper:
P. Hamedani, M. Changizian," A new hybrid predictive-PWM control for flying capacitor multilevel inverter," J. Electr. Comput. Eng. Innovations, 12(2): 353-362, 2024.

DOI: 10.22061/jecei.2024.10549.721

URL: https://jecei.sru.ac.ir/article_2070.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Integration of Clinical, Genetic, and Molecular Features in Predicting Castration Resistance Events in Prostate Cancer: A Comprehensive Machine Learning Analysis

A. Mohammadi [1], ***M. Habibi*** [1], ***F. Parandin*** [2, *]

[1] *Department of Computer Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.*

[2] *Department of Electrical Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.*

Article Info

Article History:
Received 24 November 2023
Reviewed 11 January 2024
Revised 21 February 2024
Accepted 06 March 2024

Keywords:
Metastatic castration-sensitive prostate cancer
Castration resistance
Random forest classifier
Personalized medicine
Machine learning

*Corresponding Author's Email Address: *fa.parandin@iau.ac.ir*

Abstract

Background and Objectives: Metastatic castration-sensitive prostate cancer (mCSPC) represents a critical juncture in the management of prostate cancer, where the accurate prediction of the onset of castration resistance is paramount for guiding treatment decisions.

Methods: In this study, we underscore the power and efficiency of auto-ML models, specifically the Random Forest Classifier, for their low-code, user-friendly nature, making them a practical choice for complex tasks, to develop a predictive model for the occurrence of castration resistance events (CRE) . Utilizing a comprehensive dataset from MSK (Clin Cancer Res 2020), comprising clinical, genetic, and molecular features, we conducted a comprehensive analysis to discern patterns and correlations indicative of castration resistance. A random forest classifier was employed to harness the dataset's intrinsic interactions and construct a robust predictive model.

Results: We used over 18 algorithms to find the best model, and our results showed a significant achievement, with the developed model demonstrating an impressive accuracy of 75% in predicting castration resistance events. Furthermore, the analysis highlights the importance of specific features such as 'Fraction Genome Altered 'and the role of prostate specific antigen (PSA) in castration resistance prediction.

Conclusion: Corroborating these findings, recent studies emphasize the correlation between high 'Fraction Genome Altered' and resistance and the predictive power of elevated PSA levels in castration resistance. This highlights the power of machine learning in improving outcome predictions vital for prostate cancer treatment. This study deepens our insights into metastatic castration-sensitive prostate cancer and provides a practical tool for clinicians to shape treatment strategies and potentially enhance patient results.

Introduction

Prostate cancer accounts for a significant proportion of cancer cases, making up one in every five diagnoses. It is the most common cancer among men and the second leading cause of cancer-related deaths in men in the U.S. Metastatic prostate cancer, a more advanced stage, has shown an increasing incidence despite an overall decline in prostate cancer cases since 2000. Changes in prostate-specific antigen (PSA) screening recommendations in 2008 and 2011 have played a role in this trend. However, metastatic castration-sensitive prostate cancer (mCSPC) cases showed a significant 72% increase in 2013

compared to 2004.

This rise raises concerns because mCSPC is generally considered incurable and has a lower survival rate than localized prostate cancer. Although localized prostate cancer has a 100% 5-year survival rate, mCSPC's prognosis of mCSPC is less favorable, with a 5-year survival rate of 29.8%. Patients with de novo metastases and those whose cancer spreads after being initially diagnosed with localized disease may respond differently to treatment [1]-[5]. However, Prostate cancer is prevalent among men and a significant cause of cancer-related deaths in the Western world [6]. While androgen deprivation therapy (ADT) is commonly used to manage prostate cancer, approximately one-third of patients develop resistance, leading to castration-resistant prostate cancer (CRPC) [7]. Patients typically progress to CRPC within 18-48 months, with metastatic CRPC (mCRPC) being a major contributor to short median survival times [8], [9]. Docetaxel is the primary treatment for mCRPC, and studies have shown that combined treatment with prednisone significantly enhances quality of life and survival [10], [11]. However, despite its benefits, some patients become resistant to docetaxel therapy and discontinue treatment due to adverse events As docetaxel-based chemotherapy remains crucial for managing advanced prostate cancer, the ability to predict early discontinuation based on patient characteristics remains uncertain [12]-[15]. Metastatic castration-sensitive prostate cancer (mCSPC) is a type of prostate cancer that is initially responsive to androgen deprivation therapy (ADT) but eventually progresses to castration-resistant prostate cancer (CRPC) [16]. Castration resistance events (CRE) are defined as the development of CRPC, which is characterized by disease progression despite ADT [17]. CRE is a significant challenge in managing mCSPC, and identifying patients at a high risk of developing CRE is crucial for improving treatment outcomes.

In recent years, several studies have utilized machine learning algorithms, including RFC, to predict CRE in mCSPC. A retrospective study by Pan et al. in 2019 identified patients with rapid progression from hormone-sensitive to CRPC, using machine learning techniques [18]. Another study by Park et al. in 2020 aimed to predict CRPC in patients with prostate cancer using a machine learning approach [19]. Saito et al. constructed a new prognostic prediction model for patients with prostate cancer based on longitudinal data obtained from electronic health records using machine learning techniques [20]. These studies demonstrate the potential of machine learning algorithms, such as RFC, in predicting CRE in mCSPC, which can aid personalized treatment planning and improve patient outcomes. Certainly, our work stands out from previous studies in several ways. In this study, we addressed the prediction of CRE in mCSPC using the RFC algorithm. We extensively utilized diverse data, including clinical, genetic, and molecular features, to decipher specific patterns and correlations indicative of castration resistance. We thoroughly analyzed these patterns using the RFC algorithm and constructed a predictive model for CRE occurrence. This model not only exhibits a significant level of accuracy in predicting CRE but also serves as a practical tool for clinicians to inform treatment strategies and optimize patient outcomes. Our work advances the understanding of the mCSPC and provides a tangible solution for enhancing clinical decision-making in this domain.

The rest of the paper is organized as follows: Section 2 presents the Materials and Methods, detailing the patient background and the process of building the machine learning model. Section 3 discusses the Random Forest Classifier, its performance evaluation, and the impact of hyperparameter tuning. Section 4 delves into the analysis of ROC curves and AUC values in model evaluation. Section 5 highlights the importance of features in the predictive model. Section 6 discusses the findings from recent studies on the role of 'Fraction Genome Altered' and PSA level in mCSPC patients. Section 7 explores the transformation of clinical practice through data-driven approaches for prostate cancer management. Section 8 concludes the paper, summarizing the findings and their implications for the future of medical practice. Finally, Section 9 is devoted to the author contributions and compliance with Ethical Standards.

Materials and Methods

Patient Background

The present study involves an examination of data from a cohort of 424 prostate cancer patients, focusing specifically on individuals with metastatic castration-susceptible disease. The focus was on patients with metastatic castration-susceptible disease at the time of the gene profile sampling. Comprehensive patient information, specimen details, treatment approaches, and clinical outcomes were obtained from a specialized clinical research database. The data were sourced from pathology reports, patient-reported smoking status, and complete medical records. The primary treating physician determined the castration resistance and metastasis, with potential biases carefully scrutinized by the research team. A dedicated research radiologist evaluated disease extent, location, and bone involvement using bone and CT scans. High-volume cases met specific criteria for visceral metastases or at least four bone metastases, whereas others were classified as having low-volume disease. The dataset encompassed a median age of 66 years (interquartile range, 59–72), including patients with both low-volume and high-volume de novo metastatic disease. The samples originated from the bone (14%), liver (4%), lung (6%), lymph node (22%), other soft tissue (6%), and prostate (48%). Castration-resistance events were

recorded in 139 patients with metastatic castration-susceptible disease and 184 patients without castration-resistance events [20]. (Supplementary Table 1).

Predicting Castration Resistance in Cancer Using Machine Learning

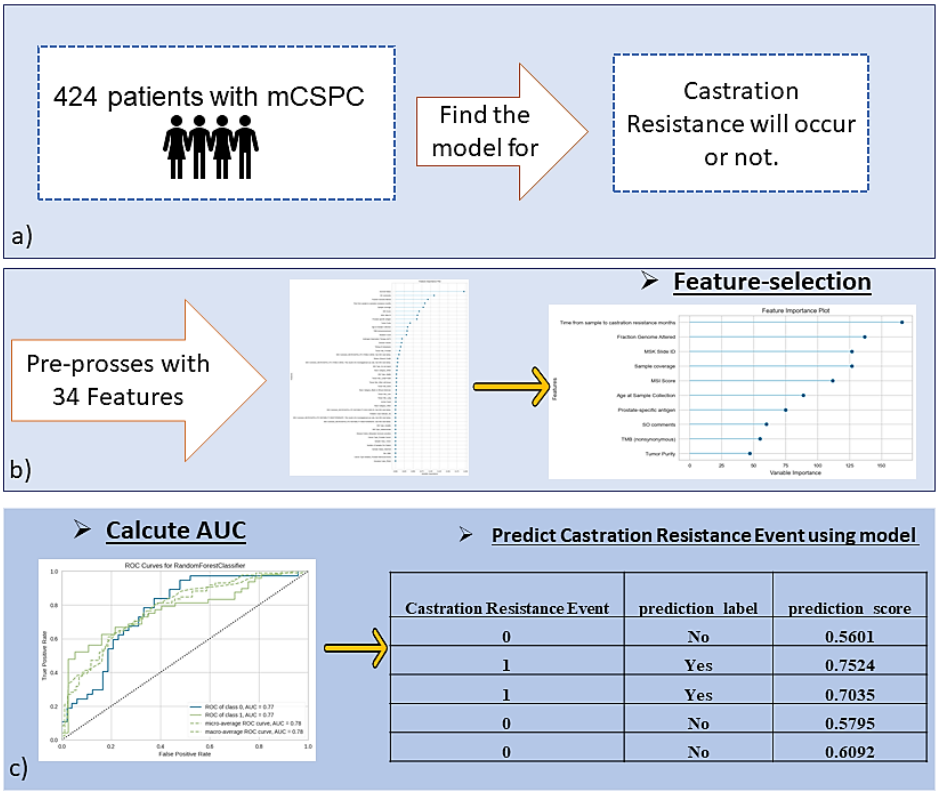
This diagram encapsulates three-phase approach for predicting Castration Resistance in mCSPC patients. Meticulous data collection, preprocessing, model selection, feature importance assessment, and model evaluation together represent a comprehensive journey towards achieving accurate predictions in a clinically significant context (Fig. 1).

In phase **a**), initially, a dataset containing information from 424 patients with mCSPC was collected.

Subsequently, these data were fed into a machine learning model, after which the best model was selected from among more than 30 different models, and a Random Forest classifier proved to be the most effective. Subsequently, during the ensuing phase, the model was fine-tuned using an optimization algorithm. The ultimate goal of this model was to predict the occurrence or absence of Castration Resistance based on 34 different patient features.

In phase **b**), important features are selected from among the various features during this stage

In phase **c**), The final model, configured according to these features, was evaluated. Ultimately, it was able to predict the occurrence of Castration Resistance with an accuracy of 75%, based on the importance of the features.



Castration Resistance Event	prediction_label	prediction_score
0	No	0.5601
1	Yes	0.7524
1	Yes	0.7035
0	No	0.5795
0	No	0.6092

Fig. 1: Visualizing the journey.

From Data Preprocessing to Modeling: Building a Machine Learning Model

In this data preprocessing and modeling journey, a series of distinct stages was employed to prepare the data and construct a machine learning model. These stages encompass various transformations, addressing missing values, converting categorical variables, combining categorical information into target categories, and removing outliers. Commencing with raw data, which comprises unprocessed information, the process unfolds step-by-step. The Label Encoder is first utilized to convert categorical variables into numerical values.

The subsequent steps involve the application of SimpleImputer to handle missing values and iteratively refine results. The OrdinalEncoder and OneHotEncoder were then employed to convert categorical variables into ordinal or binary numerical values. The TargetEncoder stage leverages information from categorical variables as target categories to enhance the model performance. Subsequently, the RemoveOutliers stage eliminates outlier data points to improve the model performance. Finally, CleanColumn Names were employed for appropriate and standardized variable naming. Following this sequential order, the RandomForestClassifier algorithm was employed to create a machine-learning

model tailored to the decision-making task. In this study, we utilized the Pycaret library, a powerful tool for automating machine learning workflows.

Table 1: Evaluation of the performance of various machine learning models

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT(Sec)
Random Forest Classifier	0.7374	0.8093	0.8058	0.7529	0.7738	0.4584	0.4682	1.003
Light Gradient Boosting Machine	0.7374	0.79	0.7692	0.7686	0.7651	0.4656	0.4709	0.689
Gradient Boosting Classifier	0.7348	0.7888	0.7853	0.7574	0.766	0.4571	0.4662	0.834
Extreme Gradient Boosting	0.72	0.7777	0.7282	0.7655	0.7415	0.4342	0.4391	0.804
Extra Trees Classifier	0.693	0.768	0.758	0.727	0.734	0.366	0.379	0.975
Linear Discriminant Analysis	0.6902	0.7537	0.7024	0.7387	0.7164	0.373	0.3775	0.518
Ada Boost Classifier	0.6844	0.7426	0.7384	0.7157	0.7227	0.3533	0.3581	0.768
Decision Tree Classifier	0.6609	0.6538	0.7018	0.6973	0.6953	0.3101	0.3148	0.46
Naive Bayes	0.6283	0.721	0.4658	0.7844	0.5665	0.2894	0.3256	0.92

The Random Forest Classifier

Machine learning algorithms have transcended the boundaries of engineering and optimization, permeating numerous other fields with their transformative capabilities [21]-[23].The Random Forest classifier has significantly affected the field of machine learning, particularly in predictive data mining. This ensemble learning technique uses labeled data samples to classify entities into distinct categories, train a model on one dataset, and evaluate its performance on an independent test dataset. It employs decision trees, which are fundamental tools in supervised learning, and enhances its performance by utilizing random sampling of data to create bootstrap samples and stochastic selection of input features for constructing individual decision trees. The power of the Random Forest technique lies in the combined strength of its decision-tree classifiers and their interaction, which significantly improves the classifier's ability to generalize beyond the training data. Comparative studies show that Random Forest accuracy rivals that of other ensemble techniques such as bagging and boosting. Its advantages include efficient operation on large databases, handling of multiple input variables without variable deletion, unbiased estimation of generalization error, effective handling of missing data, and robustness in maintaining accuracy even with significant data gaps. Additionally, its inherent parallel nature allows seamless implementation with cutting-edge technologies, such as multithreading, multicore processors, and parallel architectures [24]-[26].

This flowchart outlines the comprehensive pipeline for building, training, and evaluating machine learning models using a structured dataset. This process encompasses multiple stages, each contributing to the creation of an effective predictive model.

The objective of this study was to compare and evaluate the performance of various machine learning models for predicting a significant target variable. We used more than 18 algorithms to find the best model, and the best results are presented in the (Table 1). Interesting results were obtained, highlighting the crucial importance of selecting an appropriate model for prediction because of the valuable insights derived from this process. This is particularly vital for the design and implementation of data-driven decision-making systems (Fig. 2).

After carefully analyzing the results, the "Random Forest Classifier" was chosen as the preferred model due to its high accuracy (0.7374), substantial area under the ROC curve (AUC) value (0.8093), and commendable recall (0.8058). This model demonstrated exceptional predictive performance by finding a balance between the accuracy and recall. The F1 score (0.7738) also indicated its ability to effectively harmonize precision and recall.

Moreover, when considering other evaluation metrics such as the Kappa coefficient and Matthews Correlation Coefficient (MCC), the "Random Forest Classifier" showed significant values of 0.4584 and 0.4682, respectively. These results reinforced the selection of the "Random Forest classifier as the top-performing model for predicting the target variable. Furthermore, a comprehensive analysis highlighted that the "Light Gradient Boosting Machine" is an appealing choice. It exhibited a similar accuracy to the first model (0.7374), a reasonable AUC value (0.79), and acceptable recall (0.7692). Its F1 score (0.7651) indicated a well-balanced blend of precision and recall. Among the other models considered, the "Gradient Boosting Classifier" achieved

notable results with an average accuracy (0.7348), satisfactory AUC value (0.7888), and recall (0.7853). This model demonstrated sensible equilibrium between accuracy and recall, as evidenced by its F1 score of 0.766. In conclusion, this analysis underscores the importance of selecting the correct model based on the significance of the evaluation metrics and the specific requirements of the problem at hand.

The "Random Forest Classifier" has been recommended as the top choice in this study for predicting the target variable, primarily due to its superior accuracy and overall performance (Table 1).

The table displays the performance metrics for the different machine-learning models used for classification. The metrics included Accuracy, Area under the Curve (AUC), Recall, Precision, F1 Score, Kappa, Matthews Correlation Coefficient (MCC), and Training Time (TT).

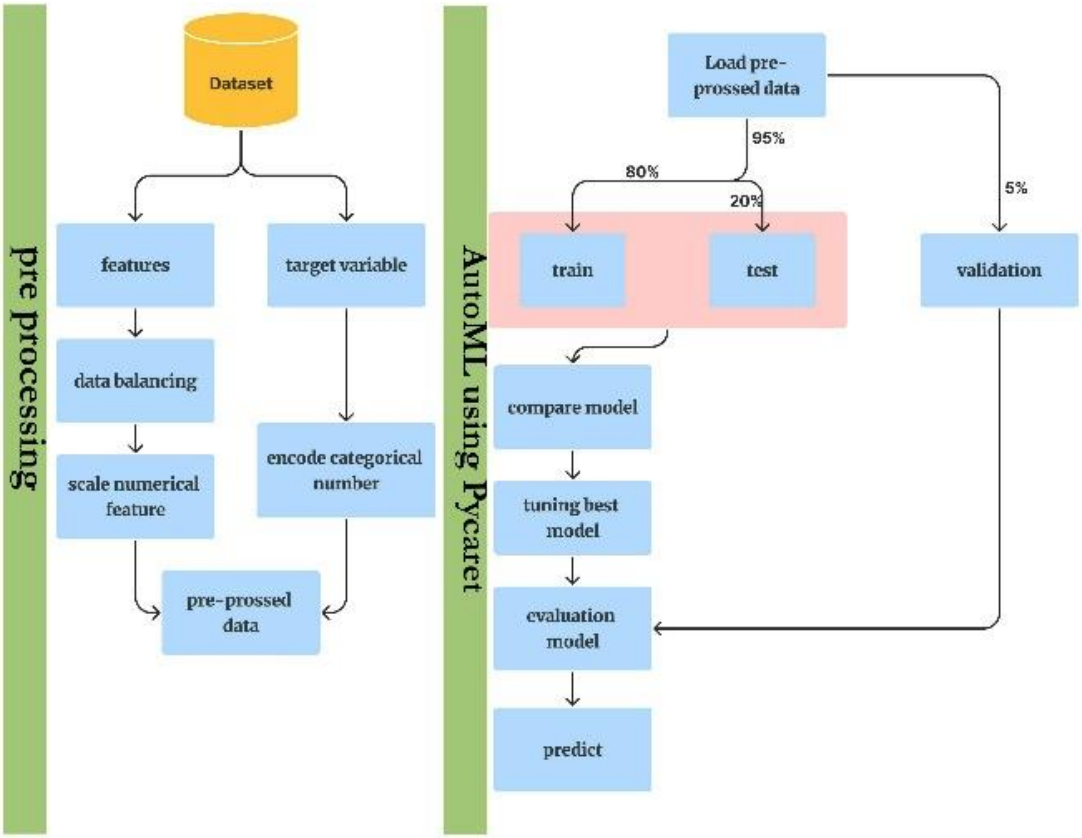


Fig. 2: flowchart of model.

Impact of Hyperparameter Tuning on Random Forest Classifier

Hyperparameter tuning plays a pivotal role in optimizing the performance of machine-learning models, enabling them to reach their full predictive potential. The Random Forest Classifier, a versatile ensemble learning algorithm, demonstrates the substantial influence of hyperparameter tuning on performance [27]-[31] . This study examines the impact of hyperparameter tuning on the performance of the Random Forest Classifier, drawing insights from the provided table that showcases various evaluation metrics (Supplementary Table 2).

The results showed varying performance across different splits. Accuracy ranged from 0.6471 to 0.8209, correctly categorizing the samples.

AUC values ranged from 0.7065 to 0.8757, indicating accurate class differentiation. Recall ranged from 0.6316 to 0.9231, highlighting the model's ability to identify positives. Precision, 0.6744–0.8250, managed the false positives well. F1 score, 0.6857–0.8571; balanced precision and recall. Kappa ranged from 0.2655 to 0.6322, suggesting good agreement, and MCC, 0.2675 to 0.6334, measured the prediction correlation. The average and standard deviation values summarize the performance metrics across the folds. The mean accuracy, AUC, recall, precision, F1 score, Kappa, and MCC were approximately 0.7377, 0.7936, 0.7649, 0.7712, 0.7642, 0.4663, and 0.4722, respectively. The standard deviations indicate the variations. In conclusion, this analysis highlighted the model's predictive probability across splits despite

performance differences. Assessing key metrics revealed the model strengths and weaknesses, aiding real-world decision-making.

Analyzing the ROC Curves and AUC Values in Model Evaluation

Various metrics were used to assess the quality of the classification model. In this section, we analyze the ROC curves and Area Under the Curve (AUC) metric. The classification model in this study was trained to differentiate between two primary categories. The evaluation results (Fig. 3) of the test data are as follows.

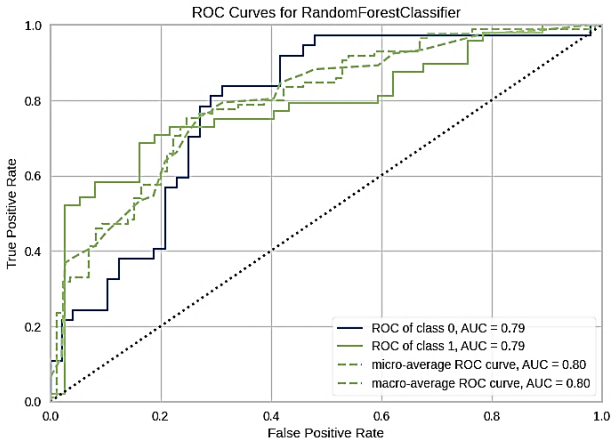


Fig. 3: ROC curves for random forest classifier.

The figure shows ROC curves for two classification models, class 0 and class 1, showing a balanced relationship between the True Positive Rate and False Positive Rate, indicating the effectiveness of the model in distinguishing between classes. A curve above this line indicates superior classification performance and informed decision making.

ROC curves serve as a powerful tool for evaluating a model's ability to distinguish between categories. Figure 4.1 displays the ROC curves for each class. As evident from the curves, the AUC for class 0 was 0.79, and that for class 1 was 0.79. These values indicate the capability of the model to discriminate between the two primary classes.

Micro-average and macro-average metrics were employed to assess the overall model performance. Here, the micro-average AUC calculates the mean of the true positive and false positive rates across all classes, providing a comprehensive assessment of the model's performance. This value was equal to 0.80, signifying the general ability of the model to distinguish between all classes. Based on the evaluation results, we conclude that the classification model exhibits high accuracy and proficiency in identifying different categories. The consistent AUC values across different classes, along with the micro-average and macro-average AUC values of 0.80, demonstrate the stability and reliability of the model's performance against test data.

Random Forest Classifier Performance Evaluation

The Random Forest classifier exhibited a satisfactory performance in the classification task. With an accuracy of 0.7529, the model showed the percentage of correctly classified instances out of the total. The Area Under the Curve (AUC) value of 0.7855 indicated the model's ability to discriminate between positive and negative classes, suggesting a promising overall discriminative capability (Table 2).

Table 2: Evaluating random forest classifier performance

Model	Accuracy	AUC	Recall	Prec	F1	Kappa	MCC
Random Forest	0.75	0.78	0.72	0.81	0.76	0.50	0.50

The table displays the performance metrics for the different machine-learning models used for classification. The metrics included Accuracy, Area under the Curve (AUC), Recall, Precision, F1 Score, Kappa, Matthews Correlation Coefficient (MCC), and Training Time (TT).

The confusion matrix provides a detailed overview of a machine-learning model's predictions versus actual outcomes, enabling a thorough evaluation of its performance. This matrix, with metrics such as accuracy and precision, highlights strengths and weaknesses, thereby guiding improvements for better predictions.
The confusion matrix provides further insight into the performance of the model (Fig. 4).

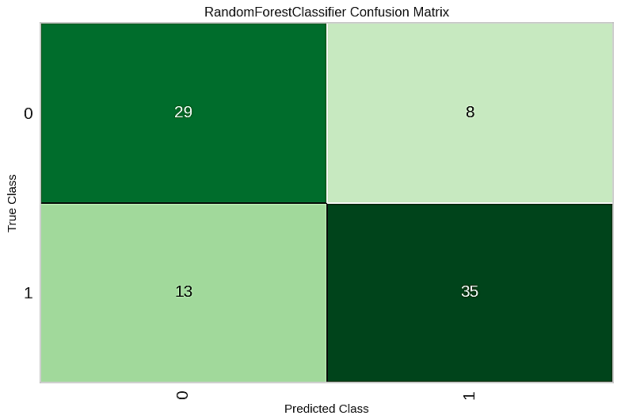


Fig. 4: Random Forest classifier confusion matrix.

The model achieved a recall score of 0.7292, reflecting its sensitivity for correctly identifying positive instances. At the same time, the precision score of 0.8140 highlights its ability to correctly identify positive instances from those that are predicted as positive. This balance between recall and precision is captured by the F1 score of 0.7692, which provides a harmonic mean and demonstrates the balanced performance of the model in handling both true positives and false negatives. A Kappa value of 0.5051 evaluates the model's agreement beyond chance with the actual outcomes, while a Matthews Correlation

Coefficient (MCC) of 0.5087 considers true positive, true negative, false positive, and false negative rates to provide a comprehensive classification measure.

In conclusion, the Random Forest classifier presents commendable accuracy and a balanced trade-off between precision and recall. The AUC value highlights the ability to discriminate between classes. However, further refinement could potentially enhance the predictive capacity and agreement with observed outcomes. This evaluation collectively provides valuable insights into the strengths and areas of the RF classifier for potential improvements in its classification capabilities.

The prediction result has been specified in the (Supplementary predict.csv) file.

In the file, the outcome of the prediction was specified using two elements: 'prediction_score' and 'prediction_label.' The term 'prediction_score' represents a numerical value that indicates the level of confidence or probability associated with the prediction. On the other hand, 'prediction_label' probably represents the assigned category or label that describes the prediction result. Together, these two pieces of information provide insights into the interpretation of the prediction."

Importance Feature

In descending order, the following features were ranked as the most important in our analysis. The significance and role of each of these features in the context of the subject are discussed below (Fig. 5).

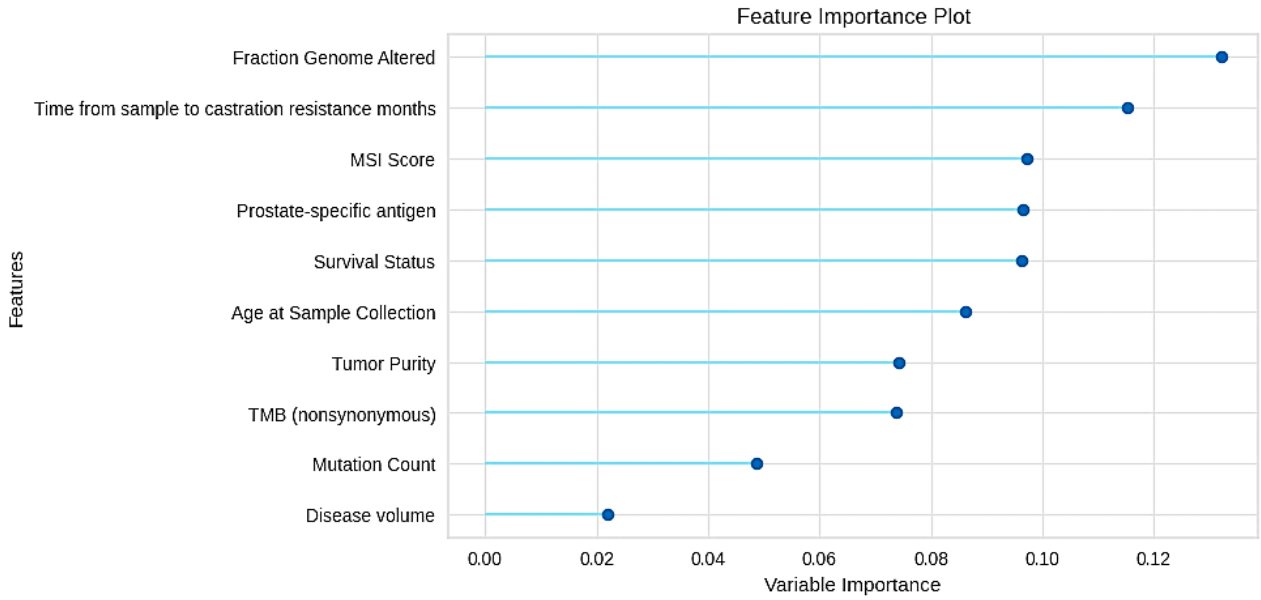


Fig. 5: feature importance plot.

This figure illustrates the results of the feature importance analysis on a tuned Random Forest classifier. This shows the varying impact of different features on the model's predictions, with higher importance scores indicating stronger influences. This insight emphasizes the significance of feature selection and tuning for optimizing the performance of the classifier.

Starting with the list of selected features, the 'Fraction Genome Altered' was introduced as the most critical feature. This feature indicates the proportion of the genome that has experienced alterations, and can serve as a representative of genomic fluctuations within the samples.

Next, the 'Time from Sample to Castration Resistance' is identified as another highly significant feature in this analysis.

The time taken for cancer to progress from the initial treatment to castration resistance offers essential insights into disease advancement and treatment responsiveness.

Following that, the 'MSI Score' is highlighted as a feature of great importance. This score reflects genetic instability in the tumor DNA, and an elevated MSI score might indicate a greater potential for responding to treatment through immunotherapy.

Prostate-specific Antigen (PSA) is deemed highly relevant as a biomarker for prostate cancer. Changes in PSA levels can provide vital information on disease progression.

Survival Status as Age at Sample Collection is another critical factor for predicting survival and understanding the impact of patient age on it. Understanding the

influence of age on survival is crucial for determining disease prognosis.

Subsequently, Tumor Purity was mentioned as a determining feature of tumor cellularity within tissue samples. This metric can impact the accuracy of genetic analyses and treatment decisions.

The 'TMB (Nonsynonymous) Mutation Count' is highlighted as another significant feature in the order of importance. The number of non-synonymous mutations in the tumor genome indicates the extent of new genetic changes, potentially correlating with the immune therapy response.

Following that, the 'Mutation Count' signifies the overall population of mutations within the tumor genome and provides insights into genomic complexity.

This hierarchical analysis of the importance of selected features allows us to better understand the roles and impacts of each feature in predicting disease progression, patient prognosis, and treatment response. These selections can enhance the depth of the analysis and yield more comprehensive insights into the subject matter. All features are shown in (Supplementary Fig 1).

Discussion

Investigating the role of fraction genome altered and PSA level in (mCSPC) patient: Findings from a recent study

Based on the findings of this study (Singla et al., 2021), it can confidently be claimed that there is a strong correlation between the level of 'Fraction Genome Altered' and the occurrence of resistance. This study emphasizes that individuals with high genomic alterations during the initial treatment of prostate cancer are likely to face increased resistance to treatment and advanced castration. A precise analysis of the data from The Cancer Genome Atlas database clearly indicated that individuals with a higher number of altered genome regions were significantly associated with disease progression. This could serve as a robust indicator for predicting resistance and disease progression in prostate cancer. According to these investigations, it can be concluded that 'Fraction Genome Altered' could potentially serve as an indicator to identify individuals at higher risk of resistance to castration in prostate cancer. This unequivocally confirms that the results of this study will aid in enhancing diagnostic accuracy and suitable treatment planning for individuals with this type of cancer [32]. Based on research conducted by Nakanishi, Shotaro et al in 2021, the findings from this study revealed a strong correlation between PSA levels and castration resistance in patients with prostate cancer. This study emphasizes that an increase in PSA levels within three months of initiating treatment can serve as an independent indicator for progression towards castration-resistant prostate cancer and survival in patients diagnosed with metastatic hormone-sensitive prostate cancer. This research has clearly indicated that if the PSA level exceeds one percent of the pre-treatment value within the first three months of treatment, it can effectively predict the progression towards castration resistance and overall survival of patients.

The use of PSA levels as a significant factor in data analysis and treatment prediction for patients with prostate cancer has been substantiated. Furthermore, this study underscores that a more detailed examination of PSA levels within a specific time frame after treatment initiation can be a potent predictive tool for treatment outcomes. This study plays a pivotal role in enhancing diagnostic accuracy and refining treatment strategies for patients with prostate cancer [33].

Transforming clinical practice: Data-driven approaches for prostate cancer management

The study reveals the potential of machine learning algorithms in predicting castration resistance in metastatic castration-sensitive prostate cancer (mCSPC). By identifying predictive markers, the model can guide treatment strategies tailored to each patient's unique profile. The Random Forest Classifier achieved an impressive 75% accuracy in predicting Castration Resistance Event occurrence, demonstrating the potential of machine learning techniques in prostate cancer treatment. The study emphasizes the role of machine learning algorithms in enhancing personalized medicine. The synergy of clinical, genetic, and molecular features within the dataset has revealed patterns and correlations crucial in predicting castration resistance. However, further refinement and optimization are needed to enhance the predictive capacity and accuracy of the model. The interdisciplinary nature of the approach has broader implications for clinical practice, as the convergence of computer engineering, data analysis, and medical expertise could reshape clinical decisions. The study highlights the transformative potential of data-driven approaches in cancer research and personalized medicine, with the integration of advanced algorithms and interdisciplinary collaboration promising improvements in patient lives and shaping the future of medical practice.

Conclusion

In our extensive study aiming to predict Castration Resistance in Prostate Cancer using the Random Forest Classifier, we embarked on a meticulous exploration that involved evaluating over 18 algorithms to unearth the best predictive model. Our exhaustive efforts culminated in a significant achievement, with the developed model showcasing an impressive accuracy rate of 75% in predicting castration resistance events. The model's robust performance metrics, such as the AUC value of 0.7855 and a balanced F1 score of 0.7692, underscore its

competence in accurately categorizing cases while maintaining precision and recall equilibrium. Further bolstering its classification capabilities, the Kappa coefficient (0.5051) and Matthews Correlation Coefficient (0.5087) validate its agreement with real-world outcomes. Additionally, our analysis of feature importance unveiled pivotal variables, with 'Fraction Genome Altered,' 'Time from Sample to Castration Resistance,' and 'MSI Score,' among others, providing crucial insights into disease complexity and patient prognosis. In conjunction with research by Singla et al. in 2021, our findings emphasize the significance of 'Fraction Genome Altered' and 'Prostate-specific Antigen (PSA)' as potent indicators for castration resistance. These insights, derived from our rigorous exploration across various algorithms, have the potential to significantly enhance diagnostic accuracy and refine treatment strategies for individuals confronting metastatic castration-sensitive prostate cancer.

Author Contributions

All authors designed, simulated, carried out the data analysis, collected the data interpreted the results, and wrote the manuscript.

Acknowledgment

This paper was written with the support of the Kermanshah Branch, Islamic Azad University.

Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

Abbreviations

MCSPC	Metastatic Castration-Sensitive Prostate Cancer
CRE	Castration Resistance Event
ADT	Androgen Deprivation Therapy
CRPC	Castration-Resistant Prostate Cancer
PSA	Prostate-Specific Antigen

Data Availability

The data analyzed in this study were obtained from Metastatic Castration-Sensitive Prostate Cancer (MSK, Clin. Cancer Res. 2020) at http://www.cbioportal.org/study/clinicalData?id=prad_mcspc_mskcc_2020

Software and Code

The codes for our model are available at: https://github.com/alirezamohamadiam/Predicting_Castration_Resistance_Event

References

[1] Y. Kim, M. Alhassan, “Analyzing factors enabling prostate cancer screening behaviors among African American males in the south region using the Andersen's behavioral model of healthcare services utilization,” J. Prev., 44(2): 253-266, 2022.

[2] H. Rahbar, P. Karabon, M. Menon, Q. D. Trinh, F. Abdollah, “trends in prostate-specific antigen screening since the implementation of the 2012 US preventive services task force recommendations,” Eur. Urol. Focus., 4(6): 1002-1004, 2017.

[3] D. C. Parker, M. S. Cookson, “The changing landscape in the management of newly diagnosed castration sensitive metastatic prostate cancer,” Investig. Clin. Urol., 61(1): S3-S7, 2020.

[4] W. K. B. Ranasinghe, N. A. Brooks, M. A. Elsheshtawi, J. W. Davis, T. K. Bathala, C. Tang, et al., “Patterns of metastases of prostatic ductal adenocarcinoma,” Cancer, 126(16): 3667-3673, 2020.

[5] W. Ranasinghe, D. D. Shapiro, M. Zhang, T. Bathala, N. Navone, T. C. Thompson, et al. “Optimizing the diagnosis and management of ductal prostate cancer,” Nat. Rev. Urol., 18(6): 337-358, 2021.

[6] E. Gupta, T. Guthrie, W. Tan, “Changing paradigms in management of metastatic castration-resistant prostate cancer(mCRPC),” BMC Urol., 14(55), 2014.

[7] D. Kristiyanto, K. E. Anderson, L. H. Hung, K. Y. Yeung, “Predicting discontinuation of docetaxel treatment for metastatic castration resistant prostate cancer (mCRPC) with randomforest,” F1000Res., 5: 8353, 2016.

[8] M. S. Cookson, W. T. Lowrance, M. H. Murad, A. S. Kibel, “Castration-Resistant prostate cancer: AUA guideline amendment,” J. Urol., 193(2): 491-499, 2015.

[9] A. Heidenreich, P. J. Bastian, J. Bellmunt, M. Bolla, S. Joniau, et al. “EAU guidelines on prostate cancer. Part II: Treatment of advanced, relapsing, and castration-resistant prostate cancer,” Eur. Urol., 65(2): 467-479, 2014.

[10] I. F. Tannock, R. de-Wit, W. R. Berry, J. Horti, A. Pluzanska, et al. “Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer,” N. Engl. J. Med., 351: 1502-1512, 2004.

[11] D. P. Petrylak, “Docetaxel-based chemotherapy trials in androgen-independent prostate cancer: first demonstration of a survival benefit,” Current Oncol. Rep., 7(3): 205-206, 2005.

[12] D. P. Petrylak, C. M. Tangen, M. H. A. Hussain, P. N. Lara Jr, J. A. Jones, M. E. Taplin, P. A. Burch, D. Berry, C. Moinpour, M. Kohli, et al., “Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer,” N. Engl. J. Med., 351: 1513-1520, 2004.

[13] A. J. Templeton, F. E. Vera-Badillo, L. Wang, M. Attalla, P. De Gouveia, R. Leibowitz-Amit, J. J. Knox, M. Moore, S. S. Sridhar, A. M. Joshua, et al. “Translating clinical trials to clinical practice: Outcomes of men with metastatic castration-resistant prostate cancer treated with docetaxel and prednisone in and out of clinical trials,” Ann. Oncol., 24: 2972-2977, 2013.

[14] C. J. Sweeney, Y. H. Chen, M. Carducci, G. Liu, D. F. Jarrard, M. Eisenberger, Y. N. Wong, N. Hahn, M. Kohli, M. M. Cooney, et al., “therapy in metastatic hormone-sensitive prostate cancer,” N. Engl. J. Med., 373: 737-746, 2015.

[15] K. Deng, H. Li, Y. Guan, “Treatment stratification of patients with metastatic castration-resistant prostate cancer by machine learning,” iScience, 23(2): 100804, 2020.

[16] D. Bansal, M. A. Reimers, E. M. Knoche, R. K. Pachynski,” immunotherapy and immunotherapy combinations in metastatic castration-resistant prostate cancer,” Cancers, 13(2): 334, 2021.

[17] L. H. Xiao, P. R. Chen, Z. P. Gou, Y. Z. Li, M. Li, L. C. Xiang, P. Feng, “Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen,” Asian J. Andrology, 19(5):586-590, 2017.

[18] A. R. Hansen, et al., “Pantoprazole Affecting Docetaxel Resistance Pathways via Autophagy (PANDORA): Phase II Trial of high dose pantoprazole (Autophagy Inhibitor) with docetaxel in Metastatic Castration-Resistant Prostate Cancer (mCRPC),” The Oncologist, 24(9): 1188-1194, 2019.

[19] S. Saito, S. Sakamoto, K. Higuchi, et al., “Machine-learning predicts time-series prognosis factors in metastatic prostate cancer patients treated with androgen deprivation therapy,” Sci. Rep., 13: 6325, 2023.

[20] K. H. Stopsack, et al. “Oncogenic genomic alterations, clinical phenotypes, and outcomes in metastatic castration-sensitive prostate cancer,” Clin. Cancer Res., 26(13): 3230-3238, 2020.

[21] F. Parandin, A. Mohamadi, "Designing and optimizing a photonic crystal-based all-optical XOR gate using machine learning,” Majlesi J. Electr. Eng., 18(1): 1-8, 2024.

[22] F. Parandin, M. R. Malmir, “Low delay time all optical NAND, XNOR and OR logic gates based on 2D photonic crystal structure,” J. Electr. Comput. Eng. Innovations, 8(1): 1-8, 2020.

[23] M. Feli, F. Parandin, “A numerical optimization of an efficient double junction InGaN/CIGS solar cell,” J. Electr. Comput. Eng. Innovations, 6(1): 53-58, 2018.

[24] V. Y. Kulkarni, P.K. Sinha, "Random forest classifiers: a survey and future research directions," Int. J. Adv. Comput., 36(1): 1144-1153, 2013.

[25] A. Parmar, R. Katariya, V. Patel, "A review on random forest: An ensemble classifier," in Proc. International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018: 758-763, 2018.

[26] A. B. Shaik, S. Srinivasan, "A brief survey on random forest ensembles in classification model," in Proc. International Conference on Innovative Computing and Communications (ICICC), 2: 253-260, 2018.

[27] P. Probst, M. N. Wright, A. Boulesteix, "Hyperparameters and tuning strategies for random forest," WIREs: Data Min. Knowl. Discovery, 9(3): e1301, 2019.

[28] J. Hancock, T. M. Khoshgoftaar, "Impact of hyperparameter tuning in classifying highly imbalanced big data," in Proc. 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI): 348-354, 2021.

[29] H. L. Le, D. H. Tran, D. V. Chau, "A survey on the impact of hyperparameters on random forest performance using multiple accelerometer datasets," Int. J. Comput. Appl., 30(4): 351-361, 2023.

[30] K. E. Hoque, H. Aljamaan, "Impact of hyperparameter tuning on machine learning models in stock price forecasting," IEEE Access, 9: 163815-163830, 2021.

[31] H. J. P. Weerts, A. C. Mueller, J. Vanschoren, "Importance of tuning hyperparameters of machine learning algorithms," arXiv preprint arXiv:2007.07588, 2020.

[32] R. K. Singla, C. S. Sai, H. Chopra, S. Behzad, H. Bansal, R. Goyal, R. K. Gautam, C. Tsagkaris, S. Joon, S. Singla, B. Shen, “Natural products for the management of castration-resistant prostate cancer: Special focus on nanoparticles based studies,” Front Cell Dev. Biol., 9: 745177, 2021.

[33] S. Nakanishi, M. Goya, et al., “Three-month early change in prostate-specific antigen levels as a predictive marker for overall survival during hormonal therapy for metastatic hormone-sensitive prostate cancer,” BMC Res. Notes, 14: 227, 2021.

Biographies

Alireza Mohamadi received his Bachelor's degree in Computer Engineering from Islamic Azad University, Kermanshah Branch. He is particularly interested in applying Machine Learning to various areas, with a focus on Photonic Gates. He is passionate about exploring the potential of Artificial Intelligence, Machine Learning, and Deep Learning.

- Email: malireza718@gmail.com
- ORCID: 0009-0001-3698-481X
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Maryam Habibi received her B.Sc. in Computer Software Engineering from Razi University in Kermanshah, Iran, in 2004. She then earned an M.Sc. in Artificial Intelligence from the Science and Research Branch of the Islamic Azad University in Tehran, Iran, in 2008. Currently, she is pursuing her Ph.D. at the same university with a primary focus on artificial intelligence. Since 2008, she has been a Lecturer at the Computer Engineering Department, Kermanshah Branch, Islamic Azad University in Kermanshah, Iran. She is a Faculty Member at the Computer Engineering Department, Kermanshah Branch, with research interests spanning artificial intelligence, machine learning, and computer vision.

- Email: ma.habibi@iau.ac.ir
- ORCID: 0000-0001-8939-8622
- Web of Science Researcher ID: NA
- Scopus Author ID: 15078505400
- Homepage: NA

Fariborz Parandin received the B.Sc. and M.Sc. degrees in Electrical Engineering from the University of Razi, Kermanshah, Iran, in 2000 and 2002, respectively. He obtained his Ph.D. degree in Optoelectronics from Razi University in 2017. He joined the Islamic Azad University, Kermanshah Branch, in 2008. He is currently the Associate Professor of Electrical Engineering at Islamic Azad University. His research interests include optoelectronics, semiconductor lasers, photonic crystals and applications of photonic integrated circuits.

- Email: fa.parandin@iau.ac.ir
- ORCID: 0000-0001-9044-3048
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
A. Mohammadi, M. Habibi, F. Parandin, “Integration of clinical, genetic, and molecular features in predicting castration resistance events in prostate cancer: A comprehensive machine learning analysis,” J. Electr. Comput. Eng. Innovations, 12(2): 363-372, 2024.

DOI: 10.22061/jecei.2024.10406.697

URL: https://jecei.sru.ac.ir/article_2088.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

CinfuMax: An Influence Maximization-Based Model for Predicting Cancer Driver Genes in Gene Regulatory Networks

M. Akhavan-Safar [1], ***B. Teimourpour*** [2,*], ***M. Ayoubi*** [3]

[1] *Department of Computer and Information Technology Engineering, Payame Noor University (PNU), Tehran, Iran.*

[2] *Department of Information Technology, Faculty of Industrial and Systems Engineering, Tarbiat Modares University (TMU), Tehran, Iran.*

[3] *Department of Data Science, Tarbiat Modares University (TMU), Tehran, Iran.*

Article Info

Article History:
Received 18 July 2023
Reviewed 13 March 2024
Revised 05 April 2024
Accepted 08 April 2024

Keywords:
Cancer drivers
Influence maximization
Independent cascade model
Influence speed
Regulatory networks

*Corresponding Author's Email Address:
b.teimourpour@modares.ac.ir

Abstract

Background and Objectives: The identification of driver genes, which initiate cancer in cells, holds immense significance within the field of oncology. Discovering these genes is crucial for identifying markers that can indicate specific conditions or diseases, as well as for developing novel systemic and molecular treatment approaches for them. Several computational methods have been developed to identify the genes responsible for cancer based on genomic data. However, many of these methods find key mutations in genomic data to predict which genes are responsible for cancer. These methods rely on mutation and genomic data, but they often exhibit a high rate of false positives in the results. In this study, we propose an influence maximization-based approach, called CinfuMax, which can predict cancer-associated genes without relying on mutation information.

Methods: In this method, the concept of influence maximization and the independent cascade model is employed. Firstly, the gene regulatory network for breast, lung and colon cancers was constructed using regulatory interactions and gene expression data. Next, we implemented an independent cascade diffusion algorithm on the networks to calculate the coverage of each gene. Ultimately, genes with the highest coverage were identified and classified as drivers.

Results: The proposed method's results were compared to those of 19 other computational and network-based methods, utilizing the F-measure and the number of predicted driver genes as evaluation metricsThe results clearly indicate that the proposed method outperforms other methods. Furthermore, CinfuMax successfully identifies 18, 19, and 22 individual driver genes in breast, lung, and colon cancers, respectively, which were not previously identified by any other methods.

Conclusion: Corrected: The results indicate that independent cascading methods for identifying driver genes outperform linear threshold methods. Driver genes were also categorized based on their influence speed, and the genes with the highest diffusion rate in each type of cancer have been identified. The identification of these genes can be valuable for molecular therapies and drug development.

Introduction

Cancer is a disease caused by oncogene activations, such as genetic mutations, chromosome rearrangements, or the deactivation of tumor suppressor genes [1], [2]. It is the second leading cause of death globally, with approximately 6.9 million people losing their lives in 2018.

Lung cancer (2.09 million cases), Breast cancer (2.09 million cases), and Colon cancer (1.80 million cases) are among the most prevalent cancers [3]. During tumor progression, most of the altered genes identified are passenger-type, meaning they do not contribute to the oncogenic process. However, a small portion of the altered genes is believed to be driver genes that disrupt normal transcriptional processes and transform the cell from normal to cancerous.

Cancer Driver Genes prediction Problem

Many computational techniques had been proposed to discover CDGs. In those techniques, it is assumed that the genes that cause cancer are more susceptible to essential modifications in genes, such as mutations. Not all mutations that arise within the cancer genome led to cancer. Therefore, most computational methods attempt to distinguish driver mutations from non-driver mutations. Most of the methods available to identify CDGs depend on transcriptomic or genomic data. For instance, OncodriveCLUST, proposed by Tamboero et al in 2013 [4], identifies genes that exhibit a significant tendency to accumulate mutations in protein sequences. It creates a model to classify genes by evaluating the encoding of silent mutations. Another computational method, Simon [5], aims to improve the identification of cancer driver genes by estimating the background mutation rate and considering the operational effect of mutations on proteins, background mutation changes in tumors, and the redundancy of the genetic code. One of the features of this method is its capability to distinguish between mutations that affect protein function and other mutations. Furthermore, it can differentiate between the number of background mutations in various samples and patients. Dendrix [6] is a computational method that integrates two coverage characteristics: identifying genes in different patient samples and exclusivity. It aims to distinguish driver mutations from passenger mutations that are infrequently observed in certain patients. The ActiveDriver method, developed in 2013 by Reimand et al [7], identifies signal sites where the mutation rate is significantly higher than the mutation level in the entire gene sequence, highlighting the importance of these sites in cancer biology. Another method, e-Driver [8], extracts the internal distribution of malignant mutations between functional regions of proteins to determine the mutation rate relative to other regions of the same protein. If the observations are positive, those genes could be the CDGs. Oncodrive-FM is another computational method based on mutation data [9]. One of the major challenges in cancer genomics lies in identifying Cancer Driver Genes (CDGs) and their associated pathways amidst various types of mutations. The method calculates a functional influence metric using three established methods and evaluates the deviation in the functional influence of variants found in a gene across multiple tumor samples. To address the limitations of traditional approaches, a new criterion termed FM bias is introduced, aiming to overcome issues related to accurately estimating the mutation rate and dependence on incremental changes.

The MDPFinder method [10] uses both mutation data and gene expression data to identify the pathways of cancer mutations and genes that cause cancer. It aims to address the issue of identifying mutant driver paths by developing a maximum weight matrix [6]. To achieve this, it utilizes a genetic algorithm and integrates gene expression data and mutation data to identify cancer mutation pathways and the genes responsible for causing cancer. The DriverML method is another computational method [11] that utilizes machine learning and the Rao test to identify cancer-causing genes, relying on mutation data. The MutsigCV computational method also leverages mutation and expression data [12] to identify abnormal changes in genes and address the issue of heterogeneity in mutation processes and mutation frequency of genes. Additionally, iPAC is a computational method that systematically performs statistical tests on a list of genes to extract the CDGs [13], using gene expression and the number of copies as input data.

Another category of methods for identifying driver genes involves leveraging the structure of biological networks alongside mutation and genomic data. For instance, the Netbox method [14] utilizes protein-protein interaction network analysis to identify frequently changing modules. This method involves creating a network of PPI interactions and signaling pathways, identifying network modules, and statistically evaluating the significance of modularity. DawnRank is another method that utilizes mutation data [15] to focus on each patient's cancer genes, aiming to discover rare and specific driver genes for individual patients. DawnRank utilizes the individual patient's specific genetic information to identify cancer genes. It ranks mutated genes in a patient based on their potential for transmission in the Molecular Interaction Network, with higher-ranking genes being more likely to be drivers. Memo is a systematic approach to identifying cancer modules based on the concept of mutually exclusive events [16], utilizing mutation data and network structure. It searches for modules characterized by three key features: (1) frequent alterations of module genes in tumor samples, (2) involvement in known biological processes, and (3) mutually exclusive change events within the modules. Additionally, Memo conducts mutation enrichment analysis to examine mutational hotspots in genes [17], hypothesizing that genes with such hotspots could act as driver genes. The method integrates diverse data types, including multidimensional disease-related data, biological functional data, and

molecular networks. It employs two approaches: simulating a random walk in sequences to provide a quantitative measure of mutation location and clustering, and evaluating whether a protein domain exhibits a higher mutation rate than the rest of the protein. DriverNet is a computational framework designed to identify driver mutations within miRNA expression networks [18]. While leveraging a network structure, this method also relies on mutation data. It extracts the relationship between genome aberrations and transcription patterns through the gene network's structure.

Another category of methods for recognizing cancer driver genes that has recently gained attention are network-based and bioinformatics methods. These approaches do not rely solely on mutational and genomic data, but also utilize biological network structures to identify CDGs. Notably, the iMaxDriver-N and iMaxDriver-W methods fall within this category [19], as they identify driver genes using gene expression data and the structure of the transcriptional regulatory network, employing the concept of influence maximization and the linear threshold model.

Previously proposed methods for identifying cancer driver genes (CDGs) have limitations. Computational methods rely on mutation data, which inherently contains a significant amount of noise, leading to high false positive results. Moreover, there is significant overlap in the genes identified by these computational methods. While previous network-based methods do not encounter the same issues as computational methods, there is room for improvement in terms of the number of identified CDGs and performance metrics. Addressing these limitations, a network-based method that does not rely on mutational data to identify cancer-causing genes was proposed in this study. This method uses the concept of influence maximization in the transcriptional regulatory network and an independent cascade model to prioritize genes. The data used include gene expression data and human regulatory interactions. In this method, the coverage of each gene is calculated in terms of diffusion power in the gene regulatory network (GRN). The Gene Regulatory Network consists of DNA fragments in a cell that interact indirectly with each other and with other molecular regulators, ultimately determining which genes in the network are transcribed into mRNA. The proposed method is capable of classifying genes based on network propagation speed, which holds significant potential for molecular therapy and drug development purposes.

Influence Maximization Problem

A social network is a social structure consisting of a set of individuals and the relationships or interactions between them. The rapid proliferation of the Internet has significantly increased the popularity of social networking among people. Consequently, social networks have emerged as a popular platform for product advertising and information dissemination [20].

Many topics are explored through the analysis of social networks, including models of diffusion and social influence. The influence maximization problem involves identifying the most influential nodes to achieve the maximum impact of diffusion in a social network, which is known as an NP-Hard problem.

The purpose of influence maximization is to leverage the network's capacity, such as social networks, to reach the widest audience and maximize the spread of information or influence. In general, inputs related to the influence maximizing problem include:

- A directed graph $G = (V, E)$, the network on which influence maximization is to be performed.
- A set of nodes as primary active nodes.
- A function $f : 2^V \rightarrow R$ that maps a set of nodes (S) to their diffusion values $(f(S))$. This shows how much it will affect the amount of diffusion on the network if we choose this set of nodes as seed (S) for the propagation process.
- a budget k

The goal is to find a set of seed(S) that

$$\max f(S) \quad |S| \leq k \tag{1}$$

Influence maximization involves identifying the minimum k nodes that can maximize diffusion in a social network, thereby enabling these nodes to exert the greatest impact on other nodes within the network. There are several models to solve the problem of influence maximization. One widely used method for modeling the diffusion process is the independent cascade model, which draws inspiration from particle movement models in physics [21]. The independent cascade model for influence maximization was initially studied by Goldenberg et al [22]. In this model, a set of nodes is chosen as initially active nodes (seeds).

At each step t, an active node v attempts to activate one of its neighboring inactive nodes with a given probability p_v. If successful, the newly activated nodes become active in the next step $(t + 1)$ and initiate a similar process to activate adjacent inactive nodes. Once a node's activation attempt succeeds or fails, it cannot attempt to activate the same node again. This process continues until it is no longer possible to activate a new node.

A. Independent Cascade Model

Cascading models for diffusion draw inspiration from particle motion and probability theory [23]. These models were initially studied in marketing, with one of the simplest being the independent cascade diffusion model

[24]. The impact of the influence maximization problem largely hinges on the selection of the diffusion model, specifically the diffusion function f. Understanding how a set of nodes (S) affects the entire network remains a significant challenge within the influence maximization problem. The concept involves establishing a random process on the network originating from S and spreading like a contagion. The expected number of infected nodes at the end of the diffusion propagation will indicate the influence f(S) through the set S. Consequently, the influence maximization problem is defined as a random transmission process in the network. The independent cascade model is a subset of cascade propagation models. In this model we have a directed graph G= (V, E). For each edge (u,v) in this graph, the weight $p_{u,v} \in [0,1]$ denotes the probability of diffusion for that edge.

At each time step $t \in N$, each node in the graph is in one of the following three states:

- ***Infected***: The node is newly infected and remains active for a period of time as it tries to infect its neighbors.
- ***Susceptible***: The nodes are not infected yet, but will probably become infected at this time step or in later step.
- ***Inactive***: These nodes were infected in the past but are now inactive. Infected nodes only become infected at a time step and then become inactive.

Fig. 1 shows the process of propagating an independent cascade model for a small network with 10 nodes and three primary active nodes.

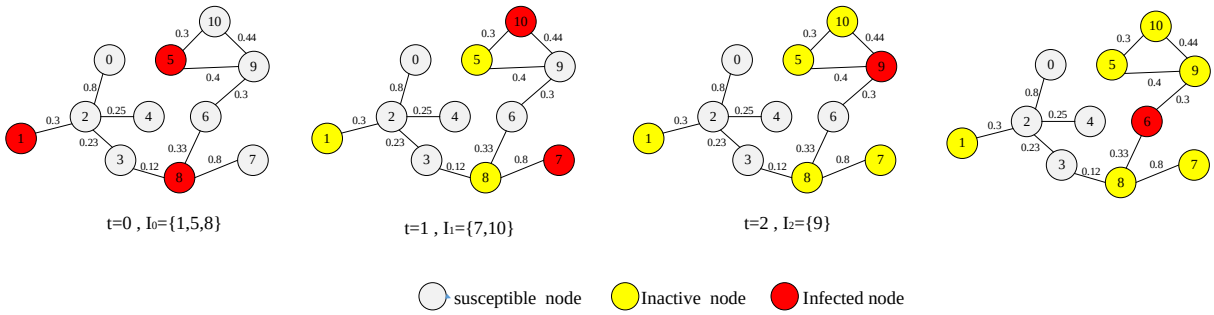


Fig. 1: The IC model example in a network with 10 nodes and 11 edges. Susceptible, infected and inactive nodes at each time step are shown in grey, red and yellow, respectively. Nodes 1, 5 and 7 are the initial active nodes in t=0.

As depicted, each infected node remains active for a time step to infect each of its susceptible neighbors. Infected node u will attempt to infect its neighbor's susceptible node v, with a success probability of $p_{u,v}$. Each attempt to activate susceptible nodes represents an independent random event. In cases where an infected node has multiple susceptible neighbors, attempts are made to activate them in a specified order. This process persists until an infected node successfully infects one of its neighbor's susceptible nodes.

Formally, for a susceptible node v in S_t, its probability of being infected at time step $t + 1$ is given by (2):

$$p(v, I_t) = 1 - \prod_{u \in P(v) \cap I_t} (1 - p_{u,v}) \quad (2)$$

where,

I_t: The infected nodes set

S_t: The susceptible nodes set

$P(v) \cap I_t$: infected parent sets of node v

Also,$P(v)$ are the parents of node v and are defined as (3):

$$P(v) = \{u \in V \mid (u, v) \in E\} \quad (3)$$

For a node $u \in (P(v) \cap I_t)$, the probability that an attempt will unsuccessful is equal to $(1 - p_{u,v})$, and v will be infected if not all attempts unsuccessful. The set of all nodes infected during a contagion process from S can define as (4):

$$I(S) = \bigcup_{t \geq 0} I_t \quad (4)$$

Finally, the penetration function $f(S)$ in the independent cascade model can be defined as (5):

$$f(S) = \mathbb{Z}[|I(S)|] \quad (5)$$

So, in general, the IC model works as follows:

The IC model starts with an initial set of infected nodes (seed). The influence process is revealed in a discrete process according to a random rule:

1. When node n becomes infected in step t, it is given a single chance to infect each currently Susceptible neighbor x; it succeeds with a probability $p(n, x)$
2. If x has several newly Susceptible neighbors, their efforts will be sorted as desired.
3. If n succeeds, x is infected in step t + 1. But whether v succeeds or not, it cannot make further effort to infect w in subsequent rounds.
4. This process runs until no more infection are possible.

Methodology

In this section, we will outline the CinfuMax pipeline, which comprises two distinct steps: the construction of a gene regulatory network and the implementation of the independent cascade algorithm to identify driver genes and determine their influence rate. The driver genes identified through the independent cascade model will be clustered based on their influence rates. Finally, we compare the proposed method with 19 other computational and network-based methodss based on three cancer regulatory networks as benchmarks and several gold standard databases.

A. Gene Regulatory Network

Biological networks represent the numerous of interactions within a cell, providing insight into how relationships between molecules regulate normal cellular behavior. Recent advances in bioinformatics and computational biology have facilitated the study of complex networks of transcriptional regulatory. These networks describe gene expression as a function of regulatory inputs characterized by interactions between proteins and DNA [24]. A gene regulatory network (GRN) is a directional graph in which gene expression regulators bind to target gene nodes through regulatory interactions. Gene expression regulators, including transcription factors (TFs) that can act as activators and repressors, RNA-binding proteins, and RNA regulators, constitute a collection of DNA fragments in a cell [25]. Analyzing and understanding the regulatory relationships between transcriptional regulators and their purposes is essential for comprehending biological phenomena, from cell growth and division to the identification and treatment of diseases, including cancer. Transcriptional regulatory networks (TRNs) are among the most important types of gene regulatory networks, playing a crucial role in the mechanism of diseases, especially cancer [26]. These types of networks are formed from the effect of a type of gene called a transcription factor on other genes. In this study, a list of confirmed regulatory interactions between the transcription factor and genes, described in the next section, was used to construct the cancer transcriptional regulatory networks.

B. Network Reconstruction

To construct transcriptional regulatory networks for each cancer, a list of regulatory interactions and gene expression data was necessary. The list of approved regulatory interactions was downloaded from the RegNetwork database [27], which is freely accessible[1]. This database reports five types of regulatory interactions related to pre-transcription and post-transcription for humans and mice. RegNetwork integrates regulatory interactions collected from various databases and extracts potential regulators based on TFBS[2]. In this study, interactions related to miRNA genes were filtered. Table 1 displays the information about the data used from the RegNetwork database.

Table 1: Characteristics of data taken from the RegNetwork

Element	Description	Number of elements
Gene	Gene regulatory network nodes	21175
interaction	Gene regulatory network edges	150202
TF	Transcription factors (a type network nodes)	1456
Gene	Target genes (a type network nodes)	19719
TF-gene	The TF-gene regulations (a type network edges)	149841
TF-TF	The TF'-TF regulations (a type network edges)	361

Gene expression data was also downloaded from the GEO database [28], which is freely available[3]. This database provides lung (GSE3268)[4], colon (GSE32323)[5], and breast (GSE15852)[6] gene expression data in the .CEL format. Expression data are reported separately in this database for normal tissue and adjacent tumour tissue for each tumour. Prior to use, these files must be processed using the Affy package in R and the RMA method. After processing, synonymous genes were isolated, and duplicate gene values were averaged. three text files corresponding to the three cancer tissues were generated, with each row containing the gene name and its expression values in normal tissue and its adjacent cancer tissue for different patients. The gene regulatory network for each cancer was then constructed by mapping the processed gene expression data to the list of regulatory interactions. This involved retaining interactions where both the source and destination genes were present in the list of related expression data, and filtering out the rest. The number of genes and regulatory interactions in each network is presented in Table 2.

Table 2: Number of nodes and edges in each cancer regulatory network

Network name	Number of nodes	Number of interactions
Breast cancer network	2499	7540
Lung cancer network	2782	8199
Colon cancer network	2500	7540

[1] http://www.regnetworkweb.org
[2] transcription factor binding sites
[3] https://www.ncbi.nlm.nih.gov/geo/
[4] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
[5] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
[6] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi

The CinfuMax Algorithm

Influence maximization models aim to select a minimal seed set that activates the largest number of nodes in the network. Given the time-intensive nature of influence maximization algorithms and the specific type of networks studied (TRNs), we focused solely on nodes of the transcription factor type as being "infected." Each node was individually considered as infected, and the influence score in the network was calculated. The algorithm was executed 300 times on each network, and the resulting coverage values were averaged , serving as the final coverage metric for each gene. The coverage score indicates the potential impact of infecting a gene on the other susceptible genes in the network. Genes with higher coverage are more likely to be associated with cancer driver genes. To optimize the algorithm's iterations for the best outcome, we tested iterations ranging from 10 to 500. In all three cancer networks, the optimal iteration value was obtained to be 300. Moreover, the proposed algorithm can determine the required iterations for each gene to achieve its maximum coverage, enabling the identification of genes with the potential to spread rapidly within the network. An overview of the proposed model is depicted in Fig. 2.

the CinfuMax method takes a cancer regulatory network as input and provides the coverage value and influence rate for each gene as output. Within the independent cascade model, a key parameter is the sensitivity of infected nodes in the network, which is typically set to 0.1 in the basic IC algorithm. To assess sensitivity, we implemented the algorithm using parameter values of 0.1, 0.2, 0.3, 0.4, and 0.5. Across all three networks, the best performance was observed when the parameter was set to 0.4.

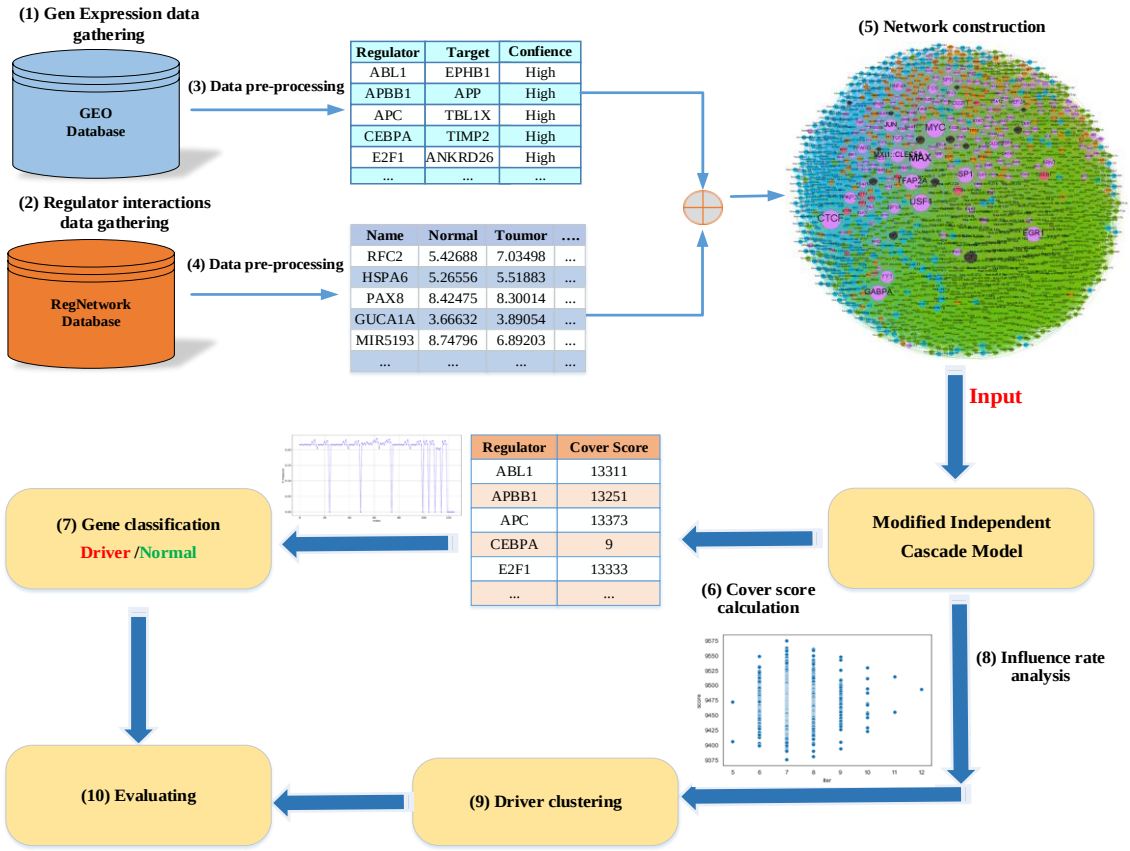


Fig. 2: (1 and 2) The CinfuMax pipeline. (3 and4) Collection of required data (gene expression and regulatory interactions) and pre-processing and data preparation. (5) Constructing transcriptional regulatory network. (6) Running the modified IC algorithm and calculating cover scores (7) Tuning of thresholds and classification of genes (8) Influence rate calculating (9) driver gene clustering (10) Evaluating.

Evaluated Method

CinfuMax results were compared to 19 previous computational and network-based methods. The DriverDBv2 database was used to obtain results on the computational methods [29].

It uses the Cancer Genome Atlas database, such as colon, lung, and breast cancers, as input for computing tools. TCGA is a project aimed at cataloging genetic mutations responsible for cancer through genome sequencing and bioinformatics [30]. It serves as a central

repository for TCGA data and is licensed by the National Cancer Institute Genomic Cancer Center [31]. The results of network-based methods were obtained from their respective papers. The driver genes provided in the Cancer Gene Census (CGC) were used as the gold standard for evaluating CinfuMax and previous methods. The CGC reports a list of cancer driver genes, and we downloaded the lists for colon (TCGA-COAD), lung (TCGA-LUSC), and breast (TCGA-BRCA) cancers from the free TCGA data portal [7]. CGC-approved driver genes were then selected and used as the gold standard of evaluation. In this standard database, 572, 572 and 566 driver genes are reported for breast, colon and lung cancers, respectively.

The confusion matrix was used to calculate the evaluation criteria. Its various values are described in Table 2. To evaluate the performance of the proposed method and compare it with other methods, we used Recall, Precision and F-measure, which are common in binary classification problems.

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Precission = \frac{TP}{TP + FP} \tag{7}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

Confusion matrix was used to calculate evaluation metrics. It shows the prediction results of a classification problem. The confusion matrix and its various values are described in Table 3.

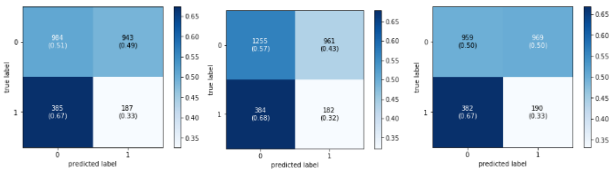
Table 3: The confusion matrix and its various parts

		Actual class	
	Samples	**Positive**	**Negative**
Predicated class	**Positive**	TP (True positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

TP: It refers to the number of genes that have been cancer drivers and the algorithm has also identified them as drivers.
FP: It refers to the number of genes that have not been introduced as cancer drivers in the dataset used, but the algorithm has mistakenly categorized them as cancer drivers.
FN: It refers to the number of genes that have been actually cancer drivers but the algorithm mistakenly categorized them as normal.
TN: It refers to the number of genes that have not been carcinogens and the algorithm has also correctly identified them as non-carcinogens.

Results and Discussion

In this study, cancer regulatory networks were constructed using gene expression data and regulatory interactions. Subsequently, an independent cascade influence algorithm was applied to the network to determine the diffusion score of each gene. To streamline computations and reduce execution time, we focused solely on individual transcription factors as infected initial nodes, based on the network structure. The implementation of the algorithm and its evaluation were carried out using the Python language. The output comprised a list of genes with their respective coverage scores and influence rates, organized in descending order of coverage score. Subsequently, based on a specified threshold value, genes were categorized as either "driver" or "normal." Fine-tuning of the threshold value was performed using the precision_recall_curve and metric packages within the Python sklearn library. The confusion matrix of the proposed method for all three cancer networks is shown in Fig. 3.



(a) (b) (c)

Fig. 3: Confusion matrixs for (a) Breast cancer, (b) Lung cancer (c) and colon cancer networks.

A. *Breast Cancer Network*

The F-measure values for CinfuMax and other computational and network-based methods in Breast cancer are shown in Fig. 4. It is evident that CinfuMax outperforms all other computational and network-based methods in terms of F-measure. Additionally, as illustrated in Fig. 4, CinfuMax has identified 187 drivers, representing the highest number of drivers compared to the previous computational and network-based methods (after iPac). Although iPac has predicted more drivers, its F-measure is notably lower.

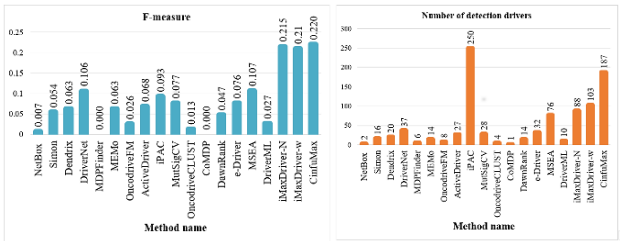


Fig. 4: Performance measures of CinfuMax and other pervious methods proposed in breast cancer.

As shown in Fig. 5, CinfuMax successfully identified 149 genes that were also identified by other methods. Furthermore, CinfuMax discovered 38 unique genes that were not predicted by any previous computational and network-based methods. Additionally, we conducted a comparative analysis of the proposed method in terms of

[7] https://portal.gdc.cancergov

the overlap of predicted drivers with computational and network methods. The Venn diagram in Fig. 6 illustrates that CinfuMax identified 64.9% (124) of the genes identified by other network-based methods, along with 63 unique genes not predicted by any of the network-based methods. Moreover, in comparison to computational methods, CinfuMax identified 93 unique genes that were not predicted by any previous computational methods.

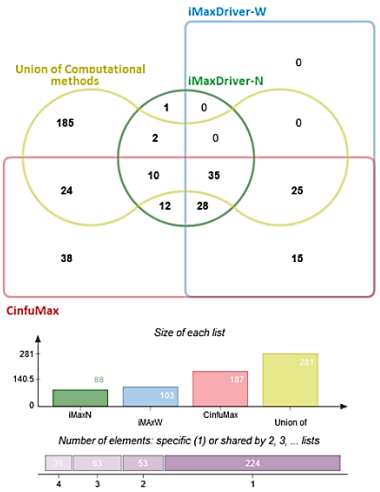


Fig. 5: The Venn diagram for predicted CDGs using CinfuMax and other computational and network-based methods in breast cancer.

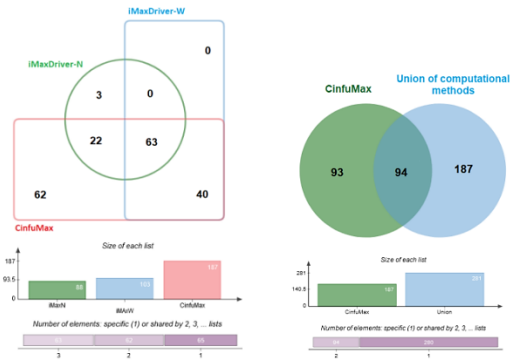


Fig. 6. The Venn diagram for predicted CDGs using CinfuMax and (A). Other network-based methods, (B) the union of all other computational methods in breast cancer

B. Colon Cancer Network

The F-measure values for CinfuMax and other computational and network-based methods in colon cancer network are shown in Fig. 6. It is evident that the proposed method outperforms all computational and network-based methods in terms of F-measure. Additionally, as illustrated in Fig. 7, CinfuMax has identified 190 drivers, representing the highest number of drivers compared to the previous methods (after the iPAC computational method). Although iPac was able to identify 286 drivers, its F-measure is notably low (0.088). We conducted a comparison between CinfuMax and 19 previous methods with respect to the overlap of CDGs predicted. As shown in Fig. 8, CinfuMax successfully identified 158 genes that were also identified by other methods. Additionally, CinfuMax discovered 32 unique genes that were not identified by any previous computational and network-based methods. Furthermore, we evaluated the proposed method in terms of the degree of overlap of predicted drivers separately with computational and network methods. The Venn diagram in Fig. 9 illustrates that CinfuMax identified 61.2% (123) of the genes identified by other network-based methods, along with 67 unique genes not predicted by any of the network-based methods. Moreover, in comparison to computational methods, CinfuMax identified 79 unique genes that were not predicted by any previous computational methods.

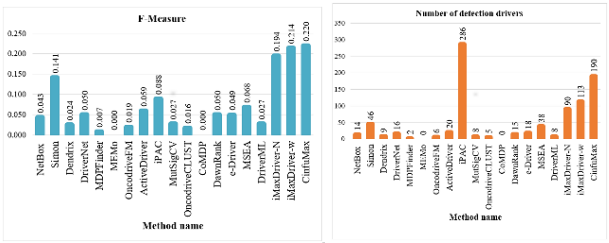


Fig. 7: Performance measures of CinfuMax and other pervious methods proposed in colon cancer.

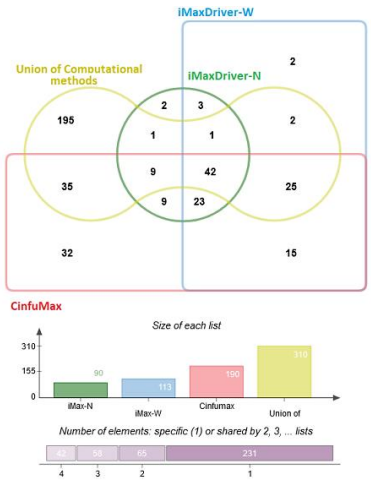


Fig. 8: The Venn diagram for predicted CDGs using CinfuMax and other computational and network-based methods in colon cancer.

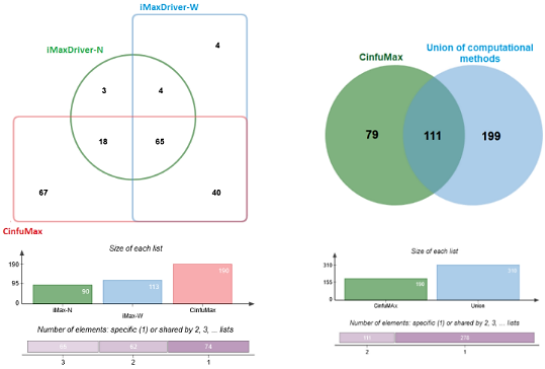


Fig. 9: The Venn diagram for predicted CDGs using CinfuMax and (A). Other network-based methods, (B) the union of all other computational methods in colon cancer.

C. *Lung Cancer Network*

F-measure values for CinfuMax and other computational and network-based methods in Lung cancer are shown in Fig. 10 . It is evident that CinfuMax surpasses all computational methods in terms of F-measure and has the highest value among network methods after iMaxDriver-W. Although its F-measure score is 0.03 less than that of the iMaxDriver-W method, it outperforms significantly in terms of the number of predicted drivers. As depicted in Fig. 10, CinfuMax has identified 182 drivers, representing the highest number of drivers compared to the previous computational and network-based methods.

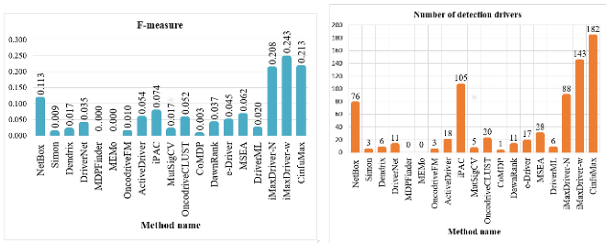


Fig. 10: Performance measures of CinfuMax and other pervious methods proposed in lung cancer.

CinfuMax and other previous methods were compared based on the overlap in predicting driver genes. As shown in Fig. 11, CinfuMax was able to identify 149 genes identified by other methods. In addition, CinfuMax identified 33 unique genes that were not identified by any of the previous computational and network-based methods. Furthermore, we assessed the proposed method's overlap in predicted drivers separately with computational and network methods.

As illustrated in the Venn diagram in Fig. 12, CinfuMax identified 58.44% (135) of genes predicted by other network-based methods and 47 unique genes not predicted by any network-based methods. Additionally, compared to computational methods, CinfuMax identified 130 unique genes not predicted by any previous computational methods. In addition to comparing the proposed method with other previous methods in terms of performance and number of predicted drivers, we also compared it with two previous methods based on linear threshold.

The ROC diagram for the CinfuMax method, which is based on the independent cascade model, and the two methods iMaxDriver-N and iMaxDriver-W, which are based on the linear threshold model, are depicted in Fig. 13.

The results show that in all three cancer networks, the ROC diagrams are almost the same, but the independent cascade model is significantly better than the linear threshold-based methods in terms of the number of predicted drivers as well as the number of unique drivers (Fig. 14).

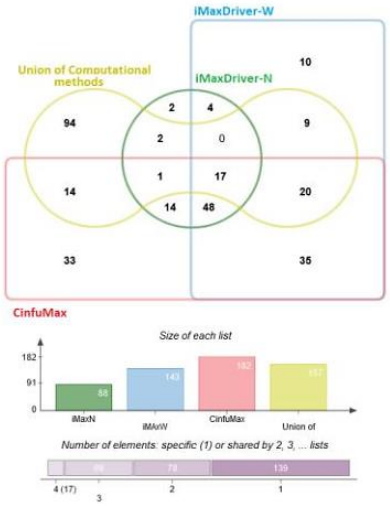


Fig. 11: The Venn diagram for predicted CDGs using CinfuMax and other computational and network-based methods in lung cancer.

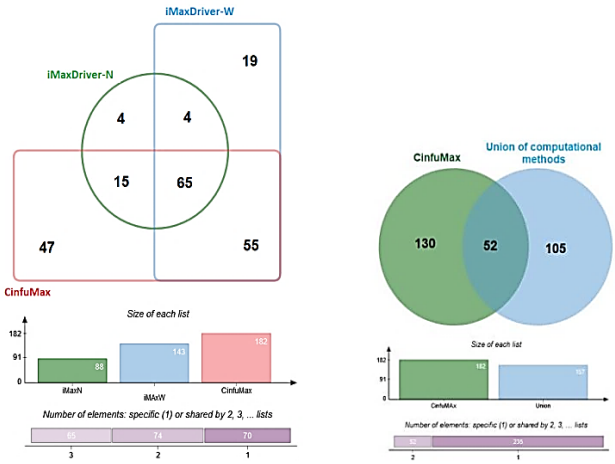


Fig. 12: The Venn diagram for predicted CDGs using CinfuMax and (A). Other network-based methods, (B) the union of all other computational methods in lung cancer.

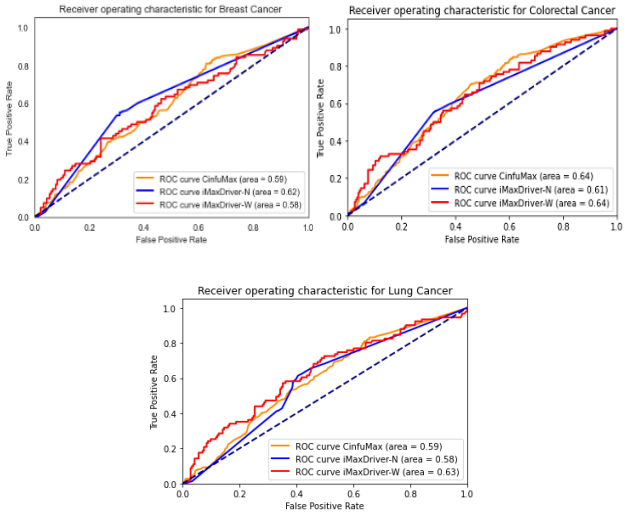


Fig. 13: The ROC diagram for the CinfuMax method and other LT based methods.

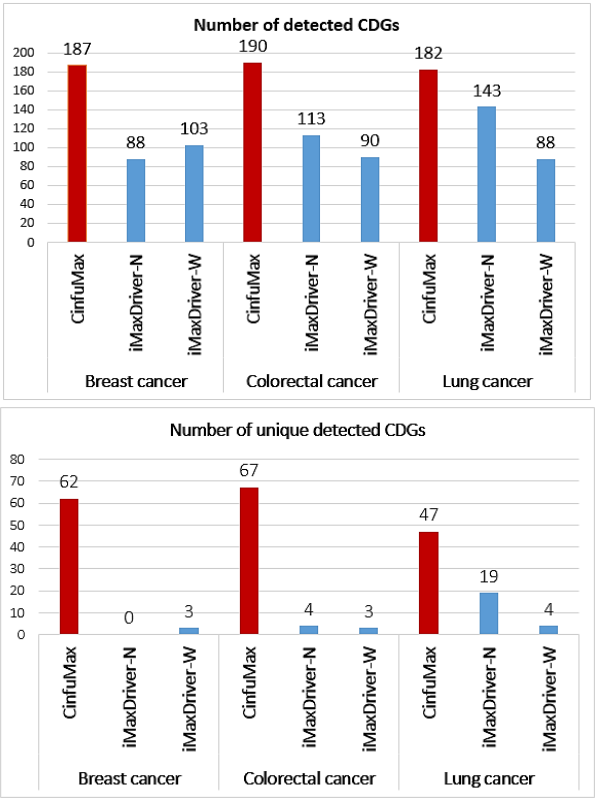


Fig. 14: Compare of the driver genes predicted and unique drivers predicted by CinfuMax and pervious LT based method.

Many researchers working in fields such as bioinformatics and biomathematics at some point face the well-known question of whether results need to be empirically confirmed [32]. Within the field of computational biology, 'experimental validation' refers to the process of replicating a scientific discovery obtained through computational methods by conducting investigations that do not heavily depend on computational resources. This procedure entails gathering additional evidence to bolster the conclusions drawn from the computational study. The integration of orthogonal sets of computational and experimental methods in a scientific study can enhance confidence in its results. However, the term 'experimental validation' may pose a hindrance to this collaborative effort [33]. As similar articles in this field also lack laboratory and experimental confirmation [5], [18], [34]-[36].

Prediction of Genes with The Fastest Speed of Influence

In addition to evaluating the proposed model in terms of the number of diagnostic drivers and comparing its performance, we also identified the driver genes with the highest influence speed compared to other drivers. Influence speed indicates which driver has achieved the highest spread rate earlier in the independent cascade model. This aspect has not yet been explored in diffusion-based methods for identifying cancer driver genes. Identifying driver genes with higher influence speed can be crucial for molecular therapy and targeted drug prescribing. The distribution diagram for iterations in all three tumor tissues is shown in Fig. 15. Nodes with low influence rates were excluded for better visualization. The diagram represents iteration values on the horizontal axis and influence values on the vertical axis, illustrating the nodes separately by iteration.

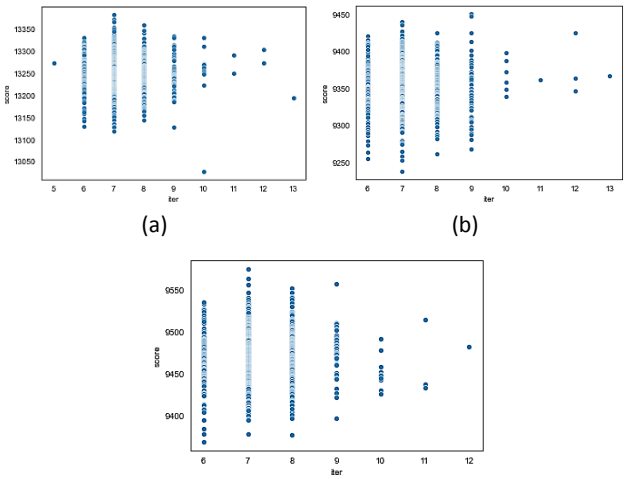


Fig. 15: Distribution diagrams related to the influence speed of each gene according to the algorithm iteration in CinfuMax algorithm. (a) colon cancer, (b) breast cancer, (c) lung cancer.

The cancer driver genes predicted by the CinfuMax method were classified into two categories based on their influence speed rates.

The first category includes genes classified as drivers by the proposed algorithm, which achieve their diffusion score faster than other drivers but do not have the highest diffusion score. These genes are cancer drivers and infect their target genes faster than other drivers, but their infection is less than that of genes with highest diffusion scores. These genes can be therapeutically significant in the early stages of the disease and in preventing the spread of cancer.

The list of these genes in the three cancerous tissues of the breast, lung, and colon is provided in Table 3. For instance, in lung cancer, 18 genes exhibit the highest influence speed in the network, with 17 of these genes not identified by any previous computational or network-based methods. In colon cancer, 15 genes have the highest penetration rate in the network, with 8 of these genes not identified as drivers by any previous computational or network-based methods. In breast cancer, for example, the AFF1 and ZNF384 genes, identified only by the CinfuMax method, are among the driver genes with the highest influence speed. The role of AFF1 in cancer metastasis has been confirmed by [32]. We have also depicted the EGO networks corresponding to

the two unique genes with the highest influence speed in breast cancer in Fig. 16 and Fig. 17. An ego network represents a set of regulatory relations from the perspective of a focal gene. As shown, the AFF1 driver gene, with the highest propagation rate, initiates propagation through only three genes The AFF1 Ego network comprises 1066 nodes and 85905 edges at a depth of 3, demonstrating that the spread speed of abnormality by a gene in the network depends on the structure and general position of that gene in the regulator network, not just its local connections. A similar result is observed for another gene with high influence speed in breast cancer, NF384.

The second category comprises driver genes with the fastest influence speed and the highest diffusion score (top 5%).

These genes are after the first category genes in terms of influence speed, but they have the most infection in the network.

The list of cancer driver genes belonging to the second category is provided in Table 4. These genes can be given special attention in order to prevent metastasis and treat cancer.

Some of these driver genes have not been identified in previous methods and have only been recognized by the proposed method. Information regarding these driver genes in the three cancer networks studied is presented in Table 5.

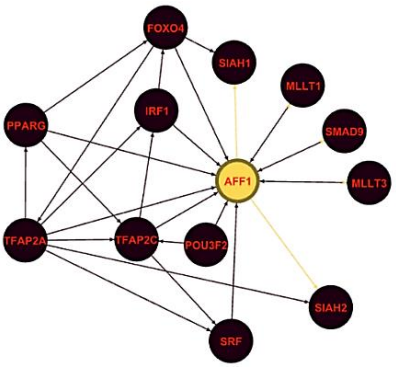


Fig. 16: AFF1 ego network up to 1 depth (13 nodes, 28 edges and Out-degree 5). This network up to depth 3 includes 1066 nodes and 85905 edges).

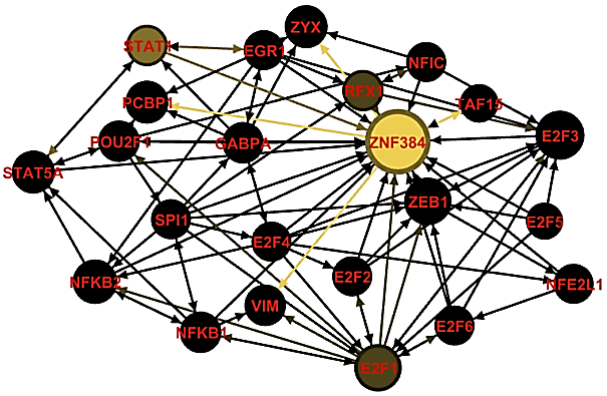


Fig. 17: ZNF384 ego network up to 1 depth (23 nodes, 91 edges and Out-degree 4) This network up to depth 3 includes 10852nodes and 86021edges.

Table 3: List and number of driver genes with the highest influence speed identified by CinfuMax

Tumor tissue	# CDGs	# Unique CDGs (Compare with other methods)	# Unique CDGs (Compare with network-based methods)	# Unique CDGs (Compare with computational methods)	Name of unique CDGs
Colon	15	8	13	8	SRSF3, FUBP1, RAD51B, AFF4, HOXA13, PLAG1, TAL2, THRAP3
Breast	15	5	11	3	HOXC11, ELF4, RAD51B, FUBP1, TNFAIP3
Lung	18	17	17	18	ERCC3, RAD51B, TMF1, HCLS1, HAX1, HMG20B, PDLIM4, ZFP36, FOXP3, SNAI1, HMGN2, CNOT8, GTF3C4, ARNTL2, CTDSPL, BAZ1B, DOT1L,

Table 4: List of genes with the highest diffusion speed and highest influence score identified by CinfuMax

Tumor tissue	# CDGs	# Unique CDGs (Compare with all other methods)	# Unique CDGs (Compare with network-based methods)	# Unique CDGs (Compare with computational methods)	Name of unique CDGs
Colon	38	0	1	12	SMARCB1
Lung	35	0	0	31	-
Breast	42	1	1	17	TAL1

Table 5: Name of genes with the highest diffusion speed and highest influence score identified by CinfuMax

Tumor tissue	Cancer driver genes
Breast	ESR1, FOXO4, FLI1, STAT3, NCOA1, TP53, DAXX, NCOR2, TAL1, RUNX1T1, ERG, KLF6, CTCF, LEF1, NCOR1, RARA, GATA1, MAX, MYOD1, RB1, FOXO1, TCF12, ARNT, HMGA1, MYB, SMAD3, NFE2L2, BCL6, WT1, CREB1, JUN, NFKB2, HMGA2, PML, EP300, SMAD4, ATF1, BRCA1, STAT6, SMARCA4, PAX5, AR
Colon	FOXO1, ABL1, CREB1, SMAD2, SMAD3, SMAD4, MITF, RB1, CTNNB1, ATF1, JUN, REL, PATZ1, CREBBP, ZBTB16, MAF, MYOD1, BRCA1, ESR1, TP53, DDX5, TCF7L2, CTCF, MAX, HNF1A, GATA2, STAT5B, CUX1, ARNT, NCOR2, SMARCB1, LMO2, CDX2, PAX5, STAT3, EP300, PPARG, CEBPA
Lung	EP300, CEBPA, REL, CREB1, HNF1A, FOXO4, PPARG, MYOD1, ABL1, GATA2, JUN, MYB, FOXO1, CREBBP, BCOR, GATA3, STAT5B, RB1, STAT3, CTCF, AR, NFKB2, PML, SMAD2, ZBTB16, RARA, ARNT, LMO2, BCL6, NCOR2, TCF7L2, MYC, SMARCA4, ATF1, MAX, LEF1, GATA1, TP53, CUX1, TAL1, BRCA1, NCOA2, SMARCB1, SMAD4, STAT6, TRRAP, SMAD3, ESR1, CTNNB1

Conclusion and Future Work

A method based on influence maximization was introduced to identify cancer driver genes in human gene regulatory networks, utilizing the independent cascade diffusion model. The independent cascade is one of the popular models in the problem of maximizing influence. This method is also able to classify the identified driver genes based on the influence speed. In this method, regulatory networks associated with breast, lung, and colon cancers are initially constructed using gene expression data and regulatory interactions. Subsequently, the modified independent cascade algorithm is independently executed on each cancer network. Genes are then sorted in descending order of influence scores. Based on a predefined threshold value, the genes are classified into two categories: driver and normal. Moreover, cancer drivers are further categorized into two classes based on their influence speed. The results showed that the CinfuMax has better performance in terms of F-measure and number of predicted drivers than other computational and network-based methods. Additionally, CinfuMax successfully identifies a considerable number of unique drivers that were not previously predicted by other computational and network-based methods. Therefore, it can be utilized as a complementary tool alongside other computational approaches. The proposed method exhibits superior performance compared to iMaxDriver methods, which are based on the linear threshold model. This result shows the use of independent cascade model is more appropriate than linear threshold model in the gene regulatory network to identify driver genes. One of the limitations of influence maximization models is the computational time and the selection of the initial active set (seed). To address this, an effective technique was proposed in this study to reduce execution times. The running time of the proposed algorithm was 35 minutes on a computer equipped with an Intel CORE i7 microprocessor and 8 GB of RAM, which is reasonable for influence maximization algorithms. However, future research could focus on providing methods to further reduce the execution time of the algorithm and ensuring the proper selection of seed nodes. Furthermore, it is noted that driver genes have not been previously classified based on influence speed in diffusion-based methods. Prioritizing genes based on infection and influence speed can be crucial for therapeutic purposes and the development of targeted drugs.

Author Contributions

B. Teimourpour and M. Akhavan-Safar contributed to the study design. M. Akhavan-Safar conducted the writing of the article and data processing. M. Ayyoubi performed data analysis and model implementation. B. Teimourpour critically revised the article and provided final approval of the version to be submitted. All authors have reviewed and approved the final manuscript.

Acknowledgment

The authors would like to thank the appreciate the anonymous reviewers and the editor of JECEI for their useful comments and suggestions.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Data availability

Data is available publicly at:
https://github.com/makhsafar/CinfuMax

Abbreviations

CDG	Cancer Driver Gene
IM	Influence Maximization
IC	Independent Cascade
LT	Linear Threshold

TF	Transcription Factor
TRN	Transcriptional Regulatory Network
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic acid
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative

References

[1] F. Cheng, J. Zhao, Z. Zhao, "Advances in computational approaches for prioritizing driver mutations and significantly mutated genes in cancer genomes," Briefings Bioinf., 17(4): 642-56, 2016.

[2] H. S. Jang, N. M. Shah, A.Y. Du, Z. Z. Dailey, E. C. Pehrsson, P. M. Godoy, D. Zhang, D. Li, X. Xing, S. Kim, D. O'Donnell "Transposable elements drive widespread expression of oncogenes in human cancers," Nat. Genet., 51(4): 611-617, 2019.

[3] World Health Organization, Cancers, 12 September 2018.

[4] D. Tamborero, A. Gonzalez-Perez, N. Lopez-Bigas, "OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes," Bioinf., 29(18): 2238-2244, 2013.

[5] A. Youn, R. Simon, "Identifying cancer driver genes in tumor genome sequencing studies," Bioinf., 27(2): 175-181, 2011.

[6] F. Vandin, E. Upfal, B. J. Raphael, "De novo discovery of mutated driver pathways in cancer," Genome Res., 22(2): 375-385, 2012.

[7] J. Reimand, O. Wagih, G. D. Bader, "The mutational landscape of phosphorylation signaling in cancer," Sci. Rep., 3(1): 2651, 2013.

[8] E. Porta-Pardo, A. Godzik, "e-Driver: a novel method to identify protein regions driving cancer," Bioinf. 30(21): 3109-3014, 2014.

[9] A. Gonzalez-Perez, N. Lopez-Bigas, "Functional impact bias reveals cancer drivers," Nucleic Acids Res., 40(21): e169, 2012.

[10] J. Zhao, S. Zhang, L. Y. Wu, X. S. Zhang, "Efficient methods for identifying mutated driver pathways in cancer," Bioinf., 28(22): 2940-2947, 2012.

[11] Y. Han, J. Yang, X. Qian, W. C. Cheng, S. H. Liu, X. Hua, L. Zhou, Y. Yang, Q. Wu, P. Liu, Y. Lu, "DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies," Nucleic Acids Res., 47(8): e45, 2019.

[12] M. S. Lawrence, P. Stojanov, P. Polak, G. V. Kryukov, K. Cibulskis, A. Sivachenko, S. L. Carter, C. Stewart, C. H. Mermel, S. A. Roberts, A. Kiezun, "Mutational heterogeneity in cancer and the search for new cancer-associated genes," Nature, 499(7457): 214-218, 2018.

[13] M. R. Aure, I. Steinfeld, L. O. Baumbusch, K. Liestøl, D. Lipson, S. Nyberg, B. Naume, K. K. Sahlberg, V. N. Kristensen, A. L. Børresen-Dale, O. C. Lingjærde, "Identifying in-trans process associated genes in breast cancer by integrated analysis of copy number and expression data," PloS one, 8(1): e53014, 2013.

[14] E. Cerami, E. Demir, N. Schultz, B. S. Taylor, C. Sander, "Automated network analysis identifies core pathways in glioblastoma," PloS one, 5(2): e8918, 2010.

[15] H. P. Hou, J. Ma, "DawnRank: discovering personalized driver genes in cancer," Genome Med., 6: 1-6, 2014.

[16] G. Ciriello, E. Cerami, C. Sander, N. Schultz, "Mutual exclusivity analysis identifies oncogenic network modules," Genome Res., 22(2): 398-406, 2012.

[17] D. Arneson, A. Bhattacharya, L. Shu, V. P. Mäkinen, X. Yang, "Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration," BMC Genomics, 17: 1-9, 2016.

[18] A. Bashashati, G. Haffari, J. Ding, G. Ha, K. Lui, J. Rosner, D. G. Huntsman, C. Caldas, S. A. Aparicio, S. P. Shah, "DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer," Genome biol., 13(12): 1-4, 2012.

[19] M. Rahimi, B. Teimourpour, S. A. Marashi, "Cancer driver gene discovery in transcriptional regulatory networks using influence maximization approach," Comput. Biol. Med., 114: 103362, 2019.

[20] T. M. Liggett, Interacting particle systems. New York: Springer; 1985 Feb 13.

[21] J. Goldenberg, B. Libai, E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," Mark. Lett., 12: 211-223, 2001.

[22] W. O. Kermack, A. G. McKendrick, "Contributions to the mathematical theory of epidemics--I. 1927," Bull. Math. Biol., 53(1-2): 33-55, 1991.

[23] D. Kempe, J. Kleinberg, É. Tardos, "Influential nodes in a diffusion model for social networks," in Proc. Automata, Languages and Programming (ICALP 2005): 1127-1138, 2005.

[24] A. Blais, B. D. Dynlacht, "Constructing transcriptional regulatory networks," Gene. Dev., 19(13): 1499, 2005.

[25] C. A. Jackson, D. M. Castro, G. A. Saldi, R. Bonneau, D. Gresham, "Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments," elife, 9: e51254, 2020.

[26] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, N. M. Luscombe, "A census of human transcription factors: function, expression and evolution," Nat. Rev. Genet., 10(4): 252-263, 2009.

[27] Z. P. Liu, C. Wu, H. Miao, H. Wu, "RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse," Database, bav095, 2015.

[28] E. Clough, T. Barrett, "The gene expression omnibus database," Statistical Genomics: Methods and Protocols: 93-110, 2016.

[29] I. F. Chung, C. Y. Chen, S. C. Su, C. Y. Li, K. J. Wu, H. W. Wang, W. C. Cheng, "DriverDBv2: a database for human cancer driver gene research," Nucleic Acids Res., 44(D1): D975-D979, 2016.

[30] K. Tomczak, P. Czerwińska, M. Wiznerowicz, "Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge," Contemp. Oncol., 19(1A): 68-77, 2015.

[31] M. R. Grever, S. A. Schepartz, B. A. Chabner, "The national cancer institute: cancer drug discovery and development program," in Proc. InSeminars in Oncology, 19(6): 622-638, 1992.

[32] Q. X. Meng, K. N. Wang, J. H. Li, H. Zhang, Z. H. Chen, X. J. Zhou, X. C. Cao, P. Wang, Y. Yu, ZNF384–ZEB1 feedback loop regulates breast cancer metastasis," Mol. Med., 28(1): 111, 2022.

[33] M. Jafari, Y. Guan, D. C. Wedge, N. Ansari-Pour, "Re-evaluating experimental validation in the Big Data Era: a conceptual argument," Genome Biol., 22(1): 1-6, 2021.

[34] Y. Lu, Y. Wang, N. Sheng, H. Wang, Y. Fu, Y. Tian, "RDDriver: A novel method based on multi-layer heterogeneous transcriptional regulation network for identifying pancreatic cancer biomarker," in Proc. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM): 497-502, 2022.

[35] F. Dietlein, D. Weghorn, A. Taylor-Weiner, A. Richters, B. Reardon, D. Liu, E. S. Lander, E. M. Van Allen, S. R. Sunyaev, "Identification of cancer driver genes based on nucleotide context," Nat. Genet., 52(2): 208-218, 2020.

[36] R. Gillman, M. A. Field, U. Schmitz, R. Karamatic, L. Hebbard, "Identifying cancer driver genes in individual tumours," Comput. Struct. Biotechnol. J., 21: 5028-5038, 2023.

Biographies

Mostafa Akhavan-Safar is an Assistant Professor of Information Technology Engineering currently at the School of Computer and Information Technology Engineering of Payame Noor University (PNU). He received his M.Sc. in Information Technology Engineering from Iran University of Science and Technology (IUST), and Ph.D. in Information Technology Engineering from Tarbiat Modares University (TMU), Tehran, Iran. His research interests include Bioinformatics, Machine learning, Information systems and Social Network Analysis.

- Email: akhavansaffar@pnu.ac.ir
- ORCID: 0000-0002-7337-712X
- Web of Science Researcher ID: NA
- Scopus Author ID: 57221234257
- Homepage: https://cv.pnu.ac.ir/HomePage/akhavansaffar

Babak Teimourpour is an Associate professor of Information Technology Engineering at the School of Industrial and Systems Engineering of Tarbiat Modares University (TMU). He obtained his Ph.D. in Industrial Engineering from Department of Industrial Engineering, Tarbiat Modares University (TMU), Tehran, Iran. He teaches Ph.D. and M.S. level courses. His research interests include 'Data Mining' and 'Social Network Analysis'. His team won the Iran Data Mining Cup in 2010.

- Email: b.teimourpour@modares.ac.ir
- ORCID: 0000-0002-9286-2286
- Web of Science Researcher ID: NA
- Scopus Author ID: 36115355600
- Homepage: https://www.modares.ac.ir/~b.teimourpour

Mahboube Ayoubi is a master student of data science in Tarbiat modares University, Tehran, Iran. She has a M.Sc. In biostatistics from Isfahan medical University, Isfahan, Iran. She's research interests include Social Network analysis, Machine Learning, Deep Learning and Survival Analysis.

- Email: m.aubi_68@yahoo.com
- ORCID: 0000-0001-6773-7996
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
M. Akhavan-Safar, B. Teimourpour, M. Ayoubi, “CinfuMax: An influence maximization-based model for predicting cancer driver genes in gene regulatory networks,” J. Electr. Comput. Eng. Innovations, 12(2): 373-386, 2024.

DOI: 10.22061/jecei.2024.10026.673

URL: https://jecei.sru.ac.ir/article_2095.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Response Surface Methodology for Behavior Analysis and Performance Improvement of Gravitational Search Algorithm

M. Amoozegar [1,*], *S. Golestani* [2]

[1] *Department of Computer and Information Technology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran.*

[2] *Ph.D. Candidate, Computer Science, University of Saskatchewan, Saskatoon, Canada.*

Article Info

Article History:
Received 26 January 2024
Reviewed 15 February 2024
Revised 30 March 2024
Accepted 08 April 2024

Keywords:
Parameter analysis
Interaction effect
Fine tuning
Response Surface Methodology (RSM)
Gravitational Search Algorithm (GSA)

*Corresponding Author's Email Address:
Amoozegar@kgut.ac.ir

Abstract

Background and Objectives: In recent years, various metaheuristic algorithms have gained popularity due to their effectiveness in solving complex optimization problems across diverse domains. These algorithms are now being utilized for an ever-increasing number of real-world applications. However, two important factors that significantly impact the performance of metaheuristic algorithms are understanding their behavior and fine-tuning their parameters. Deep understanding of an algorithm behavior assists in improving efficiency, while meticulous parameter calibration enhances optimization capability.

Methods: In this study, a response surface methodology-based approach is proposed to analyze the behavior of optimization algorithms. This approach constructs a comprehensive model to determine parameter importance and interaction effects. Although applied to the Gravitational Search Algorithm, this methodology can serve as a generally applicable strategy to gain insights into any metaheuristic algorithm's functionality through quantitative and visual analysis.

Results: Evaluation using 23 test functions exhibited that the technique identifies ideal parameter values and their comparative importance and interplays, enabling superior comprehension.

Conclusion: The proposed framework utilizes informative modeling and multi-faceted analysis to elucidate algorithm mechanics for more targeted calibration, thereby enhancing optimization performance.

Introduction

Metaheuristic approaches help in solving optimization problems; therefore, optimizing their performance is crucial, given their extensive use in a wide array of scientific and engineering problems. Different variants of these approaches have been proposed in recent years [1]-[4]; however, the primary challenge persists in ensuring the efficiency and stability of these algorithms.

The performance of these algorithms is heavily dependent on finding the right set of parameter values. However, navigating the high-dimensional space of possible parameter combinations can be computationally prohibitive. Parameters can be tuned using two different strategies: offline parameter initialization and online parameter tuning. In offline methods, parameters are initialized and fixed before the execution of the algorithm whereas online strategies dynamically and adaptively tune parameters during running time [5], [6]. On the other hand, online approaches are very complicated and time consuming, also their usage is not possible in all scenarios.

A simple method in offline category is the Design of

Experiments (DOE) method that provides a systematic framework for analyzing parameter effects and interactions. It can be very useful for gaining deeper insight into optimization algorithms and improving their performance. DOE establishes a structured approach for determining the relative importance of parameters, unveiling complex interactions, and finding optimal values. By strategically selecting experimental designs, informative models can be constructed to clarify the behavior of algorithms with minimal runs [7].

Two recently proposed methods [8], [9] employed a two-stage algorithm based on DOE to optimize the objective function and maximize the efficiency of the problem. Gunawan and Lau [10] presented an experimental sequential approach to determine the important parameters and fine-tune their values in two stages. The proposed framework relies on design of experiments (DOE) methods, where it employs factorial design in the first phase to screen parameter effects and extract key parameters. In second phase, it estimates the range of these parameters using the surface coverage method.

Response surface methodology (RSM) have gained significant traction across various scientific fields in recent times. [11] applies multi-objective heuristic optimization to generate ensemble classifiers and then utilizes factorial design and response surface methodology to study interactions between algorithm factors and performance, characterizing ensemble reliability. Actually, this paper has focused on the stability analysis of the multi-objective optimization-based classifier. The authors in present [12] an integrated approach combining Taguchi experimental design, response surface modeling, and multi-objective optimization to balance energy and time in machining. Key parameters are identified then surfaces constructed to characterize interactions, enabling navigation of tradeoffs between efficiency and rate through Pareto-based techniques.

GSA (Gravitational Search Algorithm) [13], inspired by the Newtonian gravity and the laws of motion, stands out as a powerful metaheuristic algorithm. This algorithm has shown satisfactory results in solving not only benchmark optimization problems but also in tackling diverse real-world problems. While some variants of this algorithm have been proposed in single objective, binary, and multimodal domains, there remains considerable potential for future exploration.

Particularly, the fine-tuning of the GSA algorithm has not been explored through offline strategies until our previous research [14]. Previously, initial parameter values were chosen based on the recognition of the problem space and trial-and-error mechanisms. In the mentioned study [14], a simple and systematic approach is introduced to fine-tune the parameters of the GSA algorithm using Taguchi method, which is one of existing DOE methods [15], [16]. To the best of our knowledge, there is no publication addressing the parameters' importance and the effects of their interactions. Therefore, contributions and achievements of this research are as follows:

1- Determining the importance and effect of each metaheuristic algorithms' parameter on the outcome using a systematic and analytical approach.
2- Analyzing the interaction effects between parameters.
3- Fine-tuning the parameters as a secondary objective.

Examining the behavior of the GSA can be generalized through testing across various problems. To achieve this, 23 benchmark test functions were carefully selected, and a comprehensive set of quantitative and visual analyses were conducted. The insights gained from these analyses offer valuable guidance for researchers aiming to enhance the performance of the GSA algorithm.

The remaining sections of this paper are structured as follows: Section 2 introduces basic concepts, including GSA and RSM (Response Surface Methodology). Section 3 describes the process of the proposed approach. A comprehensive analysis of 23 test functions and a detailed description of GSA's behavior are presented in section 4. The final section provides conclusions and outlines future directions for research.

Related Work

Previous works can be categorized into two groups. The first category aims at fine-tuning the parameters, while the second category focuses on understanding the parameters and investigating their effects.

In the first category, parameters have been tuned using Design of Experiments (DOE) methods [17]-[20]. Adenso-Diaz and Laguna [21] have presented a framework called CALIBRA, which uses Taguchi method and a local search procedure to find appropriate parameter settings. However, this approach is limited to handling only five parameters and focuses solely on the main effects, neglecting interaction effects between them. Hutter *et al.* [22] employed Sequential Parameter Optimization (SPO) algorithm and DOE techniques to construct a model for optimization problems, exploring the problem space in order to tune the parameters. Akbaripour and Masehian [23] introduced an approach to find the best initial values for optimization algorithms using DOE, Signal to Noise (S/N) ratio, Shannon entropy, and VIKRO methods. This paper considered both the quality of the solution and the running time of algorithm.

Gunawan and Lau [10] proposed a two phase sequential experimental method to determine important parameters and fine-tune them. They used a framework based on DoE, employing the Factorial method to assess

the parameters' importance and extract effective parameters in the first step. In the second step, they used Response Surface Model (RSM) [19] to estimate the promising initial value range for the important parameters. In a recent work by Pereira *et al.* [24], the Lichtenberg algorithm, a metaheuristic algorithm, was tuned and accelerated using RSM methodology and chaos theory. Comparisons on benchmark functions revealed that their proposed chaotic Lichtenberg algorithm achieved superior accuracy, lower cost, and stability compared to genetic algorithms and other bio-inspired metaheuristics. This demonstrates that carefully tuning parameters and integrating acceleration techniques can substantially enhance metaheuristic performance.

The limited research in the second category has specifically explored the effects of parameters using statistical methods. Some of these studies, as a secondary purpose, have delved into the effects of one or two parameters. For example, Kapoor *et al.* [25] has evaluated the effect of the mutation operator of the genetic algorithm (GA) and its importance on both simple and complicated problems. Haines *et al.* [26], proposed an approach to determine the relative importance of GA parameters using Fractional factorial as a DOE method. After running the experiments, the optimal (best) parameter values are provided, and the statistical significance for each parameter is calculated to determine and rank its importance. Arenas et al. [27] has selected four parameters of the genetic considered different values for them, and designed several experiments covering all combinations of parameter values. The proposed approach did not use any DOE methods, making it less suitable, especially when the number of parameters and their values are increased. This paper utilized analysis of variance table (ANOVA) to determine the importance of the selected parameters.

Background

A. Gravitational Search Algorithm (GSA)

One of the most fundamental laws in physics is Newton's Law of Gravity, which states: ''every particle in the universe attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.''. Based on this law, the Gravitational Search Algorithm (GSA), as a meta-heuristic optimization algorithm, is formulated [13]. This algorithm is applicable to many optimization problems [28]-[31], and the obtained results confirm its performance.

Agents are considered as objects, and their fitness is measured based on their masses. According to the force of gravity, all agents attract each other, resulting in a global movement of all objects towards agents with heavier masses. The heavier masses signify the good solutions to the problem.

The GSA algorithm can be described as follows: Consider a system with N masses (agents), in which the position of the i^{th} mass is defined as follows:

$$X_i = (x_i^1, ..., x_i^d, ..., x_i^n) \quad for\ i = 1,2,...,N \tag{1}$$

where x_i^d presents the position of ith agent in the dth dimension where n is dimension of the search space. It's important to note that the positions of the masses correspond to the solutions of the problem. The mass of each agent is calculated after computing the current fitness of the population as follows:

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{2}$$

$$M_i(t) = \frac{q_i(t)}{\sum_{j=1}^{N} q_j(t)} \tag{3}$$

where $fit_i(t)$ represents the fitness value of the agent i in tth iteration and $worst(t)$ and $best(t)$ are defined as follows (for a maximization problem):

$$best(t) = \min_{j\in\{1,..,N\}} fit_j(t) \tag{4}$$

$$worst(t) = \max_{j\in\{1,..,N\}} fit_j(t) \tag{5}$$

At a specific time "t", the force acting on mass "i" from mass "j" is defined as follows:

$$F_{ij}^d(t) = G(t)\frac{M_{pi}(t)\times M_{aj}(t)}{R_{ij}^{rpower}(t) + \varepsilon}(x_j^d(t) - x_i^d(t)) \tag{6}$$

To calculate the distance between ith and jth agent, "Rpower" plays an important role, which was set to one in the original version. To compute the acceleration of an agent, the total forces from a set of heavier masses that apply to it should be considered based on law of gravity (13), followed by the calculation of the agent's acceleration using the law of motion (14).

$$F_i^d(t) = \sum_{j\in Kbest,\, j\neq i} rand_j F_{ij}^d(t) \tag{7}$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \tag{8}$$

Afterward, the next velocity of an agent is calculated as a fraction of its current velocity added to its acceleration (15). Then, its next position could be calculated using (16).

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{9}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (10)$$

where $rand_i$ and $rand_j$ are two uniformly distributed random numbers in the interval [0, 1], and e is a small constant. $G(t)$ is the gravitational constant at time t, which is a decreasing function of time; it will take an initial value G_0, and it will be reduced over time. "kbest" is a function of time, initialized with K_0 at the beginning and is decreased with time, and $R_{ij}(t)$ is the Euclidean distance between two agents i and j:

$$R_{ij}(t) = \left\| X_i(t), X_j(t) \right\|_2 \quad (11)$$

Fig. 1 shows the steps of GSA.

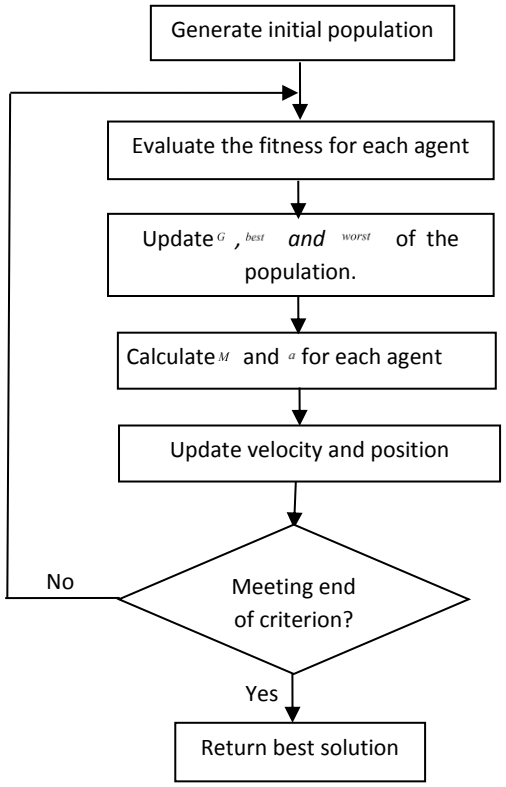


Fig. 1: Flowchart of GSA.

B. Identifying GSA Parameters

The gravitational constant plays a crucial role in maintaining an appropriate balance between exploration and exploitation. Large values of G represent high power attraction between the masses that causes more movement and complexity. Over time, G should be decreased in order to find the optima around a good solution. Therefore, the value of G significantly influences the algorithm's performance and should be carefully controlled. According to (12), G is dependent on G0 and alpha.

$$G(t) = G_0 e^{-\alpha \frac{t}{T}} \quad (12)$$

Kbest is used to control the active agents. In the early iterations, almost all agents attract others to prevent GSA from getting trapped in local optima. This behavior guarantees the exploration in the search process of GSA. In the next iterations, the value of Kbest should be decreased to emphasize on exploitation instead of exploration and improve the convergence rate. It is noteworthy that in the final iteration, only one agent attracts the others.

As mentioned in our previous research [14], a large value for Kbest decreases the convergence rate and provides the algorithm with a chance to explore the search space more thoroughly, preventing it from getting trapped in local optima. Nevertheless, with an increase in Kbest, the computational time and complexity also increase. Therefore, (10) is designed to control the value of Kbest during the progress of the algorithm.

$$Kbest = \beta * \left(\frac{(1-p)}{T} * t + 1 \right), \quad (13)$$

where t is the current iteration, T is the maximum number of iterations, and p is the fraction of total agents in the last iteration. In this equation 0.2<β≤1, whereas in the original version of GSA the value of β is equal to one. When the value of β is 0.5, only half of the agents that are applied in the original GSA will be considered. This reduction decreases the complexity of GSA. Excessively low beta values reduced randomness and exploration causing premature convergence to poor local optima. Excessively high values introduced randomness that hindered convergence, with searches failing to refine and exploit detected promising areas. An intermediate balanced beta enables sufficient diversification to escape local traps yet allows adequate intensification when promising regions are uncovered to drive solutions towards global optimality.

C. Response Surface Methodology (RSM)

In experiments and simulations, many influential factors exist, and their influences should be examined. Moreover, each factor can be initialized with a different variety of values. Therefore, deliberate factor investigation and initialization is very expensive and time consuming. However, it becomes even more complex and challenging when the number of factors and their possible values is increase. In recent years, some automated approaches have been proposed, such as Design of Experiments (DOE), which is an approach for systematizing the process of designing experiments [16].

To Design experiments using DOE, three elements should be addressed: the factors to be tested, the levels of those factors, and the plan of the experiments. Factors represent the parameters of the problem, and levels are different values for them. The plan of the experiments is a series of tests or runs, selected in a way that the effect of different levels of each factor on the response will be

investigated [19].

Factorial design [16] and Taguchi design [32] are two well-known approaches of DOE. The factorial design method tests all possible factor-level combinations, but it becomes inefficient in terms of time and cost when many factors and levels exist in the problem. On the other hand, the Taguchi method designs a small number of experiments, which is not proficient in analyzing the factors' effect.

Response Surface Methodology (RSM) is a model-based approach of DOE. RSM constructs a surface model from the designed experiments to extrapolate previously-unseen regions of the factor space and analyze the relations between input factors and responses [19]. Fig. 2 illustrates the process of RSM.

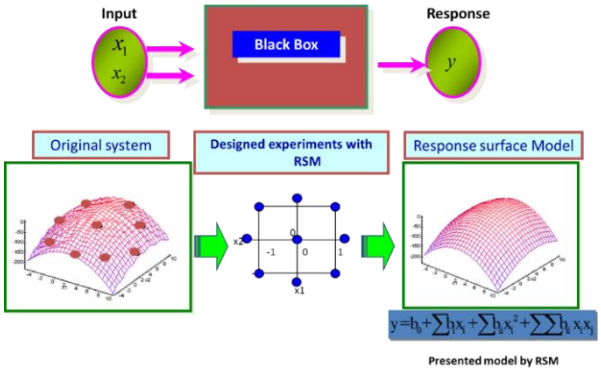


Fig. 2: The RSM method.

Some important advantages of RSM supporting its usage are:

1- Simplifying the problem by providing the surface model.
2- Determining the sensitivity of each factor.
3- Analyzing continuous variables.

Despite these advantages, RSM has a prediction error, as depicted in Fig. 3.

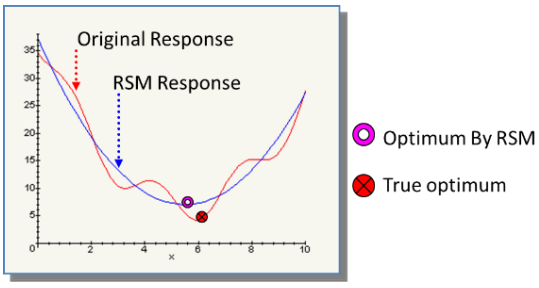


Fig. 3: Original response vs. RSM optimum response.

Two designs based on RSM are presented: central composite design (CCD) and Box-Behnken design (BBD). CCD is routable and popular, providing three to five levels for each factor. In BBD, factors can have only three levels and the provided model is not always routable. After designing and running the experiments, all the results are used to construct the statistical model.

Fig. 4 shows the designed experiments by CCD. There are three factors, x1, x2 and x3, each of them can be initiated with 5 levels, 1, -1, 0, α, -α. CCD includes three groups of experiments:

1- Factorial points: all levels of factors are 1 or -1.
2- Axial points: the value of one factor is α or -α, and other factors are initiated whit 0.
3- Central points: the values of all factors are zero. Additional center points are considered to construct an appropriate surface model.

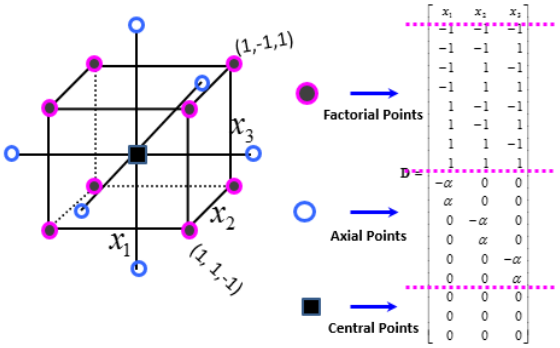


Fig. 4: Designed experiments by CCD.

The current level's values are called coded values, which can be replaced with real values of factors in the problem space. Alpha is distance of each axial point from the center in a central composite design.

RSM is supported by popular statistical tools such as SAS and Minitab. These tools analyze the results and provide plots that are very proficient in determining the impact of each factor on response.

Proposed Method

This section describes the proposed approach, employing Design of Experiments (DOE) to evaluate the parameters' importance of GSA and understand their interaction effects. The objective is to comprehend the behavior of GSA.

Initially, Key parameters of GSA (as factors) and different ranges of their values (as levels) are identified. The parameters we chose to focus on - population size (N), power of the gravitational constant (Rpower), alpha (α), initial gravitational constant (G0), and beta (β) - directly control the major components of how GSA operates. More specifically:

- Population size (N) determines the number of agent solutions in each iteration of the algorithm. This impacts the exploration of the search space, with larger N enabling sampling from more diverse areas.

- Rpower controls the rate of decay of the gravitational constant over iterations. This decay rate affects the exploration-exploitation tradeoff, with higher Rpower providing more thorough exploration initially.

- Alpha (α) controls the sharing of information between agents, with higher alpha meaning that poorer solutions receive stronger gravitational forces from better solutions. This guides the search direction.

- G0 sets the starting gravitational constant, determining the initial search scope and exploration range. Larger G0 leads to wider exploration early on.

- Beta (β) controls the stochastic contribution for moving agents. This introduces randomness to avoid local optima traps.

In this way, N, Rpower, α, G0, and β together determine the exploration-exploitation tradeoff, sharing of information, escape from local optima, and other core components of GSA operation. We selected these parameters specifically because they offer comprehensive control over the search behavior, convergence, and solution quality obtained by the algorithm. Adjusting these parameters provides insight into improving GSA's usage across problem domains.

Subsequently, experiments are designed. The next step involves the selection of an optimization problem and conducting experiments to generate results. The final step includes constructing a model to analyze the parameters' importance and their interactions. All of the abovementioned steps are thoroughly described in the following subsections.

A comprehensive analysis of an algorithm's behavior is only possible when its performance is observed across a diverse set of optimization problems. To achieve this goal, 23 standard functions were meticulously chosen, and the described proposed method was applied to each. The following section provides a summary of GSA's behavior based on the comprehensive analysis conducted.

A. Factor Identification and Level Selection

Studying GSA, independent parameters are extracted and listed in Table 1. Along parameter selection, determining their range value is important. Proper identification of parameter ranges leads to a more comprehensive statistical model and, consequently, better results. Each parameter's value range is chosen and presented in Table 1 based to a literature review, algorithm's behavior, and the role of each parameter. Hereafter, parameters and their value ranges will be referred to as Factor and Level, respectively, following the Design of Experiments (DoE) literature.

Table 1: Selected Factors and corresponding Levels for GSA parameters

Factor name	Level 1	Level 2	Level 3	Level 4	Level 5
N (Population Size)	20	30	40	50	60
Rpower	0.25	0.5	0.75	1	1.25
α (Alpha)	10	20	30	40	50
G0	10	40	70	100	130
β (Beta)	0.2	0.4	0.6	0.8	1

B. Design of Experiments Using RSM

RSM optimization navigates to optimal areas rapidly through sequential experimentation leveraging model predictions. This facilitates optimization with reduced sampling requirements versus. Additionally, RSM visualizations unlock intuitive comprehension of performance drivers, guiding reasoning about parameter tuning. The interpolated response surfaces expose sensitivity not discernible from individual DOE samples. In summary, RSM facilitates optimized enhancement of GSA through revealing performance interactions plus efficiently focusing sampling to navigate towards improved solutions. The combination of modeling, visualization, and sequential experimentation in RSM surpasses fixed sampling techniques like DOE by enabling more insightful navigation of high-dimensional spaces.

Designing appropriate experiments is crucial for creating an accurate analytical model. Considering the mentioned advantages of Response Surface Methodology (RSM) and the objectives of this research, RSM has been chosen to design experiments. Since Central Composite Design (CCD) offers five levels for each factor, it can explore more situations compared to Box-Behnken Design (BBD). Therefore, CCD design has been selected.

Configuring CCD parameters, Alpha is set to 2; therefore, five levels' values will be adjusted to -2, -1, 0, 1, and 2. In this situation, 52 experiments are designed. Having more replicas for each experiment in CCD enhances the model's accuracy, leading to better analysis. Considering the stochastic nature of the gravitational algorithm, 15 replicas for each experiment are taken into account.

Considering the number of experiments and replicas, a total of 780 experiments are designed, which are wisely selected by the model. All Steps were conducted in Minitab and Table 2 represents the first ten of these designed experiments.

Table 2: Initial ten experiments designed by RSM

EXP number	N	Rpower	A	G0	B
1	30	0.75	40	100	0.8
2	40	1	30	70	0.2
3	40	1.5	30	70	0.6
4	40	1	30	70	0.6
5	50	0.75	40	100	0.4
6	50	1.25	20	100	0.8
7	40	1	30	70	0.6
8	30	1.25	40	40	0.4
9	40	1	30	10	0.6
10	40	1	30	70	0.6

C. Selection of Optimization Problems and Experiment Execution

In the initial step, an optimization problem needs to be chosen. In this paper, the Ackley function is chosen from benchmark functions outlined in Appendix 1 of [13]. Denoted as F10, the Ackley function represents one of the multimodal functions and is illustrated in Fig. 5.

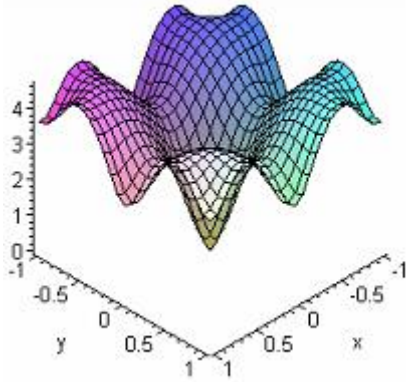


Fig. 5: The Ackley Function plot.

In the next step, parameters of GSA need to be initialized according to Table 2. In addition, the number of iterations of GSA will be set to 500.

During the execution phase, all 780 designed experiments are executed in Matlab, and the obtained results are subsequently imported back into Minitab.

D. Experimental Analysis of the Selected Function

This section outlines the study and analysis of the parameters' effect on the Ackley function.

I) Optimal Parameter Setting

A straightforward analysis involves identifying the optimal values for the parameters, a task facilitated by examining the results of experiments. From the outcomes of the 780 experiments conducted on the Ackley function, it was determined that experiment no. 141 yielded the best (optimal) solution. The parameter values for this particular experiment are presented in Table 3. The initial outcome from applying the proposed method involves fine-tuning the gravitational search algorithm's parameters, resulting in improved performance responses on the test functions.

Table 3: Response Surface Methodology (RSM) results for the Ackley function

		Tuned GSA with RSM	GSA[13]
Result		4.44×10^{-15}	6.9×10^{-6}
Parameters	Population size	50	50
	Rpower	0.75	1
	Alpha	40	20
	G_0	100	100
	Beta	0.8	-
Experiment number		141	-

II) Parameter Importance Analysis

In this paper, a two-way ANOVA is employed to determine the statistical significance of each main effect and interaction effect. The p-value serves as an indicator of whether a factor has a significant impact on the response. If the p-value is less than 0.001, the factor is considered effective; otherwise, it is deemed not effective.

Assessing the effect size of a significant factor is of great importance. In statistics, an effect size serves as a quantitative measure of the strength of a phenomenon [33]. Relative Importance and Absolute Importance measures are employed for this purpose [26]. Absolute importance of a parameter quantifies the difference between the highest and lowest responses achieved through the tuning of that parameter.

Initially, the average of responses ($\bar{y}_{ij}$) resulting from adjusting x_i in level L_i is calculated according to (14). For problem modeling, assume set S consists of all combinations of factors based on the RSM model; therefore, $S_{ij} = \left\{ x \in s \mid x_i = l_j \right\}$.

$$\bar{y}_{ij} = \frac{\sum_{d \in s_{ij}} y_d}{\left| s_{ij} \right|} \tag{14}$$

Absolute Importance and Relative Importance are calculated based on (15) and (16), respectively.

$$E_i = \max\left\{ \bar{y}_{ij} \right\} - \min\left\{ \bar{y}_{ij} \right\} \tag{15}$$

$$RE_i = 100.E_i / \bar{\bar{y}} \tag{16}$$

In these equations, $\bar{\bar{y}}$ is calculated according to (17):

$$\bar{\bar{y}} = \frac{\sum_{d=1}^{n} y_d}{n} \tag{17}$$

Finally, the Absolute Importance, Relative Importance, and p-values of the parameters of the Ackley function are presented in Table 4.

Table 4: Absolute Importance, Relative Importance, and p-values of the Ackley function

Factor	P-value	E_i	RE_i	rate
population size	0.000	3.29	139.35	5
$Rpower$	0.000	14.01	592.97	1
α	0.000	4.34	183.69	4
G_0	0.000	11.08	469.17	2
β	0.000	3.91	165.70	3

In Table 4, the P-value column reveals that all five factors significantly impact the GSA results. A comparison between E_i and RE_i makes it clear that Rpower is the most effective factor, with G0 i closely following as the second most impactful factor with a slight difference. This analysis, applied to problems similar to the Ackley function, helps identify the focal factors that contribute to more effective results.

In addition to these findings, Fig. 6 illustrates the main effect graph of the Ackley function generated in Minitab. The graph reveals the following observations:

- **Effect size of a factor on response.** A factor is more effective if it induces significant changes in the response. For instance, Rpower led to a wide range of responses, establishing it as the most influential factor, while G0 holds the second position.
- **Identifying the optimal value for each parameter and observing the behavior of the function around it.** The optimal value for Rpower is 0.8, and for G0, it is 100, leading to improved responses.
- **The direction of parameters changes.** Rpower and G0 exhibit changes in the same direction, resembling a quadratic curve. Consequently, determining values for these parameters requires more sensitivity. On the other hand, Population and alpha follow linear detrimental and incremental trends, making tuning for these parameters simpler.

III) Analysis of Interaction Effects

Analyzing the interplay between parameters, referred to as the interaction effect, is valuable. When the impact of one factor is contingent on the level of another factor, an interaction plot can provide insights. This graph illustrates possible interactions, which consists of the average of responses. The graph is constructed by fixing the first parameter and plotting the changes in the second parameter against the response. This process is then repeated for different levels of the first parameter. The interaction plot for each pair of parameters and their effects on the response of the Ackley function is illustrated in Fig. 7.

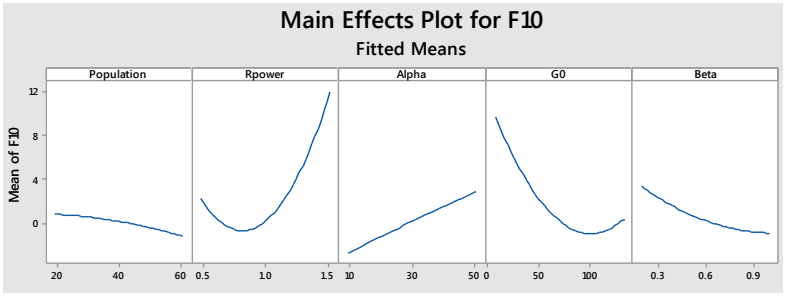


Fig. 6: main effect plot for the Ackley function.

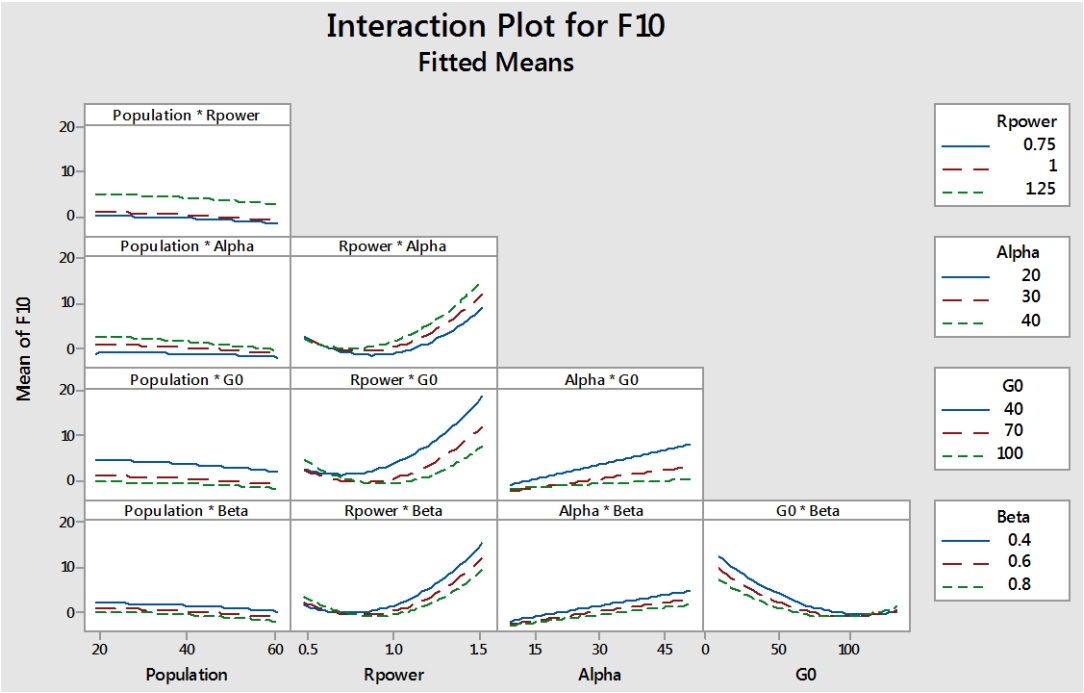


Fig. 7: Interaction plot of each parameter's pair and their effect on the Ackley function.

Deviation from parallel mode and the extent of this deviation respectively indicate interaction effects and the degree of this interaction. The second column of Fig. 7 indicates that Rpower exhibits interaction effects with all other factors, although the magnitude of this interaction is not significant. Additionally, the G0 and Beta pair (G0*Beta) also shows a slight interaction at their ends. The Population factor shows no sign of interaction with other factors.

Moreover, the P-value measure of paired factors is utilized to determine whether the interaction effects are significant or not. This P-value is extracted from ANOVA, confirming the analysis of the aforementioned plot. For instance, ANOVA reported the interaction between Rpower and Alpha (Rpower*Alpha) is significant at the 0.001 level.

Other types of plots, specifically contour and surface plots, prove valuable in analyzing parameters' interactions. Fig. 8 and Fig. 9 illustrate these plots for the Rpower and G0 parameter pair in the context of the Ackley function. According to these plots, the worst response is obtained when Rpower is set to the maximum and G0 is set to the minimum. The best response occurs under conditions where Rpower is low and G0 is high.

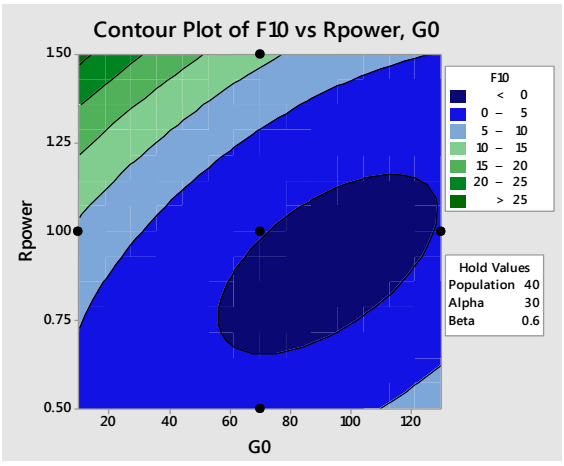


Fig. 8: Counter plot of G0 and Rpower parameter pair for the Ackley function.

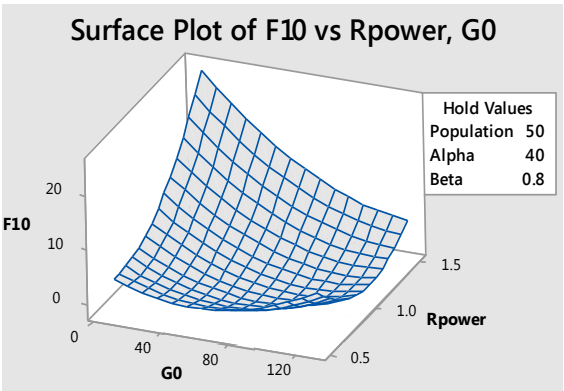


Fig. 9: Surface plot of G0 and Rpower parameter pair for the Ackley function.

Analyzing the Behavior of Gravitational Search Algorithm (GSA)

As previously mentioned, to enhance our understanding of the behavior of GSA, we will examine its performance across 23 standard test functions detailed in Appendix 1 of [13]. These benchmark objective functions serve as the performance evaluation suite for analyzing the gravitational search algorithm. These specific functions were chosen because they were originally introduced and utilized by the authors who developed the gravitational search algorithm (GSA) in their seminal paper [13], [34]. Additionally, these benchmark functions offer a diverse set of complex multi-modal landscapes possessing varied characteristics. These benchmark functions have been widely adopted to characterize, validate, and compare the performance of GSA across subsequent research. All these functions are selected from minimization problems, where n represents the number of dimensions, f_{opt} is the minimum response, and S is a subset of $\mathbb{R}^n$.

These test functions are categorized into three main categories. The first category comprises the unimodal test functions (F1 through F7), each featuring a single local and global optimum. The second category encompasses high-dimensional multimodal functions, F8 through F13, while the third category consists of low-dimensional multimodal functions, F14 through F23. Multimodal functions exhibit multiple local optima but only one global optimum. As the number of dimensions increases, the number of local optima solutions also rises. Successfully navigating through local optima to reach the global optimum is crucial in these problems.

The process previously applied to the Ackley function will now be extended to all the test functions mentioned. Consequently, 23 models will be provided. The impact of parameter tuning on the performance of the GSA will be investigated in the following subsections. This will be followed by a discussion on each parameter's importance, main effect, and interaction effect. Finally, a summarization will rank the levels of parameters, and the optimal value for each factor will be proposed.

A. Presentation of Optimal Solutions

The optimal solution for each function along with the appropriate parameter values are presented in Table 5. The second column of this table (“GSA”) displays the solutions obtained from the original GSA [13], averaged over 15 runs. The third column (“Best Tuned GSA with RSM”) showcases the best solution among all 780 experiments designed by RSM, with the experiment number identified in the "Expnum" column. The columns labeled “Parameters” showcase the values of parameters associated with the experiment, reflecting the parameter values of the best (optimal) solution. The fourth column presents the average of 15 rounds for this experiment, allowing for comparability with the results of the original GSA.

Table 5: Results for 23 test functions, comparing original GSA with the proposed method

Test Function	GSA	Best Tuned GSA with RSM	Mean Tuned GSA with RSM	Expnum	Parameters				
					Population size	Rpower	Alpha	G0	Beta
F1	7.3×10^{-11}	1.04E-43	**3.21e-24**	295	40	0.5	30	70	0.6
F2	4.03×10^{-5}	2.29E-21	**1.08e-12**	338	50	0.75	40	40	0.8
F3	$0.16\times10^{+3}$	21.61158	52.71	634	40	1	10	70	0.6
F4	3.7×10^{-6}	3.40E-11	**7.25e-11**	40	50	0.75	20	100	0.8
F5	25.16	18.93211	24.42	326	20	1	30	70	0.6
F6	8.3×10^{-11}	0	**1.027e-33**	5	50	0.75	40	100	0.4
F7	0.018	0.011099	0.037	707	50	1.25	20	100	0.8
F8	$-2.8\times10^{+3}$	-4522.85	-2969.49	112	40	1	10	70	0.6
F9	15.32	3.979836	**10.28**	646	40	1	30	10	0.6
F10	6.9×10^{-6}	4.44E-15	**1.178e-14**	141	50	0.75	40	100	0.8
F11	0.29	0.021534	1.500	438	50	0.75	20	100	0.8
F12	0.01	1.57E-32	0.020	754	50	0.75	40	100	0.4
F13	3.2×10^{-32}	1.64E-89	2.659e-32	437	50	0.75	20	40	0.4
F14	3.70	0.998004	**2.805**	368	50	0.75	20	100	0.4
F15	8.0×10^{-3}	0.000581	0.0021	646	40	1	30	10	0.6
F16	-1.0316	-1.03163	-1.031	1	30	0.75	40	100	0.8
F17	0.3979	0.397887	0.397	1	30	0.75	40	100	0.8
F18	3.0	3.0	2.999	216	50	0.75	20	100	0.8
F19	-3.7357	-3.86278	-3.862	9	40	1	30	10	0.6
F20	-2.0569	-3.322	-3.200	272	50	0.75	40	100	0.8
F21	-6.0748	-10.1532	-8.161	4	40	1	30	70	0.6
F22	-9.3399	-10.4029	**-10.402**	7	40	1	30	70	0.6
F23	-9.4548	-10.5364	**-10.536**	419	40	1	50	70	0.6

Overall, the performance of the GSA algorithm has significantly improved across various functions; for instance, F6 has shown noticeable enhancements. As with other metaheuristic approaches, the fine-tuning of parameters plays a crucial role in optimizing the algorithm's performance.

B. Parameter Ranking

In this section, the analysis of results obtained from 23 standard functions is presented, focusing on determining and rating effective parameters based on their effect size for each function. Finally, an assessment of the importance of each parameter is conducted across all 23 test functions.

Effective parameters are identified from the ANOVA table and presented in Table 6 which includes five rows for main parameters, ten rows for two-way interactions between parameters, and 23 columns for test functions.

Parameters with a p-value less than 0.001 are considered to have a significant effect, and these significant parameters are highlighted in the table.

Table 6, reveals that the most effective parameters are Rpower and G0 sequentially. Additionally, when investigating interaction effects, Rpower plays a meaningful role in combination with other parameters, while population has an effective role in only a few functions.

Although the importance of parameters is well represented in Table 6, a quantitative technique is needed for a more thorough analysis. Therefore, Table 7 rates the importance of parameters based on their significant contribution to how many functions. Although the importance of parameters is well represented in Table 6, a quantitative technique is needed for a more thorough analysis.

Therefore, Table 7 rates the importance of parameters based on their significant contribution to how many functions. For instance, population demonstrated a significant effect in 16 test functions, effective in 69.75% of them. According to this table, Rpower and G0 are the most effective parameters, standing in first and second place, respectively, and population, Beta, and Alpha follow in subsequent positions.

In addition to the quantitative criterion, the effect size of parameters is crucial. Effect size is measured with the absolute effect measure, calculated for each parameter in each test function, summarized, ranked, and presented in Table 7. According to this table, Rpower exhibits the highest variance in response values in 15 functions (65.22% of all functions) and is the second most impactful in 3 functions. In five functions, population ranks first, suggesting a considerable role overall. G0 secures the second place in 10 functions, maintaining its position from the previous analysis. Considering both first and second place analyses, it is concluded that Rpower holds the top spot in terms of importance, followed by G0 in the second place.

Table 6: Significance matrix for the 23 benchmark test functions

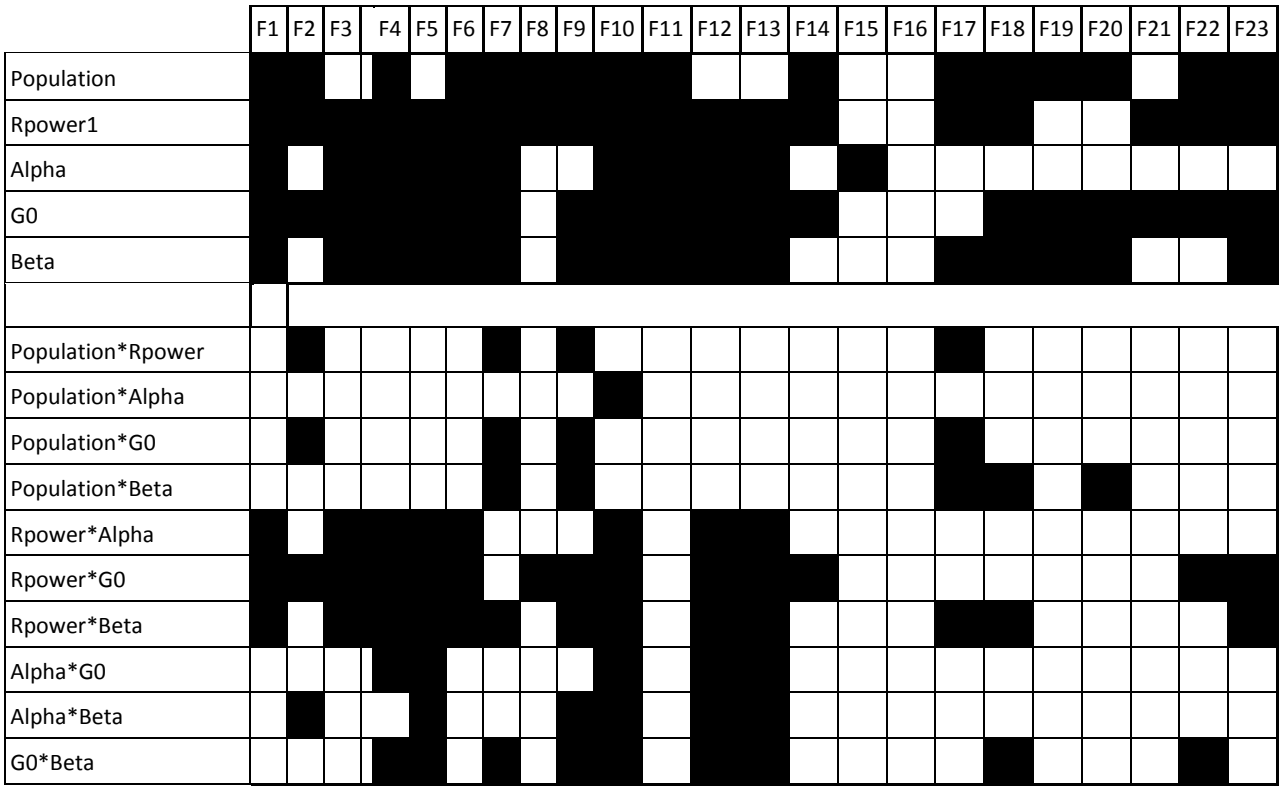


Table 7: The number and percentage of functions (from23 test function) where each parameter showed a significant effect for on the optimization, ranked by effect size

Factor	significance (p<0.001)	%	E_i in first stand	%	E_i in second stand	%	First and second Stand
Rpower	19	82.61	15	65.22	3	13.04	78.26
G0	19	82.61	3	13.04	10	43.48	56.52
Population	16	69.57	5	21.74	5	21.74	43.48
Beta	16	69.57	0	0.00	2	8.70	8.70
Alpha	11	47.83	0	0.00	3	13.04	13.04

C. Investigation of Parameter Levels

Table 8 displays distribution of factor levels across test functions. The optimal (best) levels detected for each parameter based on the number of test functions. For instance, Level 2 (equal to 0.75) is identified as the best level for Rpower, with the participation of 12 test functions.

Similarly, the best levels for other parameters are determined. In summary, the optimal values for the population, Rpower, Alpha, G0, and Beta parameters of GSA are 50, 0.75, 30 (or 40), 100, and 0.6, respectively, as presented in Table 9. For a more in-depth analysis, contour, surface, and interaction plots can be generated and investigated.

Table 8: Distribution of Optimal Factor Levels Across Test Functions

	Level 1		Level 2		Level 3		Level 4		Level 5	
Factor	N.	%	N.	%	N.	%	N.	%	N.	%
Population	1	4.35	2	8.70	9	39.13	11	**47.83**	0	0.00
Rpower	1	4.35	12	**52.17**	9	39.13	1	4.35	0	0.00
Alpha	2	8.70	6	26.09	7	**30.43**	7	**30.43**	1	4.35
G0	3	13.04	2	8.70	7	30.43	11	**47.83**	0	0.00
Beta	0	0.00	4	17.39	10	**43.48**	9	39.13	0	0.00

Table 9: Optimal levels for each parameter across 23 benchmark test functions

Factor	Optimal Level	Value
Population	Level 4	50
Rpower1	Level 2	0.75
Alpha	Level 3 ,Level 4	30 , 40
G0	Level 4	100
Beta	Level 3	0.6

Conclusion

The performance of metaheuristic algorithms relies on parameter tuning. While several approaches have been proposed for parameter tuning, a comprehensive analysis of the parameters' influence on the algorithm's behavior is of paramount importance. The approach presented in this paper utilizes Response Surface Methodology (RSM) to systematically design experiments. The model created by RSM offers valuable insights through efficient measures and plots, enabling a deeper understanding of the algorithm's behavior.

This study focused on the Gravitational Search Algorithm (GSA), a valuable metaheuristic algorithm. The proposed approach is applied to 23 single and multimodal test functions, revealing the relative importance of each parameter and analyzing interactions between parameters.

Ultimately, this study provides valuable insights for researchers seeking to enhance the performance of not only the GSA algorithm but also other metaheuristic algorithms using the proposed method. The approach is adaptable and can be applied to fine-tune parameters in various optimization problems.

For future work, this approach can be extended for sensitivity analysis of GSA.

This paper has focused on single and multimodal GSA; however, future work could extend the analysis to include multi-objective versions as well. Additionally, the methodology can be adapted for a more in-depth understanding and fine-tuning of other metaheuristic algorithms.

Since heuristic algorithms share similarities in controlling exploration and exploitation of problem spaces, this method can be applied when a close resemblance exists between parameters. Additionally, designing experiments spanning aligned domains facilitates performance assessments under equivalent conditions. Thereby, equitable evaluations can be conducted for closely related algorithms. Furthermore, in other situations using normalized and standardized metrics maps measures to a common scale. This enables impartial contrasting through statistically analyzing unified data. Consequently, the methodology provides pathways for fair benchmarking even with inherent dissimilarities through mapping to a shared assessment platform.

Author Contributions

M. Amoozegar carried out: Implemented algorithm, Interpreted the results and wrote the manuscript.

S. Golestani: Data analysis, interpreted the results and wrote the manuscript.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

GSA	Gravitational Search Algorithm
RSM	Response Surface Methodology
DOE	Design of Experiments
GA	Genetic Algorithm
CCD	Central Composite Design
BBD	Box-Behnken Design

References

[1] D. Karaboga, "Artificial bee colony algorithm," scholarpedia, 5: 6915, 2010.

[2] V. Sharma, A. K. Tripathi, "A systematic review of meta-heuristic algorithms in IoT based application," Array, 14: 100164, 2022.

[3] H. Rajabi Moshtaghi, A. Toloie Eshlaghy, M. R. Motadel, "A comprehensive review on meta-heuristic algorithms and their classification with novel approach," J. Appl. Res. Ind. Eng., 8: 63-89, 2021.

[4] S. Kaur, Y. Kumar, A. Koul, S. Kumar Kamboj, "A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: open issues and challenges," Arch. Comput. Methods Eng., 30: 1863-1895, 2022.

[5] E. G. Talbi, Metaheuristics: From Design to Implementation, vol. 74, John Wiley & Sons, 2009.

[6] G. Xu, "An adaptive parameter tuning of particle swarm optimization algorithm," Appl. Math. Comput., 219: 4560-4569, 2013.

[7] T. I. de Paula, G. F. Gomes, J. H. de Freitas Gomes, A. P. de Paiva, "A mixture design of experiments approach for genetic algorithm tuning applied to multi-objective optimization," in Proc. Optimization of Complex Systems: Theory, Models, Algorithms and Applications: 600-610, 2020.

[8] E. Shadkam, "Cuckoo optimization algorithm in reverse logistics: a network design for COVID-19 waste management," Waste Manage. Res., 40: 458-469, 2022.

[9] E. Shadkam, "A novel two-phase algorithm for a centralised production planning problem by symmetric weighted DEA approach: a case study in energy efficiency," Eur. J. Ind. Eng., 16: 732-756, 2022.

[10] A. Gunawan, H. C. Lau, "Fine-Tuning algorithm parameters using the design of experiments approach," in Proc. Learning and Intelligent Optimization: 278-292, 2011.

[11] Z. K. Pourtaheri, S. H. Zahiri, S. M. Razavi, "Stability investigation of multi-objective heuristic ensemble classifiers," Int. J. Mach. Learn. Cybern., 10: 1109-1121, 2019.

[12] C. Li, Q. Xiao, Y. Tang, L. Li, "A method integrating Taguchi, RSM and MOPSO to CNC machining parameters optimization for energy saving," J. Cleaner Prod., 135: 263-275, 2016.

[13] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, "GSA: a gravitational search algorithm," Inf. Sci., 179: 2232-2248, 2009.

[14] M. Amoozegar, E. Rashedi, "Parameter tuning of GSA using DOE," in Proc. 2014 4th International eConference on Computer and Knowledge Engineering (ICCKE): 431-436, 2014.

[15] S. Fraley, M. Oom, B. Terrien, J. Date, "Design of experiments via Taguchi methods: orthogonal arrays," Michigan Chemical Process Dynamic and Controls Open Text Book, 2006.

[16] D. C. Montgomery, Design and analysis of experiments: John Wiley & Sons, 2008.

[17] G. F. Gomes, F. A. de Almeida, "Tuning metaheuristic algorithms using mixture design: Application of sunflower optimization for structural damage identification," Adv. Eng. Software, 149: 102877, 2020.

[18] E. B. d. M. Barbosa, E. L. F. Senne, "Improving the fine-tuning of metaheuristics: An approach combining design of experiments and racing algorithms," J. Optim., 8042436, 2017.

[19] R. H. Myers, D. C. Montgomery, C. M. Anderson-Cook, Response surface Methodology: Process and Product Optimization Using Designed Experiments, vol. 705: John Wiley & Sons, 2009.

[20] J. h. Wu, X. w. Zhen, G. Liu, Y. Huang, "Optimization design on the riser system of next generation subsea production system with the assistance of DOE and surrogate model techniques," Appl. Ocean Res., 85: 34-44, 2019.

[21] B. Adenso-Diaz, M. Laguna, "Fine-tuning of algorithms using fractional experimental designs and local search," Oper. Res., 54: 99-114, 2006.

[22] F. Hutter, H. H. Hoos, K. Leyton-Brown, K. Murphy, "Time-bounded sequential parameter optimization," Learning and Intelligent Optimization, ed: Springer, pp: 281-298, 2010.

[23] H. Akbaripour, E. Masehian, "Efficient and robust parameter tuning for heuristic algorithms," Int. J. Ind. Eng, 24: 143-150, 2013.

[24] J. L. J. Pereira, M. B. Francisco, F. A. de Almeida, B. J. Ma, S. S. Cunha Jr, G. F. Gomes, "Enhanced lichtenberg algorithm: A discussion on improving meta-heuristics," Soft Comput., 27: 15619-15647, 2023.

[25] V. Kapoor, S. Dey, A. Khurana, "An empirical study of the role of control parameters of genetic algorithms in function optimization problems," Int. J. Comput. Appl., 31: 20-26, 2011.

[26] A. Haines, K. Mills, J. Filliben, "Determining relative importance and best settings for genetic algorithm control parameters," NIST Pub, 912472: 1-22, 2012.

[27] M. I. G. Arenas, P. Á. C. Valdivieso, A. M. M. García, J. J. M. Guervós, J. L. J. Laredo, P. García-Sánchez, "Statistical analysis of parameter setting in real-coded evolutionary algorithms," Parallel Problem Solving from Nature, PPSN XI, ed: Springer, pp: 452-461, 2010.

[28] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, "Filter modeling using gravitational search algorithm," Eng. Appl. Artif. Intell., 24: 117-122, 2011.

[29] M. Amoozegar, H. Nezamabadi-pour, "Software performance optimization based on constrained GSA," in Proc. 2012 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP): 134-139, 2012.

[30] A. Chatterjee, G. K. Mahanti, N. N. Pathak, "Comparative performance of gravitational search algorithm and modified particle swarm optimization algorithm for synthesis of thinned scanned concentric ring array antenna," Prog. Electromagn. Res. B, 25: 331-348, 2010.

[31] M. Yin, Y. Hu, F. Yang, X. Li, W. Gu, "A novel hybrid K-harmonic means and gravitational search algorithm approach for clustering," Expert Syst. Appl., 38: 9319-9324, 2011.

[32] R. K. Roy, Design of experiments using the Taguchi approach: 16 steps to product and process improvement: John Wiley & Sons, 2001.

[33] K. Kelley, K. J. Preacher, "On effect size," Psychol. Methods, 17: 137, 2012.

[34] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, "BGSA: binary gravitational search algorithm," Nat. Comput., 9: 727-745, 2010.

Biographies

Maryam Amoozegar received the Ph.D. degree in Software Engineering from the Iran University of Science and Technology, Tehran, Iran, in 2021. She is now an assistant professor at the Computer and Information Technology department of the Institute of Science and High Technology and Environmental Sciences at the Graduate University of Advanced Technology in Kerman. Her research interests are anomaly detection, streaming data processing, data mining, machine learning, and Swarm intelligence

- Email: amoozegar@kgut.ac.ir
- ORCID: 0000-0001-7161-8623
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: https://kgut.ac.ir/index.jsp?fkeyid=&siteid=1&pageid=1190&tid=219

Shahrzad Golestani received her B.Sc. degree in Computer Software in 2007 and her M.Sc. in Information Technology Engineering in 2012. Currently, she is a Ph.D. candidate in Computer Science at the University of Saskatchewan and a researcher at Saskatchewan Polytechnic. Her research interests include Network Security, Internet of Things Security, Machine Learning, Federated Learning, and Data Science.

- Email: shg550@usask.ca
- ORCID: 0009-0008-3198-3034
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
M. Amoozegar, S. Golestani, "Response surface methodology for behavior analysis and performance improvement of gravitational search algorithm," J. Electr. Comput. Eng. Innovations, 12(2): 387-400, 2024.

DOI: 10.22061/jecei.2024.10385.698

URL: https://jecei.sru.ac.ir/article_2093.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Image Recreating in Improving the Performance of Architectures for Person Re-Identification

S. H. Zahiri [*]***, R. Iranpoor***

Department of Electrical Engineering, Faculty of Engineering, University of Birjand, Birjand, Iran.

Article Info

Article History:
Received 12 January 2024
Reviewed 27 February 2024
Revised 30 March 2024
Accepted 03 April 2024

Keywords:
Person re-identification
Deep learning
Image processing
Convolutional neural network
Computer vision
Image detection

[*]Corresponding Author's Email Address:
hzahiri@birjand.ac.ir

Abstract

Background and Objectives: Re-identifying individuals due to its capability to match a person across non-overlapping cameras is a significant application in computer vision. However, it presents a challenging task because of the large number of pedestrians with various poses and appearances appearing at different camera viewpoints. Consequently, various learning approaches have been employed to overcome these challenges. The use of methods that can strike an appropriate balance between speed and accuracy is also a key consideration in this research.

Methods: Since one of the key challenges is reducing computational costs, the initial focus is on evaluating various methods. Subsequently, improvements to these methods have been made by adding components to networks that have low computational costs. The most significant of these modifications is the addition of an Image Re-Retrieval Layer (IRL) to the Backbone network to investigate changes in accuracy.

Results: Given that increasing computational speed is a fundamental goal of this work, the use of MobileNetV2 architecture as the Backbone network has been considered. The IRL block has been designed for minimal impact on computational speed. By examining this component, specifically for the CUHK03 dataset, there was a 5% increase in mAP and a 3% increase in @Rank1. For the Market-1501 dataset, the improvement is partially evident. Comparisons with more complex architectures have shown a significant increase in computational speed in these methods.

Conclusion: Reducing computational costs while increasing relative recognition accuracy are interdependent objectives. Depending on the specific context and priorities, one might emphasize one over the other when selecting an appropriate method. The changes applied in this research can lead to more optimal results in method selection, striking a balance between computational efficiency and recognition accuracy.

Introduction

With the development of deep learning and continuous improvement in computational power, significant progress has been achieved in the field of object-detection, which is one of the crucial issues in the realm of machine vision. By introducing the first solutions for addressing this problem using convolutional neural networks, notable methods gradually emerged to better tackle the object-detection challenge. Person re-identification (ReID), which has recently gained substantial attention due to its extensive applications in various domains, serves as a fundamental and essential

function in intelligent surveillance systems. An essential task for a network-based surveillance system consisting of distributed cameras is to associate individuals across camera views at different locations and times. This is referred to as the person ReID problem and forms the basis for many other important application programs. A comprehensive person ReID system comprises three main components: person identification, person tracking, and person retrieval [1]. Current research efforts to solve the ReID problem have mainly focused on two aspects:

1. Feature learning, or in other words, developing feature representations that are distinct for identity while being invariant to viewpoint and constant lighting conditions [2].
2. Metric learning, which practically involves developing machine learning methods to optimize the discriminative parameters of a ReID model [3].

However, achieving automatic ReID remains a significant challenge due to inherent limitations in most visually generated features from individuals' appearances. The pedestrian images captured by non-overlapping cameras usually under an uncontrolled environment and most of the images are low quality, which leads to some conventional biometrics features like gait and face are not feasible to be used for the task. In these circumstances, the appearance features of the pedestrians, which are extracted from their clothes' colors or objects carried with them, seem to be more suitable for the person ReID task. In addition, different pedestrians may share similar appearances, which often leads to the appearances of different pedestrians looked quite similar [4]. In this work, given the importance of feature extraction alongside the need for sufficient speed for real-time applications, faster architectures are employed for feature extraction. As the main component of the work, to enhance performance, a module called the Image Recreation Layer (IRL) is added to the Backbone, injecting input images between the blocks of the Backbone network. This component can be incorporated alongside various architectures. The obtained results demonstrate an improvement in the utilized method.

Related Works

The primary objective of the research conducted by SUN et al. [5] is to acquire superior features using a Part-based Convolutional Neural Network (PCB). In this section, a convolutional descriptor is generated from the input image, capturing features related to various image components. Subsequently, an enhanced method for Part Pooling (RPP) is introduced.

Additionally, Zhong et al. [6] address camera style adaptation (camStyle) as a solution to enhance data diversity and improve the performance differences among cameras. They initially utilize CycleGAN, which not only increases data variety but also introduces a significant level of noise. Zheng and colleagues propose a learning framework that employs the DG-NET network for the pairing of a generator module and a discriminator module for re-identification purposes.

Mohammed et al. [7] introduce the real-time ReID-DeePNet system, where they utilize the fusion of scores from two different deep learning models. In another study, Y. Zhu et al. [8] mimic the human visual perception process, which involves transitioning from coarse to fine-grained observation. The core of this approach involves a multi-scale structure consisting of two key elements: the Global Channel-Aware Attention (GCA) module for capturing global structural information and the Adaptively Spatial Feature Fusion (ASFF) module for highlighting distinctive features. Furthermore, they introduce a Bidirectional Pairwise Metric (BPM) loss function. The topological relationship between global and local features forms a framework that derives a novel feature representation through a graph transformation network. This representation is then trained and tested, based on the work of Wang et al. [9].

Various other approaches have also been presented to address some challenges such as attention to non-human sections, high complexity of the model and inference time, use of false labels for unsupervised methods, etc. For example, In the work of Zhu et al. [10], alignment scheme in transformer architecture is discussed and automatic transformer (AAformer) is introduced for automatic localization of human and non-human parts. Cho et al. [11], proposed a new framework for Part-based Pseudo Label Refinement (PPLR) proposed to decrease label noises with a reliable complementary relationship. In another work, two types of attention maps have been used to inform feature maps about the individual and relevant body parts, for which a holistic attention branch (HAB) and a partial attention branch (PAB) have been proposed [12].

Datasets

Research in the field of deep learning technology requires a substantial amount of data for model training. To exploit robust person ReID models, it is crucial to have the available ReID datasets with the characteristics of cluttered background, occlusions and overlapped bodies, etc [4].

Given that person ReID is a thoroughly investigated problem, numerous datasets have been made available for research purposes. Datasets like VIPeR [13], GRID [14], and CUHK01 [15] not only have a limited number of individuals but also a small number of images per individual. Most of them employ handcrafted labeling methods for person identification.

With the advancement of deep learning, small-scale datasets can no longer satisfy its training requirements. Therefore, large-scale datasets such as CUHK03[16],

Market1501 [17], and DukeMTMC-reid[18] have been proposed and accepted.

The Market-1501 dataset is one of the most widely used datasets in the field of person detection and identification in images. This dataset includes 1501 different individuals in an open environment, with a total of approximately 32,217 images collected from 6 surveillance cameras equipped with various sensors. Each individual has around 6 to 20 images in the dataset, captured as full-body images. The dataset is divided into two parts: training and testing (query and gallery). The training part comprises images of the first 751 individuals, each having around 12 images. The testing part involves another set of 750 individuals, with only one query image per individual and approximately 4 to 18 gallery images for each individual [17].

CUHK03 is much larger dataset which includes 13164 images of 1360 pedestrians. Unlike existing datasets, which only provide manually cropped pedestrian images, this dataset provides automatically detected bounding boxes for evaluation close to practical applications [16].

In Fig. 1 an example of images from the two datasets used is presented.

Fig. 1: Samples of images in the datasets, with part (a) being images captured from two separate cameras in the first and second rows from the cuhk03 dataset, and part (b) from the market-1501 dataset.

To evaluate a Re-ID system, Cumulative Matching Characteristics (CMC) [19] and mean Average Precision (mAP) [17] are two widely used measurements. CMC is accurate when only one ground truth exists for each query, since it only considers the first match in evaluation process. However, the gallery set usually contains multiple groundtruths in a large camera network, and CMC cannot completely reflect the discrim- inability of a model across multiple cameras [20].

Backbone Networks

Backbone networks serve as the primary feature extractors for object-detection and person ReID tasks. These networks take images as input and output feature maps corresponding to the input image. Most backbone networks used for object detection are originally designed for classification tasks, with the fully connected layers often removed. Enhanced versions of classification networks are also available. For instance, researchers have designed new backbones tailored to better address the specific challenges of object-detection.

specific needs in tasks like [21] and [22] are addressed by modifying backbone networks, such as adding or removing layers or replacing some layers with other specialized design layers.

Due to various requirements regarding accuracy alongside efficiency, individuals can choose deeper and more compact backbones like ResNet [23], ResNeXt [24], AmoebaNet, or lightweight backbones like MobileNet [25], ShuffleNet [26], SqueezeNet [27], Xception[28], MobileNetV2 [28]. When targeting mobile devices, lightweight backbones can meet the requirements effectively.

The MobileNetV2 architecture is a deep neural network design for image processing with the goal of reducing the number of parameters and computational operations. This architecture is particularly beneficial for applications and devices with limited resources. As input images in many applications are larger than traditional matrices, the use of architectures with a high number of parameters and computational operations can lead to excessive computational resource consumption. This is where the MobileNetV2 architecture employs specific techniques to minimize resource consumption.

The main structure of MobileNetV2 is composed of Depthwise Separable Convolution layers. These layers are divided into two main stages:

Depthwise Convolution: In this stage, each input channel is convolved separately with a smaller kernel filter (usually 3x3 or 5x5). These filters are applied to each input channel and ultimately fulfill the role of a regular convolution.

Pointwise Convolution (1x1 Convolution): In this stage, 1x1 kernel filters are applied to the output of the previous stage. These filters are used to combine and transfer information between channels. Each input channel is transformed into an output channel.

The combination of these two stages allows the network to aggregate information in a concise manner while significantly reducing the number of parameters and computational operations. Therefore, MobileNetV2 is recognized as a lightweight architecture suitable for image processing on mobile devices and constrained resources [28].

Methods

In this section of the research plan, the focus is on reviewing the efforts made in the field of person ReID using feature-based partitioning methods. Feature

partitioning is a common approach in person ReID identification, where feature maps are divided into several non-overlapping regions or predefined blocks based on learning local features.

Fig. 2 illustrates the overall workflow of the person ReID process. Initially, the CNN of interest is trained using the dataset to perform feature extraction. The images are divided into training and evaluation sets, with 80% used for training and 20% for evaluation. Subsequently, the network is tested using images from the query and gallery sets. One image is selected as the main image, and based on the extracted features, a decision is made regarding the similarity of individuals in the recognition task.

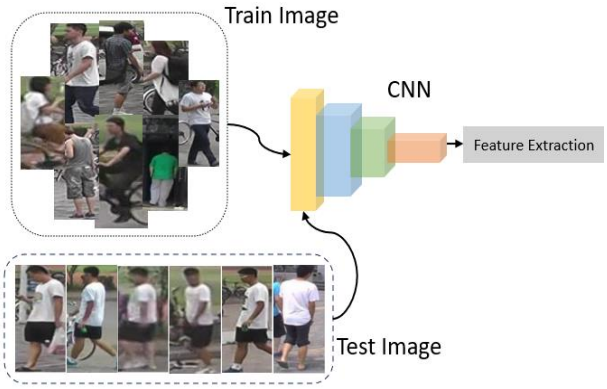


Fig. 2: General approach in person re-identification.

Here, the PCB (Part-based Convolutional Baseline) architecture has been employed to enhance common methods and align features [29]. This architecture takes the whole image as input and divides the resulting feature map into regular convolutional layers to segment it into different regions. Furthermore, the RPP (Random Partitioning Pooling) [5] method has been proposed, aiming to enhance the re-allocation of key points within each local area and strengthen stability to mitigate non-alignment issues.

In all these cases, Backbone networks that have been pre-trained on large datasets like ImageNet have been used. Architectures like ResNet [23] and DenseNet [31] are employed for tasks with larger dimensions and a need for higher computational and time resources. Additionally, lightweight architectures like MobileNetV2 [29] are used to reduce computational costs.

In this section, by modifying certain structures of the Backbone networks, the modified architectures are compared with the pre-trained ones. Due to the requirement for reduced computational costs in real-time applications, one of the main objectives of this comparison is to reduce training time. For this reason, the MobileNetV2 architecture has been implemented with modifications for use as a Backbone network for the purpose of conducting comparisons.

In the design of the MobileNetV2 architecture, key parameters such as kernel sizes and channel numbers are carefully adjusted to strike a balance between accuracy and performance. Moreover, different versions of MobileNetV2 with varying depths and complexities can be used for different tasks.

As depicted in Fig. 3, the MobileNetV2 architecture is utilized for the purpose of result comparison. The reason for choosing this architecture is its faster performance compared to other available architectures, ensuring lower computational costs and thus making it suitable for broader usage. Alongside this section, a block for image re-creation is injected into intermediate layers of the main architecture. In the Backbone network, the input image with dimensions of 256x128 and 3 channels from the color image is transformed through five stages to dimensions of 8x4 with a channel length of 1280. Finally, it is converted to dimensions of 1x1280 through average pooling. Subsequently, the final class is determined using a fully connected layer, PCB, or RPP method.

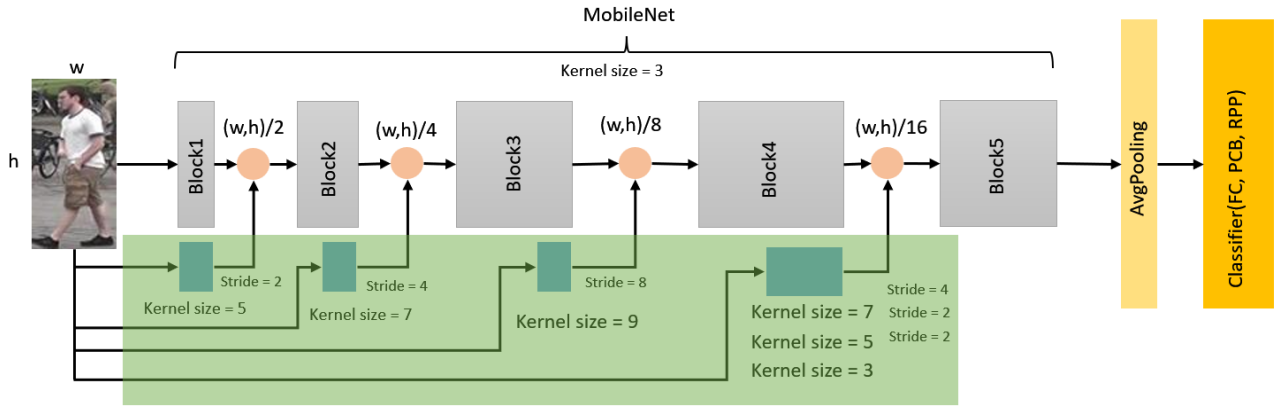


Fig. 3: The MobileNetV2 architecture, coupled with the Image Recreating Layers (IRL) block.

Image Recreating Layers (IRL):

Image recreating layers are injected between the Backbone network layers. In this process, the input image passes through several sections in parallel with kernel sizes from different layers. The kernel sizes in the first layer of each section are 5, 7, and 9 respectively, and in the last section, two layers with kernel sizes of 7 and 5 are used. In subsequent layers of each section, a kernel size of 3 is used, and the output of each section is used among the blocks of the Backbone network.

Table 1 reports details of the data used in training the network. This section contains the total count of images in the dataset and also accounts for the query and gallery images. Training images are typically divided into two groups: the training set and the evaluation set, which helps in validating the performance of the model. The table lists the number of classes utilized in training the network.

Table 1: Information related to the data used in training the network

Dataset	Number of image	Train image	Valid image	Quary image	Gallery image	Num train ID
Market-1501	36036	12185	751	3368	19732	751
Cuhk03	14097	6260	1105	1400	5332	767

Results

After applying the mentioned changes, the results obtained from the implemented experiments are summarized in Table 2 For the purpose of comparison, architectures like DenseNet and ResNet50 are pitted against MobileNetV2 along with fully connected layers.

In Fig. 4, the visual output is presented. A single image from the Query set is chosen, and other images from the same class or those closely related are selected. Finally, it's determined which image belongs to the class of the original image. This distinction is indicated by the colors green and red.

Table 2: Comparison of methods for market1501 and cuhk03 datasets

Methods	Dataset	Rank@1	Rank@5	mAP
DenseNet	market	0.877	0.955	0.698
	Cuhk03	0.498	0.517	0.362
ResNet50	market	0.869	0.943	0.702
	Cuhk03	0.502	0.518	0.373
ResNet50 +IRL	market	0.862	0.938	0.701
	Cuhk03	0.508	0.526	0.401
MobileNetV2	market	0.685	0.854	0.444
	Cuhk03	0.407	0.467	0.213
MobileNetV2 +RPP+IRL	market	0.686	0.857	0.451
	Cuhk03	0.437	0.473	0.261

Fig. 4: The visual output of the utilized method.

Additionally, the Table 2 includes results for these architectures with the added image recreating layers for CUHK03 and Market-1501 datasets. All these results were obtained under similar conditions for the Market-1501 and CUHK03 datasets.

A noteworthy point regarding the use of more complex networks like ResNet50 and DenseNet in this context is that these architectures have been pre-trained on large datasets. Moreover, they involve more layers, which significantly increases the training time as a result.

The issue of processing time is highly critical to ensure the feasibility of using these systems, especially in resource-constrained environments. Therefore, a trade-off between accuracy and processing time is considered. Although the fundamental issue in this architecture lies in the number of its layers, it has been designed to increase speed for real-world applications and for use on hardware-constrained devices. Consequently, the accuracy of this architecture is lower compared to other architectures. The incorporation of the image recreation layer alongside it can bring about a partial improvement in the network's performance. If we use the image reformation layer in larger architectures, the results may not change significantly, and in some cases, it might

deviate from the primary goal. According to the results in Table 2, it can be stated that the image reformation layer is more effective on architectures with fewer layers.

According to Table 3, the execution time for each epoch during the training of various methods is presented in seconds. All these stages for all methods have been obtained under similar conditions with comparable hardware. The use of the MobileNetV2 architecture is faster by more than 50%, and this speed increase also holds true concerning the addition of the image recreation layer.

In Fig. 5 and Fig. 6, the loss and accuracy curves related to the training are presented. These curves pertain to both the training and evaluation datasets.

During training, all the mentioned cases have used the CrossEntropy loss function. The number of epochs for training is set to 60, and a batch size of 32 is considered.

Table 3: Implementation time of different methods

Dataset	Dens-Net	Res-Net50	ResNet +RPP	Mobil-NetV2	Mobile-NetV2 +RPP+IRL
Market-1501	63.7 (s)	50.3 (s)	85.5 (s)	23.2 (s)	29.3 (s)
Cuhk03	90.4 (s)	79.9 (s)	110 (s)	50.9 (s)	50.2 (s)

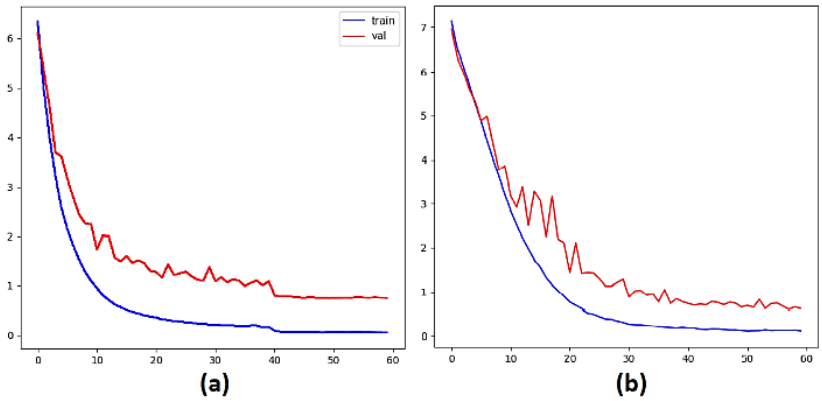


Fig. 5: The loss charts for training and evaluation data. Section (a) is for market-1501 and (b) is for cuhk03.

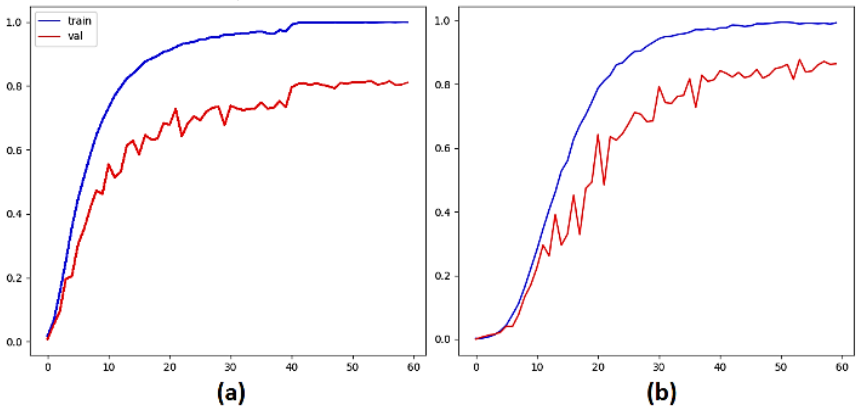


Fig. 6: The accuracy charts for training and evaluation data. Section (a) is for market-1501 and (b) is for cuhk03.

Conclusion

Person ReID is a problem with numerous practical applications in our daily lives. It is crucial to employ methods that can consider its applicability by reducing computational costs. The trade-off between speed and accuracy is also a challenge that should be addressed, and the field should be flexible enough to use various methods depending on specific needs.

Various challenges and issues exist in improving person ReID methods. One of these concerns is the limited quantity of data compared to other machine vision domains. Given the significance of input data in deep learning, challenges related to classification, identification, etc., involve vast datasets, while the person ReID problem is relatively constrained in terms of data. Furthermore, important architectures in this field have

already been trained on these datasets. In the case of the MobileNetV2 architecture, it has been fully implemented and not pretrained. Another significant challenge is the quality of the data. Since these data are collected by cameras that often lack decent quality, using methods capable of establishing meaningful connections between image components is of special importance.

The use of appropriate methods such as data augmentation and the application of generative adversarial networks (GANs), among others, can be employed to increase input data and enhance the performance of networks. Additionally, the adoption of techniques like Vision Transformers, which can establish meaningful relationships across the entire image, holds promise for future endeavors.

In general, this work aims to establish an interactive relationship between hardware constraints and adequate accuracy. It is inevitable to consider that many existing methods may achieve higher accuracy, but they often come at the cost of increased computational requirements. This study strives to explore certain modifications and their implementations to effectively enhance results while taking into account this trade-off between accuracy and computational cost.

Author Contributions

Dr. Zahiri has drawn the general road map. R .Iranpoor has searched for important articles in this field. Then, by checking the results and collecting the necessary data, the implementation of the proposed method has been done. Dr. Zahiri reviewed the results and made changes in the way of implementation and final editing of the work.

Acknowledgment

The authors would like to thank the editor and anonymous reviewers.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

ReID	Person re-Identification
IRL	Image Recreating Layers
PPLR	Pseudo Label Refinement
HAB	Holistic Attention Branch
PAB	Partial Attention Branch
PCB	Part-based Convolutional Baseline
RPP	Random Partitioning Pooling
CMC	Cumulative Matching Characteristics
mAP	mean Average Precision

References

[1] W. Wei, W. Yang, E. Zuo, Y. Qian, L. Wang, "Person re-identification based on deep learning—An overview," J. Visual Commun. Image Represent., 82: 103418, 2022.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition: 2360-2367, 2020.

[3] W. S. Zheng, S. Gong, T. Xiang, "Person re-identification by probabilistic relative distance comparison," in Proc. CVPR 2011: 649-656, 2011.

[4] D. Wu et al., "Deep learning-based methods for person re-identification: A comprehensive review," Neurocomput., 337: 354-371, 2019.

[5] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in Proc. the European Conference on Computer Vision (ECCV): 480-496, 2018.

[6] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, "Camera style adaptation for person re-identification," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 5157-5166, 2018.

[7] H. J. Mohammed et al., "ReID-DeePNet: A hybrid deep learning system for person re-identification," Math., 10(19): 3530, 2022.

[8] Y. Zhu et al., "Multiscale global-aware channel attention for person re-identification," J. Visual Commun. Image Represent., 90: 103714, 2023.

[9] L. Zhao, X. Li, Y. Zhuang, J. Wang, "Deeply-learned part-aligned representations for person re-identification," in Proc. the IEEE International Conference on Computer Vision: 3219-3228, 2017.

[10] K. Zhu et al., "Aaformer: Auto-aligned transformer for person re-identification," IEEE Trans. Neural Networks Learn. Syst., 2023.

[11] Y. Cho, W. J. Kim, S. Hong, S. E. Yoon, "Part-based pseudo label refinement for unsupervised person re-identification," in Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 7308-7318, 2022.

[12] Y. Chen, H. Wang, X. Sun, B. Fan, C. Tang, H. Zeng, "Deep attention aware feature learning for person re-identification," Pattern Recognit., 126: 108567, 2022.

[13] D. Gray, H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in Proc. ECCV 2008: 262-275, 2008.

[14] C. C. Loy, T. Xiang, S. Gong, "Multi-camera activity correlation analysis," in Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition: 1988-1995, 2009.

[15] W. Li, R. Zhao, X. Wang, "Human reidentification with transferred metric learning," in Proc. 11th Asian Conference on Computer Vision, Part I 11: 31-44, 2013.

[16] W. Li, R. Zhao, T. Xiao, X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 152-159, 2014.

[17] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, "Scalable person re-identification: A benchmark," in Proc. the IEEE International Conference on Computer Vision: 1116-1124, 2015.

[18] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in Proc. European Conference on Computer Vision: 17-35, 2016.

[19] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, “Shape and appearance context modeling,” in Proc. ICCV 2007: 1–8, 2007.

[20] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," IEEE Trans. Pattern Anal. Mach. Intell., 44(6): 2872-2893, 2021.

[21] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 779-788, 2006.

[22] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, J. Sun, "Detnet: A backbone network for object detection," arXiv preprint arXiv:1804.06215, 2018.

[23] S. Targ, D. Almeida, K. Lyman, "Resnet in resnet: Generalizing residual architectures," arXiv preprint arXiv:1603.08029, 2016.

[24] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, "Aggregated residual transformations for deep neural networks," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 1492-1500, 2017.

[25] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[26] J. Zang, L. Wang, Z. Liu, Q. Zhang, G. Hua, N. Zheng, "Attention-based temporal weighted convolutional neural network for action recognition," in Proc. 14th IFIP WG 12.5 International Conference Artificial Intelligence Applications and Innovations (AIAI 2018): 97-108, 2018.

[27] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," arXiv preprint arXiv:1602.07360, 2016.

[28] C. Fran, "Deep learning with depth wise separable convolutions," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 4510-4520, 2018.

[30] Y. Sun, L. Zheng, Y. Li, Y. Yang, Q. Tian, S. Wang, "Learning part-based convolutional features for person re-identification," IEEE Trans. Pattern Anal. Mach. Intell., 43(3): 902-917, 2019.

[31] Y. Zhu, S. Newsam, "Densenet for dense flow," in Proc. 2017 IEEE International Conference on Image Processing (ICIP): 790-794, 2017.

Biographies

Seyed Hamid Zahiri received the B.Sc., M.Sc. and Ph.D. degrees in Electronics Engineering from Sharif University of Technology, Tehran, Tarbiat Modarres University, Tehran, and Mashhad Ferdowsi University, Mashhad, Iran, in 1993, 1995, and 2005, respectively. Currently, he is a Professor with the Department of Electronics Engineering, University of Birjand, Birjand, Iran. His research interests include pattern recognition, evolutionary algorithms, swarm intelligence algorithms, and soft computing.

- Email: hzahiri@birjand.ac.ir
- ORCID: 0000-0002-1280-8133
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Rasool Iranpoor was born on July 19, 1991. He received M.Sc. degree in Electronic Engineering from Birjand University, Birjand, Iran, in 2018. He is currently a Ph.D. student at Birjand University to receive a Ph.D. degree in Electronics Engineering. His research interests include Machine Learning, Image Processing, Computer Vision, and Deep Learning Algorithms.

- Email: rasool.iranpoor@birjand.ac.ir
- ORCID: 0000-0002-7769-259X
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
S. H. Zahiri, R. Iranpoor, “Image recreating in improving the performance of architectures for person re-identification,” J. Electr. Comput. Eng. Innovations, 12(2): 401- 408, 2024.

DOI: 10.22061/jecei.2024.10446.706

URL: https://jecei.sru.ac.ir/article_2094.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

An Effective Ensemble of Deep and Machine Learning Methods for Classifying the Expertise Shape of CQA Users

S. Nemati *

Department of Computer Engineering, Shahrekord University, Shahrekord, Iran.

Article Info

Article History:
Received 22 January 2024
Reviewed 23 March 2024
Revised 09 April 2024
Accepted 15 April 2024

Keywords:
Shape of expertise
Deep learning
Machine learning
Ensemble method
Community question answering

*Corresponding Author's Email Address: s.nemati@sku.ac.ir

Abstract

Background and Objectives: Community question-answering (CQA) websites have become increasingly popular as platforms for individuals to seek and share knowledge. Identifying users with a special shape of expertise on CQA websites is a beneficial task for both companies and individuals. Specifically, finding those who have a general understanding of certain areas but lack expertise in other fields is crucial for companies who are planning internship programs. These users, called dash-shaped users, are willing to work for low wages and have the potential to quickly develop into skilled professionals, thus minimizing the risk of unsuccessful recruitment. Due to the vast number of users on CQA websites, they provide valuable resources for finding individuals with various levels of expertise. This study is the first of its kind to directly classify CQA users based solely on the textual content of their posts.

Methods: To achieve this objective, we propose an ensemble of advanced deep learning algorithms and traditional machine learning methods for the binary classification of CQA users into two categories: those with dash-shaped expertise and those without. In the proposed method, we used the stack generalization to fuse the results of the dep and machine learning methods. To evaluate the effectiveness of our approach, we conducted an extensive experiment on three large datasets focused on Android, C#, and Java topics extracted from the Stack Overflow website.

Results: The results on four datasets of the Stack Overflow, demonstrate that our ensemble method not only outperforms baseline methods including seven traditional machine learning and six deep models, but it achieves higher performance than state-of-the-art deep models by an average of 10% accuracy and F1-measure.

Conclusion: The proposed model showed promising results in confirming that by using only their textual content of questions, we can classify the users in CQA websites. Specifically, the results showed that using the contextual content of the questions, the proposed model can be used for detecting the dash-shaped users precisely. Moreover, the proposed model is not limited to detecting dash-shaped users. It can also classify other shapes of expertise, such as T- and C-shaped users, which are valuable for forming agile software teams. Additionally, our model can be used as a filter method for downstream applications, like intern recommendations.

Introduction

Community question-answering (CQA) websites have become increasingly popular as platforms for individuals to seek and share knowledge. Notable examples include Stack Overflow and Quora, which have experienced

significant success in the realm of CQA websites [1]. These platforms allow users to ask questions and offer answers to queries posed by other users. To improve the overall content quality, users can also comment and vote on both questions and answers. Additionally, these websites incorporate competitive elements such as reputation scores and badges to encourage active participation from users [1].

Recently, there has been a significant focus on conducting various research studies to identify proficient individuals within the domain of CQA platforms [2]. The prime aim of these investigations is to locate and rank users who possess the necessary knowledge and expertise to effectively address the questions being raised. Providing expert recommendations makes it possible to improve the quality of answers and reduce the waiting time for receiving responses. Additionally, in platforms such as Stack Overflow that involve job positions, the exploration of the most suitable individual for a specific job role serves as an additional motivation for expert-finding studies [3].

Apart from identifying experts, it is also crucial to understand the nature of their expertise. Researchers have proposed various expertise classifications based on the breadth and depth of an expert's knowledge across different fields [4]-[6]. These classifications include (see Fig. 1):

- I-shaped: Experts with advanced knowledge limited to a single field.
- T-shaped: Experts with advanced knowledge in one field and a broad understanding of other fields.
- C-shaped or M-shaped: Experts with advanced and broad knowledge spanning multiple fields.
- Dash-shaped or Hyphen-shaped: Individuals lacking advanced knowledge in any field, but have general knowledge in some fields.

These expertise shapes allow for a better understanding and characterization of experts within CQA platforms. Also, this understanding can help companies and organizations identify and hire trainees for their positions. In recent years, with the development of emerging technologies, companies have become more interested in hiring interns and using on-the-job training methods to prepare them for professional positions. The internship is defined in different ways in different sources, but the most general definition is the conditional employment of people on a part-time or full-time basis for a limited period with a focus on learning specific skills [7].

Every company has its own set of criteria when it comes to choosing an intern. However, in general, an ideal intern should possess the basic knowledge necessary for fulfilling the company's requirements and should be capable of handling the specific work areas (represented by the Dash-shaped users in Fig. 1).

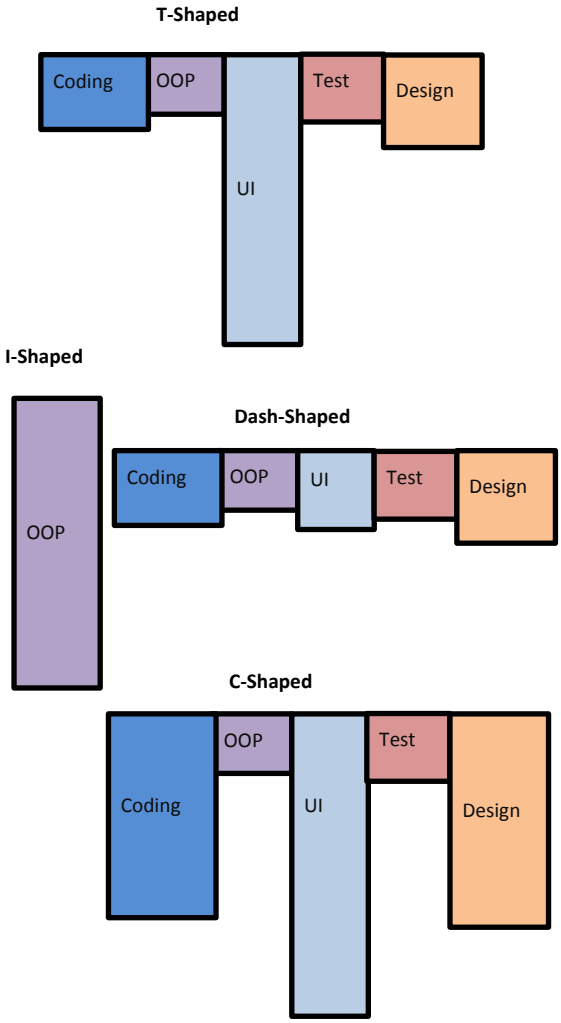


Fig. 1: The types of users in CQAs based on their breadth of knowledge.

Moreover, the internship period should not significantly burden the company financially, as there is a possibility that the intern may not end up being hired and may leave the company. Consequently, it is not advisable to select interns from individuals who are already experts or have extensive experience (represented by the I-, T-, or C-shaped users in Fig. 1), as these experienced individuals typically demand higher wages.

Previous research has primarily focused on identifying T-shaped users in CQA platforms. However, there is only one study that specifically addresses the issue of identifying suitable candidates for internship positions based on their expertise shape [1]. This study suggests that users with dash-shaped expertise have potential and are suitable choices for internship programs. However, the study does not clearly define the expertise shape and proposes statistical features to identify suitable users. Specifically, they propose two methods that utilize the

concept of entropy and the number of skills possessed by the candidates to identify dash-shaped users as suitable candidates for internships. However, their approach has two main limitations. Firstly, it only identifies users who have previously posted comments and does not apply to new users. Secondly, it ignores the most valuable aspect of a post, which is the textual content, when evaluating the expertise shape of users.

To address the issues identified in previous studies, we initially frame the problem of identifying dash-shaped users in CQA platforms as a binary classification problem. By focusing solely on the content of posts, we eliminate the need for user profiles and other statistical features associated with their questions and answers. This approach ensures the effectiveness of our proposed method even in the presence of the cold start problem. To tackle this problem, we introduce a novel ensemble method combining deep learning and traditional machine learning (ML) models. This fusion approach aims to enhance the accuracy and reliability of our solution. To demonstrate the effectiveness of our proposed model, we conduct experiments on three extensive datasets consisting of Stack Overflow questions. The primary contributions of our study can be summarized as follows:

- We introduced the problem of classifying users' shape of expertise only based on their comments.
- We proposed an ensemble method that utilizes the power of both traditional and deep learning models.
- We conducted extensive experiments and compared our method with seven machine learning and twelve deep models using three extensive datasets comprising Stack Overflow questions.

The remainder of the paper continues as follows. In the next section, a brief overview of related studies will be presented. Then, the proposed model will be described. Finally, experimental results, conclusions, and directions for future work will be discussed in the last section.

Literature Review

This section concisely summarizes relevant studies and is divided into three subsections as outlined below. Initially, we examine a selection of related studies regarding CQAs. Subsequently, we explore previous research that delves into the identification of expertise shapes. Finally, we briefly review deep learning models for the expert-finding problem.

A. Community Question Answering (CQA)

CQA platforms such as Stack Overflow are valuable repositories of knowledge. In recent times, there has been significant focus on the task of identifying experts within these platforms. The main concern lies in the low participation rate of users. To tackle this issue, various question routing methods have been devised to determine and suggest the most appropriate answer for new inquiries. A notable example is the work of Fu et al. [8], who introduced a recurrent memory reasoning network. This network utilizes the implicit relevance of the question and the history of the candidate user to locate experts. Another approach, proposed by Wang et al. [9], involves employing user profiles as input for a convolutional neural network. This network predicts the ideal candidate who can provide an answer to a new question. Furthermore, Kundu et al. [10] devised a method to estimate expertise scores by considering factors such as expert knowledge, reputation, and authority. Lastly, Sorkhani et al. [11] introduced a learning-to-rank framework for question routing. This framework incorporates a set of content-based and social-based features to rank and recommend suitable answers.

Researchers have also focused on studying the time-dependent and changing aspects of expertise. In a study by Neshati et al. [3], they introduced the concept of "future experts finding." This concept leverages existing evidence of expertise to predict the likelihood of users becoming experts in the future. The study explored four groups of features, including user behavior, emerging topics, topic similarity, and topic transitions. Another study by Zhang et al. [12] examined the temporal dynamics of answering behaviors in question routing. They developed a context-aware representation for each individual answering a question, taking into account the temporal context. Expertise was estimated by measuring the similarity between the representation of the answerer and the encoding of the question. In more recent research conducted by Liu et al. [13], a user-interest drift model was proposed. This model aimed to capture the dynamic nature of user interests over different periods.

B. Shape of Expertise

Over the past few years, there has been a growing focus on the idea of finding experts who possess specific forms and depths of expertise. This has become an important aspect of the overall problem of identifying and locating experts in various fields. In a study conducted by Rostami and Neshati in 2021, they introduced two retrieval models that are designed to effectively locate and rank individuals who possess dash-shaped expertise [1]. These individuals have an intermediate knowledge that matches the requirements of specific internship programs.

In 2018, Gharebagh and colleagues utilized a clustering method to analyze and extract various skill areas from the tags used in Stack Overflow [6]. They proposed two probabilistic models that are based on entropy calculations, which help in identifying T-shaped users within specific skill domains. In another study conducted by Rostami and Neshati in 2019, they developed two

retrieval models that focus on creating agile teams consisting of T-shaped experts [6]. These models aim to bring together individuals who possess a deep level of expertise in one area (the vertical part of the T) while also having a broader range of knowledge in other related fields (the horizontal part of the T).

In 2023, Rostami and Shakery introduced a deep learning algorithm that evaluates the likelihood of a candidate being a good fit for a particular role within an agile team [14]. Additionally, they implemented an integer linear programming model to identify the optimal members for an agile team with T-shaped experts, selecting them from a pool of highly qualified candidates.

Unlike previous studies in this specific domain, our research endeavors to classify individuals possessing specialized expertise with dash-shaped to fill an internship position. To the best of our knowledge, this particular aspect has not yet been explored or examined.

C. *Deep* Learning

In previous years, methods for identifying experts mainly relied on probabilistic language models [15]-[17], link analysis [18], [19], latent topic modeling [20]-[22], and other approaches. However, with the rise of deep learning, current expert-finding methods predominantly leverage deep learning techniques [12], [23].

In recent times, there has been considerable focus on the application of deep learning in the field of expert finding. Researchers such as Zhao et al. have developed frameworks that utilize random walk and LSTM neural networks to effectively rank candidates who can provide answers to specific questions [24]. Wang et al. have proposed a model based on CNN, which aims to identify experts on platforms like Stack Overflow [9]. Azzam et al. have generated a list of candidates ranked according to their ability to answer a given question by evaluating the cosine similarity between latent semantic vectors associated with each candidate and the question [25]. They have employed fully connected neural networks to learn these latent semantic vectors. Dehghan et al. have utilized an LSTM neural network that processes the breadth-first and depth-first traversal of candidates' expertise tree to find T-shaped experts who specialize in a specific skill area [26].

Li et al. have introduced a model called NeRank, which initially generates embedding representations of answerers and a given question using an LSTM-based model, and then uses a convolutional recommender system to compute the rank of answerers [27]. Tang et al. have proposed an attention-based factorization machine that generates a ranked list of experts in CQAs [28]. Lastly, Dehghan et al. have presented a CNN-based model that generates a ranked list of T-shaped experts who possess expertise in a particular skill area [26].

In a recent study conducted by Nikzad-Khasmakhi et al., they introduced BERTERS, a model that uses transformers and graph embedding techniques to identify potential expert candidates [23]. Similarly, our approach also involves deep learning, but with a different research objective. Unlike previous methods that focused on ranking experts or T-shaped experts with expertise in a particular query, we aim to use deep learning techniques to identify dash-shaped experts who are suitable for internship programs. Hence, the approaches discussed earlier are not applicable to address the specific problem we are trying to solve.

Problem and Data

A. *Problem Statement*

In the previous section, we discussed how most research has focused on finding people who have expertise in a particular field. However, our study is different because we are trying to identify users who have a variety of skills that would be useful in an internship. This is a binary classification challenge where we categorize the data based on the user's proficiency, which is represented by (1) in our investigation.

$$class(u_i) = \begin{cases} 0 & shape(u_i) \in \{I, T, C\} \\ 1 & otherwise \end{cases} \quad (1)$$

where, u_i is the *i*-th user, $shape(u_i)$ is the expertise shape based on the category shown in Fig. 1, and $class$ represents the user class label. It should be noted that the initial labeling of the dataset was done manually by Gharebagh et al. [6] and used in [1]. In the current research, the collection of all the texts related to the answers of the users is in the form of:

$$D = \bigcup_{i \in U} D_i \quad (2)$$

where, $D_i = \{d_{i1}, d_{i2}, \dots, d_{in}\}$ represents the set of answers' texts of the *i*-th user. Each d_i can be shown as:

$$d_i = \bigcup_{j=1}^{k} d_{sa_j,i} \quad (3)$$

to where $sa_j \in S = \{sa_1, sa_2, \dots, sa_m\}$ is a skill area. In order to identify the dash-shaped users the following probability is estimated [1]:

$$P(H = 1, i) \quad (4)$$

This shows the probability of user i being dash-shaped and can be estimated as:

$$P(H = 1, i) \propto \frac{Entropy(i)}{log\,|D_i + 1|} \quad (5)$$

where $Entropy(i)$ can be determined only based on the documents written by the *i-th* user as follows.

$$Entropy(i) = -\sum_{j=1}^{k} P_{sa_j,i}\, log\, P_{sa_j,i} \quad (6)$$

where k denotes the number of skill areas as before, and $P_{sa_{j,i}}$ is calculated using the following equation:

$$P_{sa_{j,i}} = \frac{|D_{sa_{j,i}}|}{|D_i|} \quad (7)$$

As discussed in [1], in addition to dash-shaped users, C-shaped users also have high values for the probability obtained using (6) due to their high diversity of documents while dash-shape users have lower values for the denominator of (5). Consequently, this distinction in probability values can serve as a means to differentiate dash-shaped users from those with T-, I-, and C-shaped behaviors.

B. Dataset

This research employs three datasets originally introduced by Gharebagh et al. [6]. These datasets are derived from queries made on the Stack Overflow website between August 2008 and March 2015, focusing on three specific categories: C#, Java, and Android. Each category comprises questions with relevant tags and their corresponding answers. Each dataset has identified a set of key skill areas based on the 200 most frequent tags within that particular set. Additionally, the level of knowledge possessed by users in each skill area has been established by considering the number of accepted answers provided by the users in those areas. The users within each dataset are then categorized based on their knowledge level in each skill area.

We had to obtain the text of users' questions and answers from the Stack Exchange Data Explorer database since the original datasets didn't have it. We used the user ID, question ID, and answer ID to get this information. Once we had the text, we added it to the datasets. To assess the impact of different category types on the models' performance, we combined the three datasets into one, which we call the All dataset. We also made sure that the datasets were balanced by adjusting the number of records based on the minority class. Specifically, we used random under-sampling to remove samples from the training dataset that belong to the majority class to balance the class distribution. This was done to ensure fair comparison results. You can find more details about the datasets in Table 1.

Table 1: Specification of datasets used in the current study. Max length and Avg. length represent the maximum and average length of texts in number of words

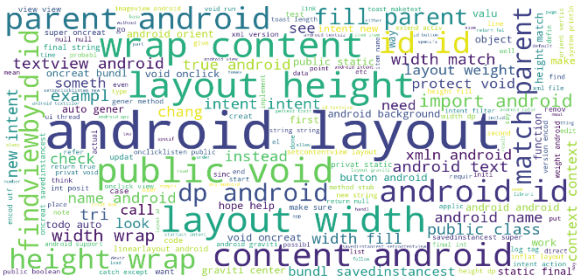
Dataset	# of Rec	Max length	Avg. length	# of unique words
Android	22902	247	95.959	70918
C#	31522	189	102.732	116203
Java	31302	214	100.689	102076
All	85726	278	100.138	237092

To show an overall view of the content of datasets based on their words, word clouds of datasets are shown in Fig. 2.

Proposed Model

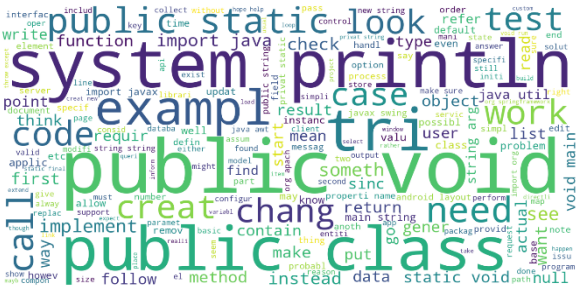
As part of our investigation, we have used a combination of advanced techniques, including deep learning, pre-training, and machine learning. These techniques are classified into three groups of diverse classifiers, aimed at producing more precise results. Our model is based on the concept of meta-learning, where a set of initial classifiers are trained using the given training data to generate predictions. These predictions are then used to train a meta-classifier, which makes the final prediction. Fig. 3 contains a diagram that illustrates the overall structure of our proposed model.

(a)



(b)

(c)



(d)

Fig. 2: Word clouds of the positive class for the (a) All, (b) Android, (c) C#, and (d) Java datasets.

In our proposed approach, we have utilized three different types of classifiers. These include classical machine learning classifiers, popular deep learning models, and a pre-trained transformer-based Bert model. The subsequent sections provide a brief overview of these models.

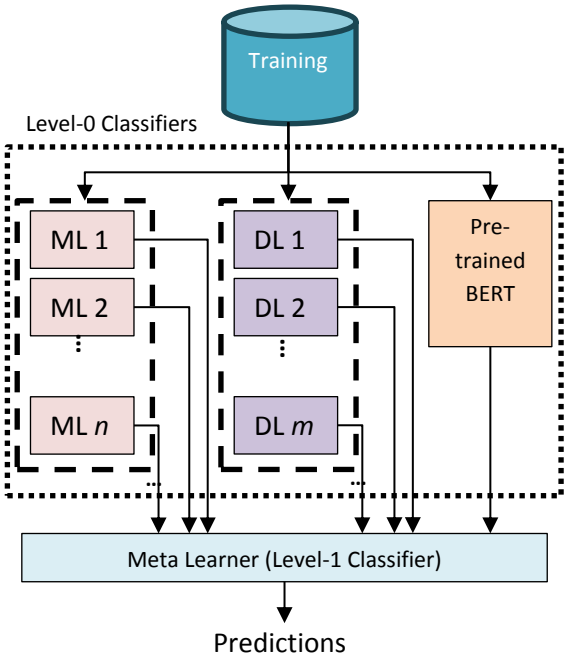


Fig. 3: The overall structure of the proposed ensemble model.

A. Classical Machine Learning Models

In the classical machine learning part of the proposed model, we utilized a total of seven techniques (i.e., n = 7 in Fig. 3): random forest (RF), support vector machine (SVM), decision tree (DT), logistic regression (LR), multi-layer perceptron (MLP), and two boosting classifiers including Adaboost (Ada), and XGBoost (XGB). Presented below is a brief outline of each of these methodologies.

- RF is a powerful methodology that synergizes the strengths of numerous decision trees, each trained on different subsets of data. This approach significantly boosts the precision and reliability of a specific dataset by leveraging the collective consensus derived from these trees. Instead of relying on a single tree's prediction, the RF algorithm calculates the average result generated by an ensemble of trees, thereby ensuring enhanced accuracy [29].
- SVM has gained extensive usage across various domains over a significant period for its ability to forecast outcomes and tackle classification and regression challenges. This technique effectively ascertains the optimal hyperplane to divide data into two distinct classes [29].
- DT is a type of supervised learning classifier that operates without any predetermined parameters. It comprises internal nodes responsible for making decisions, while the outcome is depicted by the leaf nodes [29]. In the current study, we used the CART (Classification And Regression Tree) variants of a decision tree that uses a greedy approach to split the data at each node.
- Logistic regression is a popular algorithm utilized in supervised learning. It aims to estimate the probability and forecast the result of a categorical dependent variable by establishing a connection between independent variables and the dependent variable [30].
- MLP is a type of feedforward neural network, consisting of three layers: input, output, and hidden. It uses a linear activation function [30].
- AdaBoost, short for adaptive boosting, is a boosting technique derived from the boosting algorithm. Its objective is to merge several weak classifiers into a powerful classifier [30].
- XGBoost is a powerful approach to gradient boosting, which encompasses a range of machine learning algorithms. It combines several weak learning models, particularly decision trees, to create a high-performing and reliable predictive model [30].

A review of deep and classical ML methods for classification tasks was presented in [31].

B. Deep Learning Models

In the deep learning part of the proposed model, we exploited five methods (i.e., m = 5 in Fig. 3): dense, GRU, CNN, BiLSTM, and CNN-LSTM models. The details of these models are as follows.

- Dense: This type of deep model is commonly used in various deep learning tasks, such as image classification, natural language processing, and speech recognition [32]. In the current study, we implemented a dense model shown in Fig. 4. It contains five fully connected dense layers with sizes shown in the figure.
- CNN: This type of deep model is primarily used for image processing tasks, but it can also be applied to text classification tasks [33]. To this aim, CNNs can be used to extract meaningful features from textual data. In the current research, we used the CNN model shown in Fig. 5. Here, the model can learn to automatically extract relevant features from the text data and capture important patterns using the convolutional layer followed by Maxpooling which is used for dimensionality reduction.
- GRU: Gated Recurrent Unit (GRU) deep models are suitable for text classification due to their ability to capture sequential dependencies, handle variable-length inputs, and efficiently process text data [34]. In the current study, we used the GRU model shown

in Fig. 6. It can effectively capture the contextual information and dependencies between words in a sentence using three GRU layers.

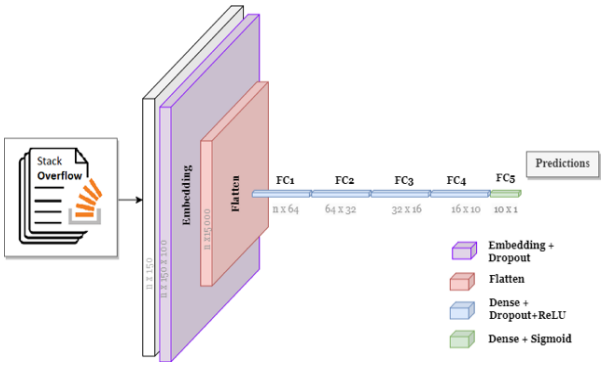


Fig. 4: The overall structure of the dense model.

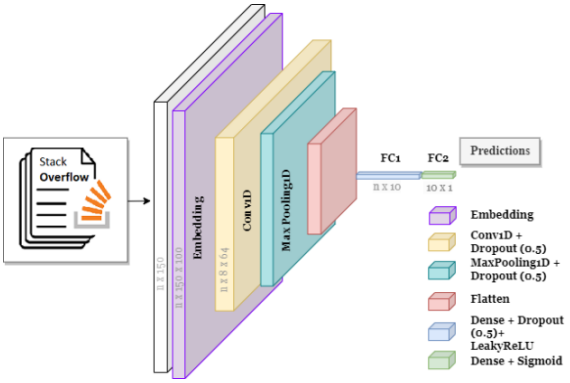


Fig. 5: The overall structure of the CNN model.

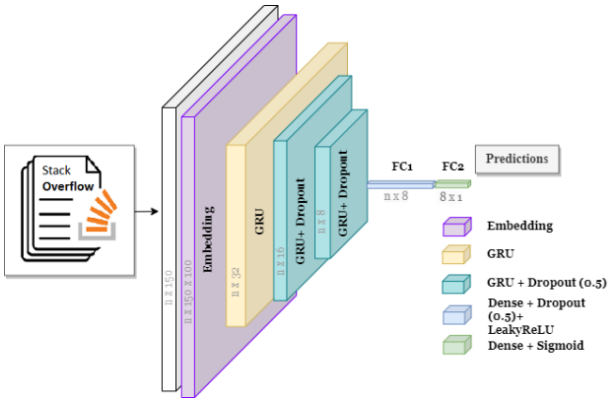


Fig. 6: The overall structure of the GRU model.

- CNN-LSTM: The combination of CNNs and Long Short-Term Memory (LSTM) networks is a popular approach for text classification tasks. This combination allows the model to capture both local and global dependencies in the text data [35]. In the current study, we used CNNs for feature extraction and LSTMs for sequence modeling as shown in Fig. 7. This allows for a more comprehensive understanding of the text data and can improve the accuracy of text classification tasks.
- BiLSTM: Bidirectional LSTM is a type of Recurrent Neural Network (RNN) that is commonly used for text classification tasks [36]. It is particularly effective in capturing contextual information from both past and future words in a sequence. In the current study, we used BiLSTM models as shown in Fig. 8.
-

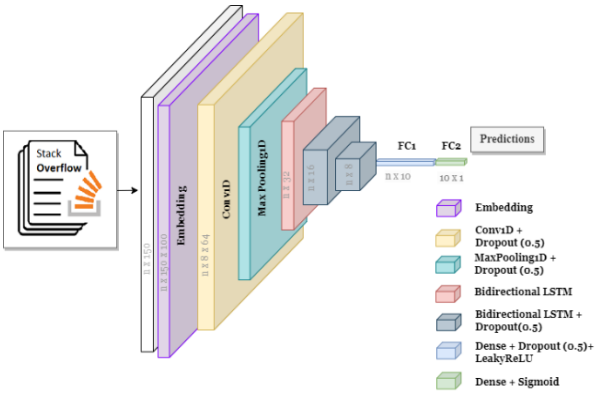


Fig. 7: The overall structure of the CNN-LSTM model.

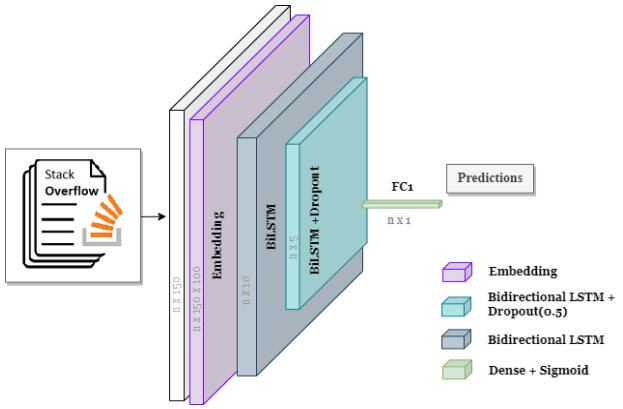


Fig. 8: The overall structure of the BiLSTM model.

C. *Bert Model*

The third level-0 classifier used in the proposed method is the pre-trained Bert classifier which is a transformer-based multi-layered encoder [37]. It uses an attention mechanism to learn the relationship between all words in a sentence. Specifically, it contains three embedding modules and 12 transformer layers each containing a dense layer and an attention layer. In the current study, we adopted the Huggingface[1] implementation of the Bert model. Bert has been previously used for text classification and a comparison of Bert and ML methods was provided in [38].

[1]https://huggingface.co/transformers/v2.10.0/model_doc/bert.html

D. Ensemble of Models

The main rationale behind employing the aforementioned three learning models in the present study lies in their utilization of different text features and their distinct mechanisms for generating predictions. This diversity holds great importance for meta-learning and stacking models [39]. In the field of data analysis, stacking is a useful technique that leverages the diverse predictions generated by base models to capture various aspects of the data, ultimately improving the accuracy of predictions. Each base model has its strengths and weaknesses, but by combining their predictions, the ensemble model can benefit from the collective expertise of these models [40]. Additionally, stacking helps to reduce bias and variance, leading to better accuracy by consolidating predictions from multiple models [41]. In addition, it assists in capturing detailed connections and patterns in the data that individual models might fail to notice. In our proposed model, we implemented a variant of the stack generalization technique, as depicted in Algorithm 1.

Algorithm 1: Pseudo-code for the stack generalization algorithm adopted from [15].

Data: Training dataset
$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$
Level-0 learning algorithms $\mathcal{L}_1, \cdots, \mathcal{L}_M$
Level-1 learning algorithms $\mathcal{L}$
Test dataset $\mathcal{X}' = \{x'_1, x'_2, \cdots, x'_T\}$
Result: Prediction vector $\mathcal{Y}' = \{y'_1, y'_2, \cdots, y'_T\}$

begin
  Randomly split $\mathcal{D}$ into I almost equal folds: $\mathcal{D}_1, \cdots, \mathcal{D}_I$
  $\mathcal{D}' = \varnothing$
  for $i = 1, \cdots, I$ **do**
    $\mathcal{D}^{-i} = \mathcal{D} - \mathcal{D}_i$
    $h = \varnothing$
    for $m = 1, \cdots, M$ **do**
      $h_m = \mathcal{L}_m(\mathcal{D}^{-i})$
    end
    $z = \varnothing$
    for $k = 1, \cdots, |\mathcal{D}^i|$ **do**
      $d = \varnothing$
      for $m = 1, \cdots, M$ **do**
        $d_m = h_m(\mathcal{D}^i_k[x])$
      end
      $z_k = (d, \mathcal{D}^i_k[y])$
    end
    $\mathcal{D}' = \mathcal{D}' \cup z$
  end
  $h' = \mathcal{L}(\mathcal{D}')$
  $\mathcal{Y}' = \varnothing$
  for $k = 1, \cdots, T$ **do**
    $z = \varnothing$
    for $m = 1, \cdots, M$ **do**
      $z_m = \mathcal{L}_m(x'_k)$
    end
    $\mathcal{Y}'_k = h'(z)$
  end
  return $\mathcal{Y}'$
end

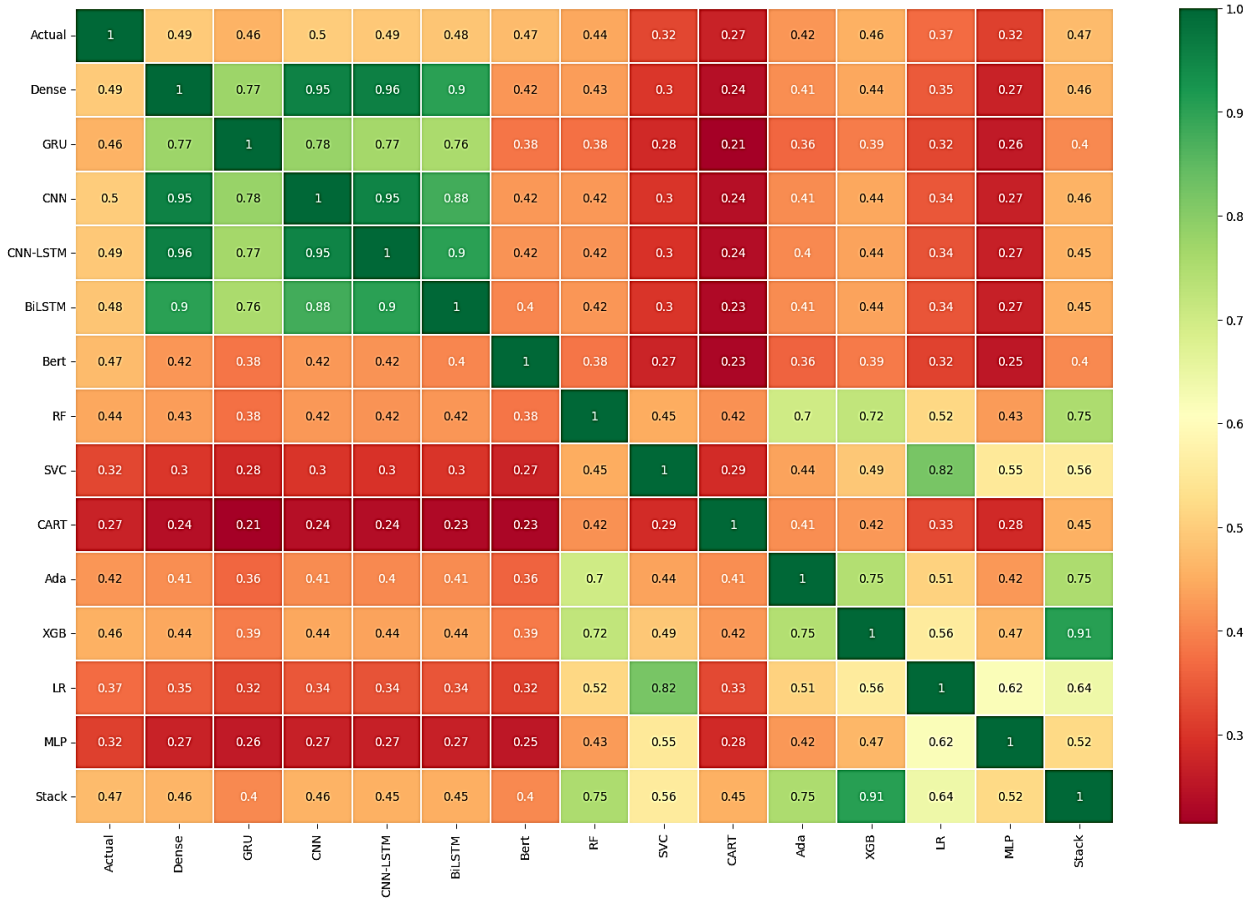


Fig. 9: Correlation between classical ML, BERT, and other deep learning models.

To showcase the effectiveness of the three different models used in the proposed method, we created a correlation chart for these methods, as shown in Fig. 9. As evident in the diagram, there is a strong correlation observed among the deep models, while the correlation between the deep and ML methods, as well as the Bert model, is relatively weaker. This suggests that combining the models in an ensemble can produce more precise outcomes as compared to using each learning method individually.

Experimental Settings

A. Compared Baselines

Our study aims to classify users' expertise based on their shape, which we frame as a binary text classification task. This differs from previous research that focused on expert identification and intern retrieval, such as the works of Gharebagh et al. [6] and Rostami and Neshati [1]. Hence, we cannot directly compare our findings with theirs due to the dissimilarity between the problems. To demonstrate the effectiveness of our proposed approach, we utilized seven deep-learning techniques that are commonly used for binary text classification:

- CRNN [42]: In this approach, every sentence is regarded as a region, and a regional CNN is utilized on the input word vectors. Subsequently, max pooling is employed to decrease the dimensionality of the local features. Finally, an LSTM layer is utilized to capture long dependencies, and a linear decoder is used to make predictions.
- IWV [43]: This model comprises three convolution layers, a max pooling layer, and a fully connected layer stacked sequentially for sentiment polarity classification.
- SS-BED [44]: This model utilizes two parallel LSTM layers on two distinct word embedding matrices to acquire knowledge about semantic and sentiment feature representations. The results obtained from the LSTM layers are then inputted into a fully connected network with one hidden layer to make the predictions.
- HAN [45]: This model comprises four essential components: a word sequence encoder, which is a bidirectional GRU, a word-level attention layer that calculates weighted sentence vectors, a sentence encoder, which is another bidirectional GRU, and a sentence-level attention layer that rewards sentences for making accurate classifications.
- ARC [46]: In this model, a one-layer bidirectional GRU is applied to the word vectors, and the outcomes are fed into an attention layer. The output of the attention mechanism is then passed through a CNN layer, followed by a max-pooling layer and a fully connected layer.
- AC-BiLSTM [47]: This model has a one-dimensional CNN layer consisting of CNNs of different filter sizes. This layer is employed for localized feature extraction. The output of the CNN layer is then fed into a bidirectional LSTM layer, followed by an attention mechanism. The output layer of this model consists of a dropout layer and a softmax layer.
- ABCDM [39]: This method utilizes a unique combination of two bidirectional LSTM and GRU layers to effectively capture contextual information from preceding and forthcoming contexts. This allows ABCDM to consider the sequential progression of information in both forward and backward directions. Additionally, ABCDM seamlessly integrates an attention mechanism within the bidirectional layers, allowing it to selectively emphasize specific words based on their varying levels of significance. Furthermore, ABCDM incorporates convolution and pooling mechanisms to reduce the complexity of features and extract localized features more efficiently.

B. Environment Setting

In our comparative analysis, we investigated the utility of the proposed model against the baseline learning models used in the proposed ensemble model as well as against seven state-of-the-art deep models described in the previous section. All the implementations were carried out using Tensorflow 2.14.0, Numpy, Sklearn, and Pandas in Python3 (version: 3.12), and Transformer (version: 4.36.2). All the models were implemented in the Google Colab environment with an Intel Xeon CPU accompanied by a 13 GB RAM, a Tesla K80 accelerator, and 12 GB GDDR5 VRAM.

C. Evaluation Criteria

To evaluate the effectiveness of models, we employed *Precision* (π), *recall* (ρ), *accuracy*, *F1*, and Area Under Curve (AUC) evaluation criteria in the experiments [39].

$$F1 = \frac{2 \times \pi \times \rho}{(\pi + \rho)} \tag{8}$$

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

$$\pi = \frac{TP}{TP + FP} \tag{10}$$

$$\rho = \frac{TP}{TP + FN} \tag{11}$$

$$AUC = \frac{\sum Rank(+) - \left(|+| \times \frac{|+|+1}{2}\right)}{|+|+|-|} \tag{12}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively and $\sum Rank(+)$ is the sum of the ranks of all positively classified samples, |+| and |-| are the number of positive and negative samples in the dataset, respectively.

Results

A. Preliminary Results

In our first round of experiments, we compared seven different machine learning methods and their ensemble (named as SG-ML) using the stack generalization method outlined in Algorithm 1 in Table 2.

Table 2: Comparison of results obtained using ML methods and their ensemble (SG-ML). Bold values indicate the best-performed models

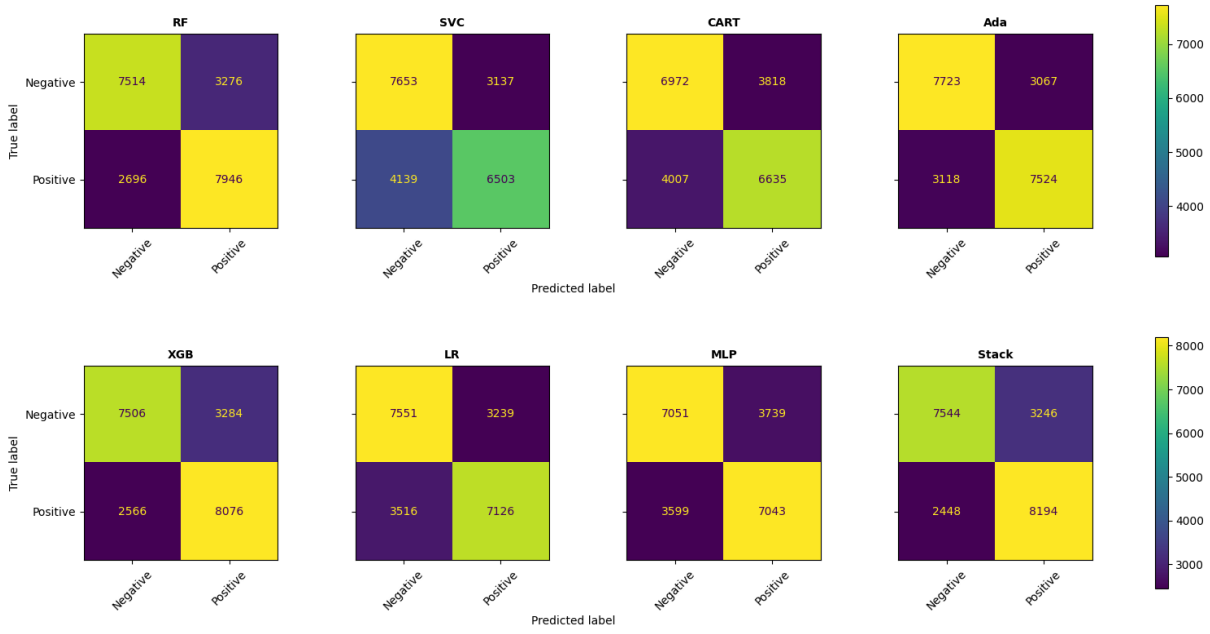
		Acc	π	ρ	F1
Android	RF	68.91	66.91	74.68	70.58
	SVM	62.54	63.97	57.19	60.39
	CART	61.49	61.50	61.18	61.34
	Ada	68.08	68.20	67.58	67.88
	XGB	68.58	67.81	70.58	69.17
	LR	64.77	65.19	63.20	64.18
	MLP	65.68	64.23	70.58	67.26
	SG-ML	**69.52**	**68.38**	**72.47**	**70.37**
C#	RF	73.28	70.74	78.43	74.39
	SVM	68.53	69.29	65.38	67.27
	CART	65.60	65.09	65.71	65.40
	Ada	73.04	72.39	73.56	72.97
	XGB	74.83	72.11	80.10	75.89
	LR	70.05	70.01	69.04	69.52
	MLP	66.69	66.01	67.38	66.68
	SG-ML	**75.32**	**72.84**	**79.92**	**76.21**
Java	RF	72.67	70.70	76.42	73.45
	SVM	65.55	66.52	61.16	63.72
	CART	64.82	64.41	64.59	64.50
	Ada	72.28	71.99	72.00	71.99
	XGB	72.94	71.11	76.29	73.61
	LR	67.29	67.40	65.63	66.50
	MLP	73.36	71.59	76.52	73.97
	SG-ML	**73.36**	**71.59**	**76.52**	**73.97**
All	RF	72.14	70.81	74.67	72.69
	SVM	66.05	67.46	61.11	64.13
	CART	63.49	63.47	62.35	62.91
	Ada	71.14	71.04	70.70	70.87
	XGB	72.70	71.09	75.89	73.41
	LR	68.48	68.75	66.96	67.84
	MLP	65.76	65.32	66.18	65.75
	SG-ML	**73.43**	**71.63**	**77.00**	**74.21**

Our analysis revealed that the RF, Ada, and XGB classifiers scored higher in terms of accuracy and F1 scores than the other methods. Additionally, the SG-ML model outperformed all level-0 models across all four datasets. We also compared the deep models and their ensemble using the same stack generalization method shown in Algorithm 1 in Table 3. The results indicate that the CNN model and the ensemble model achieved higher scores overall, but the differences between the individual model performances and their ensemble were less pronounced compared to the ML algorithms. This suggests that the variance of the deep models is lower than that of ML models.

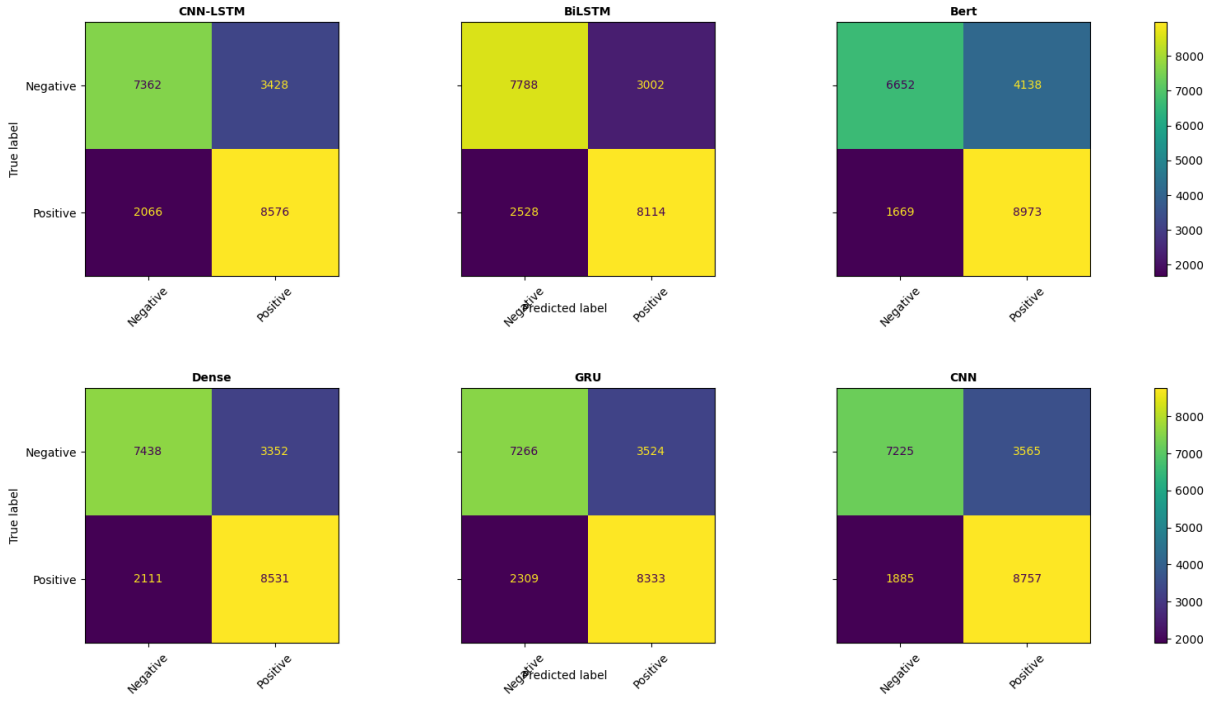
Table 3: Comparison of results obtained using deep methods and their ensemble (SG-Deep). Bold values indicate the best-performed models

		Acc	π	ρ	F1
Android	Dense	71.90	71.87	71.84	71.86
	GRU	71.92	71.07	**73.80**	72.41
	BiLSTM	**72.09**	70.67	75.41	**72.96**
	CNN	71.83	70.40	75.20	72.72
	CNN-LSTM	71.50	**73.53**	67.05	70.14
	SG-Deep	72.02	72.10	72.00	72.02
C#	Dense	76.58	73.71	81.84	77.56
	GRU	76.48	72.10	85.56	78.25
	BiLSTM	75.93	72.03	83.94	77.53
	CNN	**76.74**	72.64	**85.00**	**78.34**
	CNN-LSTM	76.08	**73.89**	79.89	76.77
	SG-Deep	76.64	77.42	76.64	76.52
Java	Dense	75.02	71.77	**81.61**	76.37
	GRU	74.83	72.13	80.06	75.89
	BiLSTM	73.15	69.20	82.44	75.24
	CNN	74.81	71.17	82.52	**76.43**
	CNN-LSTM	74.39	73.77	74.87	74.31
	SG-Deep	**75.13**	**75.55**	75.13	75.07
All	Dense	74.51	71.79	80.16	75.75
	GRU	72.78	70.28	78.30	74.07
	BiLSTM	74.20	72.99	76.25	74.58
	CNN	**74.57**	71.07	**82.29**	**76.27**
	CNN-LSTM	74.37	71.44	80.59	75.74
	SG-Deep	74.54	**74.99**	74.54	74.46

To provide more detailed information about the performance of different models on positive and negative classes, we have presented the confusion matrix of the All dataset in Fig. 10. We obtained similar results for the other three datasets, but we could not show them due to space limitations. As indicated in the figure, the XGB method had the best true positive result among the ML methods, while the Ada method had the best true negative. The Ada and XGB methods had the best false positive and false negative results, respectively, highlighting their effectiveness for classification tasks. Among deep models, the CNN model provided the best true positives, while the BiLSTM model provided the best true negatives. The BiLSTM and Bert models had the best false positives and false negatives, respectively.



(a)

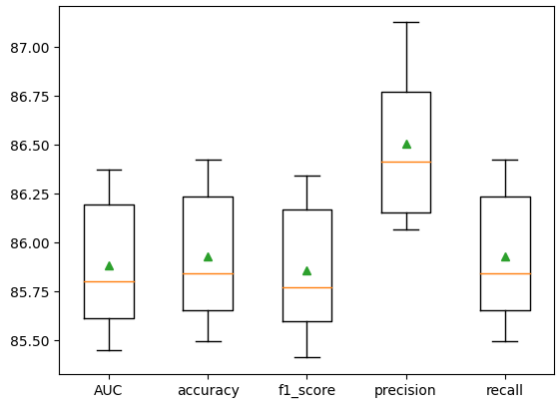


(b)

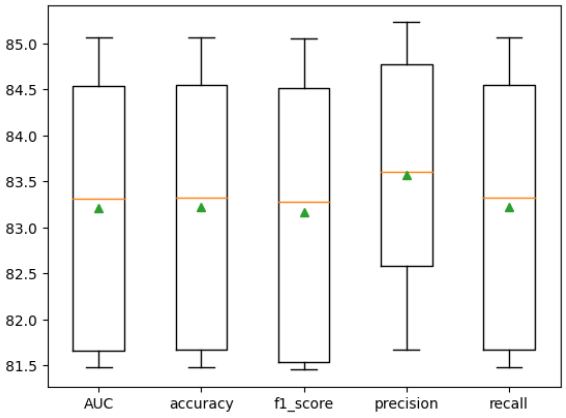
Fig. 10: Confusion matrix for (a) classical ML, and (b) Bert and other deep learning models on the All dataset.

B. *Main Results*

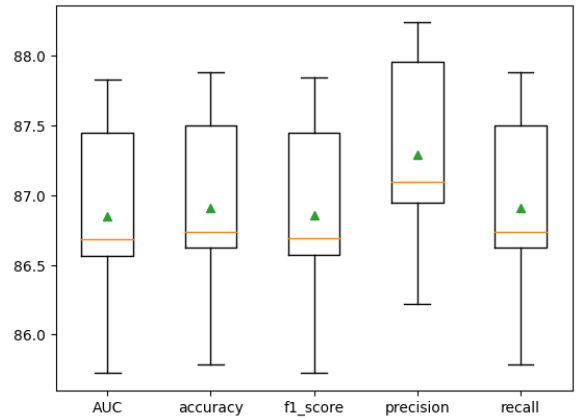
In the second part of our experiments, we presented the performance of our proposed model through 5-fold cross-validation, as displayed in Fig. 11 as box plots. The results indicate that the model's performance on the Android dataset is comparatively lower than that of other datasets, possibly due to the nature of the texts in this dataset or its fewer records when compared to other datasets (refer to Table 1 for more information).



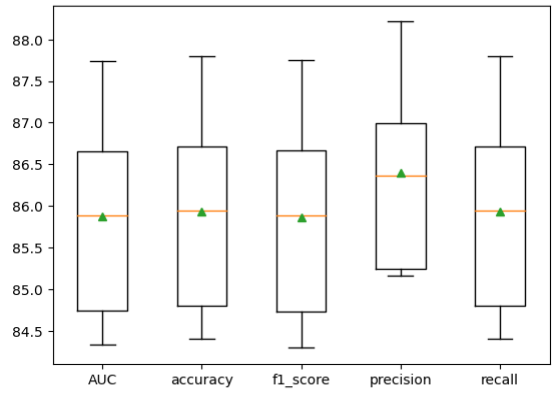
(a)



(b)



(c)



(d)

Fig. 11: Comparison of the results obtained using the proposed method with 5-fold cross-validation on the (a) All, (b) Android, (c) C#, and (d) Java datasets.

Table 4: Comparison of results obtained using the proposed model and state-of-the-art binary text classification models. Bold values indicate the best-performed models

		Acc	π	ρ	F1
Android	SS-BED	74.36	71.24	81.12	75.86
	ACBiLSM	71.99	67.15	85.32	75.15
	IWV	67.24	67.18	66.51	66.84
	HAN	64.30	64.03	64.12	64.08
	CRNN	74.17	70.15	83.53	76.25
	ARC	74.53	71.58	80.78	75.90
	ABCDM	74.36	71.24	81.12	75.86
	Proposed	**85.93**	**86.51**	**85.93**	**85.86**
C#	SS-BED	75.32	71.03	84.64	77.24
	ACBiLSM	72.00	70.48	74.66	72.51
	IWV	71.40	70.06	73.69	71.83
	HAN	66.78	65.25	70.27	67.67
	CRNN	76.06	70.87	87.61	78.36
	ARC	76.23	71.34	86.84	78.33
	ABCDM	66.87	64.48	73.53	68.71
	Proposed	**86.91**	**87.29**	**86.91**	**86.85**
Java	SS-BED	73.35	70.77	78.59	74.47
	ACBiLSM	70.19	68.41	73.84	71.02
	IWV	68.48	66.78	72.21	69.39
	HAN	63.61	62.11	67.82	64.84
	CRNN	73.77	69.23	84.56	76.13
	ARC	74.30	69.69	85.05	76.61
	ABCDM	65.14	64.81	64.64	64.73
	Proposed	**85.93**	**86.40**	**85.93**	**85.87**
All	SS-BED	75.32	71.03	84.64	77.24
	ACBiLSM	72.00	70.48	74.66	72.51
	IWV	71.40	70.06	73.69	71.83
	HAN	66.78	65.25	70.27	67.67
	CRNN	76.06	70.87	87.61	78.36
	ARC	76.23	71.34	86.84	78.33
	ABCDM	66.87	64.48	73.53	68.71
	Proposed	**86.91**	**87.29**	**86.91**	**86.85**

Furthermore, we compared our proposed model's performance with other state-of-the-art binary text classification methods mentioned above, and the results are shown in Table 4. The proposed model outperforms all the other models significantly. Interestingly, the CRNN, ARC, and ABCDM models, which all utilize convolutional layers in their architecture, delivered the best results, as shown in Table 3 for comparison of deep models.

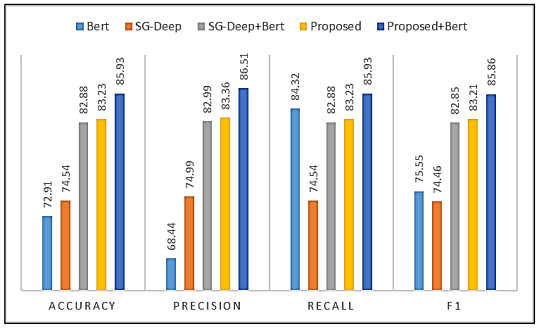
C. Ablation Study

To demonstrate the effectiveness of our proposed model, we carried out an ablation study. This involved eliminating different components of the model and evaluating the performance of the resulting models. We compared the performance of five models, as shown in Fig. 12. The first model, called Bert, only included the Bert branch of our proposed model and omitted the ML and Deep branches. The second model, SG-Deep, only included the deep learning branch and omitted the Bert and ML branches. The third model, SG-Deep+Bert, preserved the Bert and Deep branches but omitted the ML branch. The fourth model, Proposed, only preserved the ML and Deep branches and omitted the Bert branch. The fifth and final model, Proposed+Bert, utilized all three branches of our proposed model.

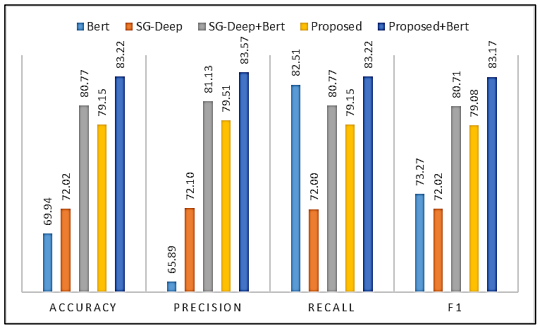
As shown in the figure, the performance of the Bert and SG-Deep models was significantly lower than the other models. However, the proposed model and SG-Deep+Bert models had similar performance, which demonstrates the effectiveness of the ensemble technique used in our proposed model. The diversity of algorithms in SG-Deep and the structural differences between deep models and Bert models make the ensemble results more accurate. Finally, the Proposed+Bert model achieved the best performance in all datasets, showing the effectiveness of using all three branches of our proposed model.

Conclusion

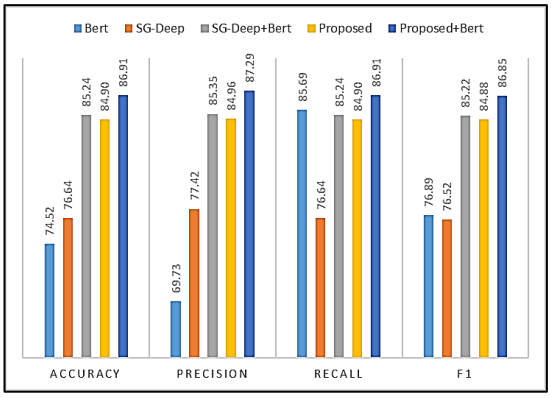
Our study proposes an ensemble model that combines deep learning and machine learning methods to detect the expertise shape of users based on their answers in Stack Overflow's CQA. To achieve this, we used seven ML models, five deep models, and a pre-trained transformer-based Bert model. Our model was able to process user answers and identify dash-shaped users. We conducted extensive experiments to evaluate our model's effectiveness, and the results across four different datasets of Stack Overflow answers demonstrate that our model outperforms both the ML and deep models used as its building blocks, as well as state-of-the-art deep models for binary classification of textual data. Our model is not limited to detecting dash-shaped users. It can also classify other shapes of expertise, such as T- and C-shaped users, which are valuable for forming agile software teams.



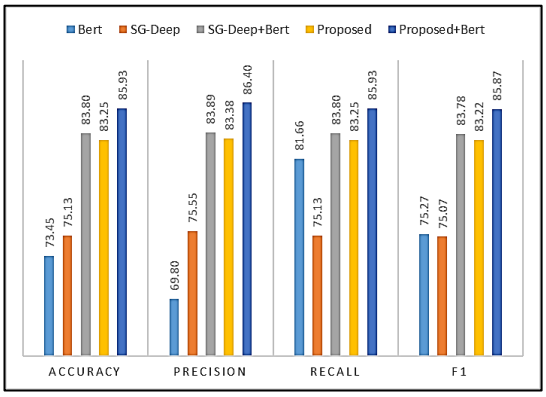
(a)



(b)



(c)



(d)

Fig. 12: Comparison of the performance of five models in the ablation study on the (a) All, (b) Android, (c) C#, and (d) Java datasets.

Additionally, our model can be used as a filter method for downstream applications, like intern recommendations. In future work, we plan to evaluate our model on similar problems in CQA texts and explore other deep ensemble models to further improve the performance of expertise shape classification problems.

Author Contributions

S. Nemati designed the experiments, analyzed the data, interpreted the results, and wrote the manuscript.

Acknowledgment

This work was financially supported by the research deputy of Shahrekord University (grant number 0GRD34M44264).

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

CQA	Community Question-Answering
BiLSTM	BiDirectional Long Short-Term Memory
RF	Random Forest
SVM	Support Vector Classifier
DT	Decision Tree
LR	Logistic Regression
MLP	Multi-Layer Perceptron
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
AUC	Area Under Curve
CART	Classification And Regression Tree
ML	Machine Learning

References

[1] P. Rostami, M. Neshati, "Intern retrieval from community question answering websites: A new variation of expert finding problem," Expert Syst. Appl., 181: 115044, 2021.

[2] S. Yuan, Y. Zhang, J. Tang, W. Hall, J. B. Cabotà, "Expert finding in community question answering: a review," Artif. Intell. Rev., 53: 843-874, 2020.

[3] A. Dargahi Nobari, S. Sotudeh Gharebagh, M. Neshati, "Skill translation models in expert finding," in Proc. the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval: 1057-1060, 2017.

[4] H. Demirkan, J. Spohrer, "T-shaped innovators: Identifying the right talent to support service innovation," Res. Technol. Manage., 58(5): 12-15, 2015.

[5] V. Kumar, N. Pedanekar, "Mining shapes of expertise in online social Q&A communities," in Proc. the 19th ACM conference on Computer Supported Cooperative Work and Social Computing Companion: 317-320, 2016.

[6] S. S. Gharebagh, P. Rostami, M. Neshati, "T-shaped mining: A novel approach to talent finding for agile software teams," in Proc. Advances in Information Retrieval: 40th European Conference on IR Research: 411-423, 2018.

[7] C. P. Maertz Jr, P. A. Stoeberl, J. Marks, "Building successful internships: lessons from the research for interns, schools, and employers," Career Dev. Int., 19(1): 123-142, 2014.

[8] X. Fu, X. Sun, H. Wu, L. Cui, J. Z. Huang, "Weakly supervised topic sentiment joint model with word embeddings," Knowl. Based. Syst., 147: 43-54, 2018.

[9] H. Wang, K. Guo, "The impact of online reviews on exhibitor behaviour: evidence from movie industry," Enterp. Inf. Syst., 11(10): 1518-1534, 2017.

[10] D. Kundu, D. P. Mandal, "Formulation of a hybrid expertise retrieval system in community question answering services," Appl. Intell., 49: 463-477, 2019.

[11] S. Sorkhani, R. Etemadi, A. Bigdeli, M. Zihayat, E. Bagheri, "Feature-based question routing in community question answering platforms," Inf. Sci. (N Y), 608: 696-717, 2022.

[12] X. Zhang et al., "Temporal context-aware representation learning for question routing," in Proc. the 13th International Conference on Web Search and Data Mining: 753-761, 2020.

[13] H. Ding, Q. Liu, G. Hu, "TDTMF: A recommendation model based on user temporal interest drift and latent review topic evolution with regularization factor," Inf. Process. Manage., 59(5): 103037, 2022.

[14] P. Rostami, A. Shakery, "A deep learning-based expert finding method to retrieve agile software teams from CQAs," Inf. Process. Manage., 60(2): 103144, 2023.

[15] K. Balog, L. Azzopardi, M. de Rijke, "A language modeling framework for expert finding," Inf. Process. Manage., 45(1): 1-19, 2009.

[16] S. Liang, M. de Rijke, "Formal language models for finding groups of experts," Inf. Process. Manage., 52(4): 529-549, 2016.

[17] D. Petkova, W. B. Croft, "Hierarchical language models for expert finding in enterprise corpora," Int. J. Artif. Intell. Tools, 17(01): 5-18, 2008.

[18] M. Bouguessa, B. Dumoulin, S. Wang, "Identifying authoritative actors in question-answering forums: the case of yahoo! answers," in Proc. the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 866-874, 2008.

[19] H. Zhu, E. Chen, H. Xiong, H. Cao, J. Tian, "Ranking user authority with relevant knowledge categories for expert finding," World Wide Web, 17: 1081-1107, 2014.

[20] A. Daud, J. Li, L. Zhou, F. Muhammad, "Temporal expert finding through generalized time topic modeling," Knowl. Based Syst., 23(6): 615-625, 2010.

[21] S. Momtazi, F. Naumann, "Topic modeling for expert finding using latent Dirichlet allocation," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 3(5): 346-353, 2013.

[22] L. Yang et al., "Cqarank: jointly model topics and expertise in community question answering," in Proc. the 22nd ACM International Conference on Information & Knowledge Management: 99-108, 2013.

[23] N. Nikzad-Khasmakhi, M. Balafar, M. R. Feizi-Derakhshi, C. Motamed, "ExEm: Expert embedding using dominating set theory with deep learning approaches," Expert Syst. Appl., 177: 114913, 2021.

[24] M. Zhao, F. Javed, F. Jacob, M. McNair, "SKILL: A system for skill identification and normalization," in Proc. the AAAI Conference on Artificial Intelligence: 4012-4017, 2015.

[25] A. Azzam, N. Tazi, A. Hossny, "Text-based question routing for question answering communities via deep learning," in Proc. the Symposium on Applied Computing: 1674-1678, 2017.

[26] M. Dehghan, H. A. Rahmani, A. A. Abin, V. V. Vu, "Mining shape of expertise: A novel approach based on convolutional neural network," Inf. Process. Manage., 57(4): 102239, 2020.

[27] Z. Li, J. Y. Jiang, Y. Sun, W. Wang, "Personalized question routing via heterogeneous network embedding," in Proc. the AAAI Conference on Artificial Intelligence: 192-199, 2019.

[28] W. Tang, T. Lu, D. Li, H. Gu, N. Gu, "Hierarchical attentional factorization machines for expert recommendation in community question answering," IEEE Access, 8: 35331-35343, 2020.

[29] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," Artif. Intell. Rev., 52(1): 273-292, 2019.

[30] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, "Text classification algorithms: A survey," Information, 10(4): 150, 2019.

[31] Q. Li et al., "A survey on text classification: From shallow to deep learning," arXiv preprint arXiv:2008.00364, 2020.

[32] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, "Deep learning–based text classification: a comprehensive review," ACM Comput. Surv. (CSUR), 54(3): 1-40, 2021.

[33] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.

[34] S. Yu, D. Liu, W. Zhu, Y. Zhang, S. Zhao, "Attention-based LSTM, GRU and CNN for short text classification," J. Intell. Fuzzy Syst., 39(1): 333-340, 2020.

[35] M. Zulqarnain et al., "Text classification using deep learning models: A Comparative review," Cloud Comput. Data Sci., 5(1): 80-96, 2024.

[36] A. Ezen-Can, "A comparison of LSTM and BERT for small corpus," arXiv preprint arXiv:2009.05451, 2020.

[37] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[38] S. González-Carvajal, E. C. Garrido-Merchán, "Comparing BERT against traditional machine learning text classification," arXiv preprint arXiv:2005.13012, 2020.

[39] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," Future Gener. Comput. Syst., 115: 279-294, 2021.

[40] S. Nemati, "Canonical correlation analysis for data fusion in multimodal emotion recognition," in Proc. 9th International Symposium on Telecommunication: With Emphasis on Information and Communication Technology, IST 2018, 2019.

[41] A. Mohammed, R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," J. King Saud Univ. Comput. Inf. Sci., 35(2): 754-774, 2023.

[42] J. Wang, L. C. Yu, K. R. Lai, X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in Proc. the 54th Annual Meeting of the Association for Computational Linguistics, (2: Short papers): 225-230, 2016.

[43] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," Expert Syst. Appl., 117: 139-147, 2019.

[44] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, P. Agrawal, "Understanding emotions in text using deep learning and big data," Comput. Human Behav., 93: 309-317, 2019.

[45] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, "Hierarchical attention networks for document classification," in Proc. the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: 1480-1489, 2016.

[46] S. Wen, J. Li, "Recurrent convolutional neural network with attention for twitter and yelp sentiment classification: ARC model for sentiment classification," in Proc, the 2018 International Conference on Algorithms, Computing and Artificial Intelligence: 1-7, 2018.

[47] G. Liu, J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," Neurocomput., 337: 325-338, 2019.

Biographies

Shahla Nemati was born in Shiraz, Iran, in 1982. She received the B.S. degree in Hardware Engineering from Shiraz University, Shiraz, in 2005, the M.S. degree from the Isfahan University of Technology, Isfahan, Iran, in 2008, and the Ph.D. degree in Computer Engineering from Isfahan University, Isfahan, in 2016. Since 2017, she has been an Assistant Professor with the Computer Engineering Department, Shahrekord University, Shahrekord, Iran. She has written several articles in the fields of data fusion, emotion recognition, affective computing, and audio processing. Her research interests include data fusion, affective computing, and data mining.

- Email: s.nemati@sku.ac.ir
- ORCID: 0000-0003-2906-5871
- Web of Science Researcher ID: AAA-3341-2019
- Scopus Author ID: 24512475100
- Homepage: https://www.sku.ac.ir/~snemati#

How to cite this paper:
S. Nemati, "An effective ensemble of deep and machine learning methods for classifying the expertise shape of CQA users," J. Electr. Comput. Eng. Innovations, 12(2): 409-424, 2024.

DOI: 10.22061/jecei.2024.10621.724

URL: https://jecei.sru.ac.ir/article_2097.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Facial Expression Recognition through Suboptimal Filter Design Using a Metaheuristic Kidney Algorithm

E. Ghasemi [1], S. M. Razavi [1], S. Mohamadzadeh [1,], M. Taghipour-Gorjikolaie [2]*

[1] *Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.*

[2] *School of Engineering, London South Bank University, London, UK.*

Article Info

Article History:
Received 06 January 2024
Reviewed 18 March 2024
Revised 10 April 2024
Accepted 30 April 2024

Keywords:
Suboptimal filter
Kidney algorithm
Nearest neighbor classification
Neural network
Filter bank
Facial expression recognition

[*]Corresponding Author's Email Address:
s.mohamadzadeh@birjand.ac.ir

Abstract

Background and Objectives: The recognition of facial expressions using metaheuristic algorithms is a research topic in the field of computer vision. This article presents an approach to identify facial expressions using an optimized filter developed by metaheuristic algorithms.

Methods: The entire process of feature extraction hinges on using a filter suboptimally configured by metaheuristic algorithms. Essentially, the purpose of utilizing this metaheuristic algorithm is to determine the suboptimal weights for feature extraction filters. Once the suboptimal weights for the filter have been determined by the metaheuristic algorithm, suboptimal filter sizes have also been determined. As an initial step, the k-nearest neighbor classifier is employed due to its simplicity and high accuracy. Following the initial stage, a final model is presented, which integrates results from both filterbank and Multilayer Perceptron neural networks.

Results: An analysis of the existing instances in the FER2013 database has been conducted using the method proposed in this article. This model achieved a recognition rate of 78%, which is superior to other algorithms and methods while requiring less training time than other algorithms and methods**.**In addition, the JAFFE database, a Japanese women's database, was utilized for validation. On this dataset, the proposed approach achieved a 94.88% accuracy rate, outperforming other competitors.

Conclusion: The purpose of this article is to propose a method for improving facial expression recognition by using an optimized filter, which is implemented through a metaheuristic algorithm based on the KA. In this approach, optimized filters were extracted using the metaheuristic algorithms kidney, k-nearest neighbor, and multilayer perceptron. Additionally, by employing this approach, the suboptimal size and number of filters for facial state recognition were determined in order to achieve the highest level of accuracy in the extraction process.

Introduction

This principle emphasizes that every state (words, silence, activity) in human interaction carries a meaningful value, and that communication plays an important role in establishing civil society, according to Watzlawick, a psychologist [1].Generally, communication is composed of 7% verbal communication, 38% para-verbal communication (such as tone analysis), and 55% nonverbal communication (such as facial expressions, gestures, eye contact, etc.) Therefore, nonverbal communication is significant to many aspects of our daily lives and to the effectiveness of human interaction in

general [2].

A positive emotion, such as trust and agreement, is more likely to develop than a negative emotion, such as fear, mistrust, etc [3]. Consequently, facial expressions are the main means by which we communicate our emotions to the external world and interact with others, and we assess how others engage with us through them. Computers are capable of automatically recognizing facial expressions and emotions [4]. A rich and robust human-computer interaction is contingent upon automatic emotion recognition systems, namely machines and devices that can recognize mood. Recent years have seen a growing interest in integrating emotion recognition into modern human-computer interfaces. Research in this field has also found applications in a variety of fields, including animation, medicine, and security [5]. Robot-assisted therapy for children with Autism Spectrum Disorders (ASD) [6] or for elderly patients with dementia has gained attention from researchers in the field of healthcare [7], [8]. In recent years, humanoid and pet robots have become increasingly popular for therapeutic purposes, including companionship, sports coaching, and daily life assistance [9].

As research progresses in this area, machine learning techniques are being used in order to enable such robots to increasingly engage and read patient states. As a result, while efficient facial recognition systems have been developed, systems that can automatically detect emotions have not been developed [10], [11]. In order to recognize facial expressions of individuals, multiple methods have been proposed. A number of advances have been made in the field of facial recognition based on machine learning approaches as facial features are extremely complex. The extraction of high-dimensional and suboptimal features, as well as the classification of the data, becomes feasible through the utilization of heuristic algorithms and machine learning. An innovative machine learning model has been proposed to accurately and rapidly detect facial expressions in this study.

Related Work

Since their introduction in the late 1990s, Convolutional Neural Networks (CNNs) have demonstrated considerable potential in image processing [12]. An ordinary CNN typically consists of a convolutional layer, a hidden layer, and a fully connected layer. This ensures efficiency in processing static images in computer vision. In the past, the utilization of CNN was limited due to the scarcity of training data and computational power. After 2010, the exponential growth in computational power and the availability of larger datasets have transformed CNNs into highly suitable tools for feature extraction and image classification

Various techniques have been proposed for enhancing performance. For instance, the sigmoid activation function has been replaced by the rectified linear unit (ReLU) activation function to address issues such as gradient explosion and facilitate training acceleration. Various methods, such as average pooling and maximum pooling, are employed for sampling inputs and aiding in generalization [14]. Regularization and data augmentation are used to prevent overfitting. Batch normalization is employed to mitigate the issue of vanishing gradients [15].

The advancements and extensive research in Convolutional Neural Networks (CNNs) have rendered them highly desirable tools for addressing tasks in image processing, pattern recognition, and feature extraction. When a large dataset of facial expressions, known as the FER2013 dataset, was introduced at ICML in 2013, various types of CNNs achieved significant classification accuracies ranging from 65% to 72.7% [16]-[24]. For instance, Liu et al. trained three separate CNN models and grouped them to enhance their performance [16]. Minaei et al. employed an end-to-end deep learning framework utilizing an attention-based correlation network [17]. Tang et al. replaced the SoftMax layer with a Support Vector Machine in a deep neural network [22]. Shi et al. proposed a new modification representation module (ARM) to replace the hidden layer [23]. Pramerdorfer et al. compared the performance of three different architectures, VGG, Inception, and Resnet, through a comprehensive analysis [24].

Liu et al. ensembled three CNNs and demonstrated that multiple groups enhance the performance of different models [16]. Pramerdorfer et al. ensemble eight CNNs together and achieved improved performance [24].

Proposed Methodology

According to the methodology employed in this paper, facial expressions are classified into seven distinct groups using the FER2013 database. Following the block diagram depicted in Fig. 1, the proposed method consists of carrying out the following stages of facial expression recognition step-by-step.

A. Dataset Preprocessing

Data preprocessing is one of the fundamental stages in facial expression recognition, and its proper implementation significantly aids in improving the recognition rate. As the starting point in this article, the FER2013 database is invoked. It consists of 28,709 training and 3,589 testing and evaluation data with gray-scale backgrounds, which have been normalized by reading them. A series of seven emotional states have been used to categorize the images in the database.

B. Face Detection and Extraction

As a first step, all images have been analyzed using the MTCNN algorithm, which has been applied to detect and crop faces in a 32x32 pixel measurement area [25].

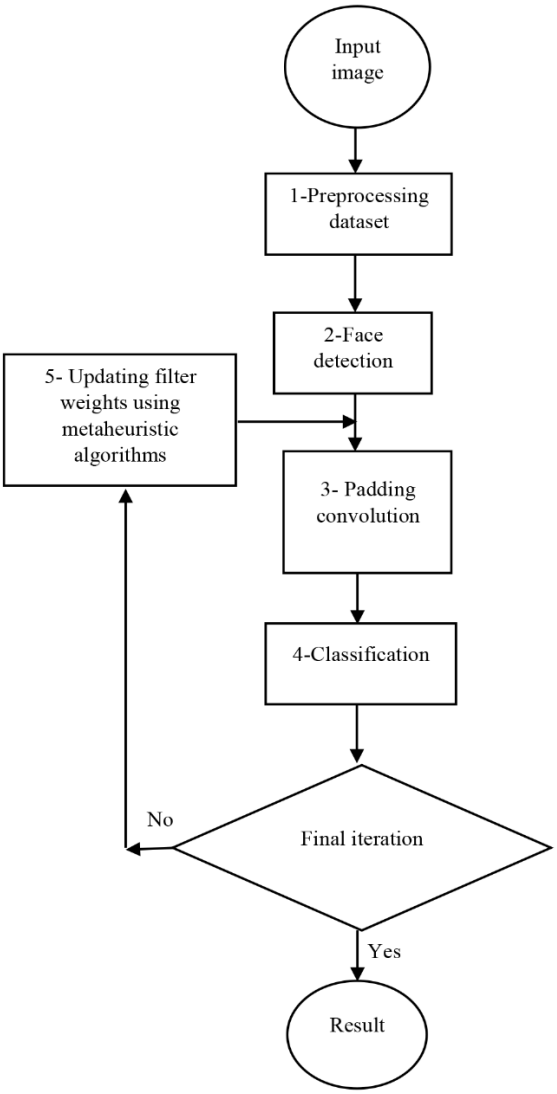


Fig. 1: Proposed methodology structure.

C. Application of Convolution Filters

A kernel can be processed over an image using two convolution methods: Valid Padding Convolution and Same Padding Convolution. We use Valid Padding Convolution in this article to minimize the amount of computational time required. Fig. 2 show that Scaling with Pooling.

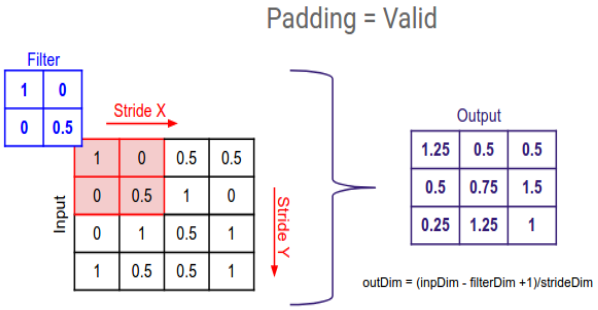


Fig. 2: Scaling with pooling.

D. Classification

In this article, it is proposed to start with the k-nearest neighbor (k-NN) classification with a k-value of 3. After the results from single filters and filter banks have been obtained, the neural network model, known as MLP (Multilayer Perceptron), which is a powerful classifier, is utilized. Furthermore, the Kidney Algorithm (KA), which belongs to the class of newer metaheuristic algorithms, is employed. The objective is to use the Kidney Algorithm (KA) to extract a single suboptimal filter or filter bank, thereby enhancing the recognition rate. As a result, by applying the filters, new features are extracted to determine how the output changes. Initially, KNN was employed, and when the best result for the nearest neighbor classifier and the overall algorithm for single-filter were obtained, a metaheuristic algorithm was utilized to extract a filter bank using an MLP neural network. In this article, employing the metaheuristic algorithm, four suboptimal single filters with sizes of 3×3, 5×5, 7×7, and 9×9 (pixels) have been extracted, and the recognition results on the data have been compared. The same method was then used to extract a filter bank, and the final stage involved implementing an MLP neural network to enhance recognition.

E. Updating Filter Weights Using Metaheuristic Algorithms

The filter weights can be related through an integral equation. This section illustrates how filter weights function as pattern recognition adaptors. Convolutional filters are commonly described based on the pattern-matching operator. Take note that, in this context, f(x) represents the pixel values or the values of lower layers of the network. The higher the f(x) value near a point, the higher the convolution value because discrete convolution is equivalent to a point-wise multiplication between the filter weights and the underlying filter values. In geometric terms, the dot product operator represents a scaled measure of vector similarity. The output of a focused convolution receives two vectors: one representing weights, the other input values. The respective values are then multiplied together and added together. This is comparable to the operation performed by the dot product operator. The magnitude of one or both vectors in the dot product can be increased by increasing their magnitudes. However, when the magnitudes are constant, the maximum value of the dot product is obtained when the vectors are in the same direction [26].

The self-adaptive filter, based on the convolution operator, aids in extracting features such as edges, blurring, and sharpening. Numerous filters are at our disposal for feature extraction, including edge detection, blurring, and sharpening. Furthermore, it has been

demonstrated that the matrix numbers of the filters undergo changes based on their respective operations. The filter extracts the main feature of the original image. Thus, our objective is to determine the suboptimal weights for the filter, ensuring suboptimal recognition results. We aspire to extract a suboptimal filter, capable of extracting the best features from the image. Fig. 3 illustrates the structure of a convolutional filter [27].

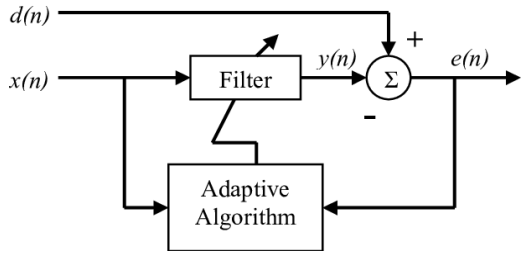


Fig. 3: Block diagram of adaptive filter [27].

F. Meta-heuristic Algorithm of the Kidney

KA is a population-based metaheuristic approach sharing certain characteristics with other population-based algorithms. Additionally, as its name implies, it emulates certain renal system methods. The primary components of the kidney process are described here. In the initial stage of KA, a random population of selected solutions is generated, and the objective performance is calculated for all of them. In each iteration, by moving towards the best-found solution so far, a new solution is generated for all candidate solutions. Then, utilizing the filtering operator, high-quality candidate solutions are directed to the FB filter repository within the population, while the remaining solutions are transferred to W. By examining the mechanisms of absorption, secretion, and excretion in the overall biological process of the kidney, the stages of the search are simulated. If a candidate solution is assigned to W in some embedded conditions of the algorithm, the algorithm provides another opportunity for this solution to improve itself before it can be transferred to FB. If this opportunity is not satisfactory, the solution is discarded from W, and a random solution is added to W. On the other hand, if, after filtration, a candidate solution is assigned to FB and the quality of this solution is not better than the worst solution found, it is discarded from FB. However, if the solution is better than the worst one, the worst solution will be discarded from FB. Eventually, the available solutions in FB are ranked, and the best solution is updated. The level of filtration is updated, and FB and W are merged. This iterative process continues until the termination criterion is achieved. The block diagram of the overall algorithm is presented in Fig. 4 [28].

Considering that the primary objective of using metaheuristic algorithms is optimization, these algorithms have been applied as a proposed approach, in which the coefficients of the extracted filters responsible for facial state recognition are optimized by the metaheuristic algorithm. These optimized coefficients obtained by metaheuristic algorithms are the primary reason for the superior recognition rate when compared to other methods. According to the results of various experiments, the size of the filters and the suboptimal number of filters were selected in the proposed method for feature extraction and achieving the superior recognition rate. In order to optimize the proposed method in various aspects, a number of optimization techniques have been employed.

Results and Discussion

A. Dataset

The FER2013 database consists of facial images with corresponding emotions, developed by Ian Goodfellow and colleagues [29]. This collection consists of 35,887 images in 8-bit grayscale format, with a size of 48×48 pixels, depicting facial emotions. The dataset is divided into three categories: 28,709 training data, 3,589 testing data, and 3,589 validation data.

All images in the database are labeled in such a way that each image falls into one of the seven main categories of facial emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

In regards to the distribution of emotions, 35,887 images are divided as follows: anger with 4,953 images, disgust with 547 images, fear with 5,121 images, joy with 8,989 images, sadness with 6,077 images, surprise with 4,002 images, and neutrality with 9,816 images. Fig. 5(a) illustrates a sample of images from the FER2013 database. In recent years, due to the popularity of FER2013 and its free accessibility, various methods have been employed to develop high-accuracy models for facial emotion recognition.

JAFFE is a grayscale dataset collected from psychological experiments. Expression is a mixture of different facial emotions collected in a laboratory-controlled environment. This dataset comprises 213 images of seven different facial expressions. Fig. 5(b) shows seven images from the dataset.

B. Performance Parameters

One of the most crucial steps following the construction and design of a model is its evaluation and analysis. After analysis, the results can be categorized into two groups: positive and negative. Subsequently, the quality of the algorithm can be assessed using relevant indicators. After analysis, the data can be divided into four groups in terms of categorization. TP (true positive.e., positive and classified as positive), FP (false positive i.e., negative but classified as positive), TN (true negative i.e., negative and classified as negative), and FN (false negative i.e., positive but classified as negative).

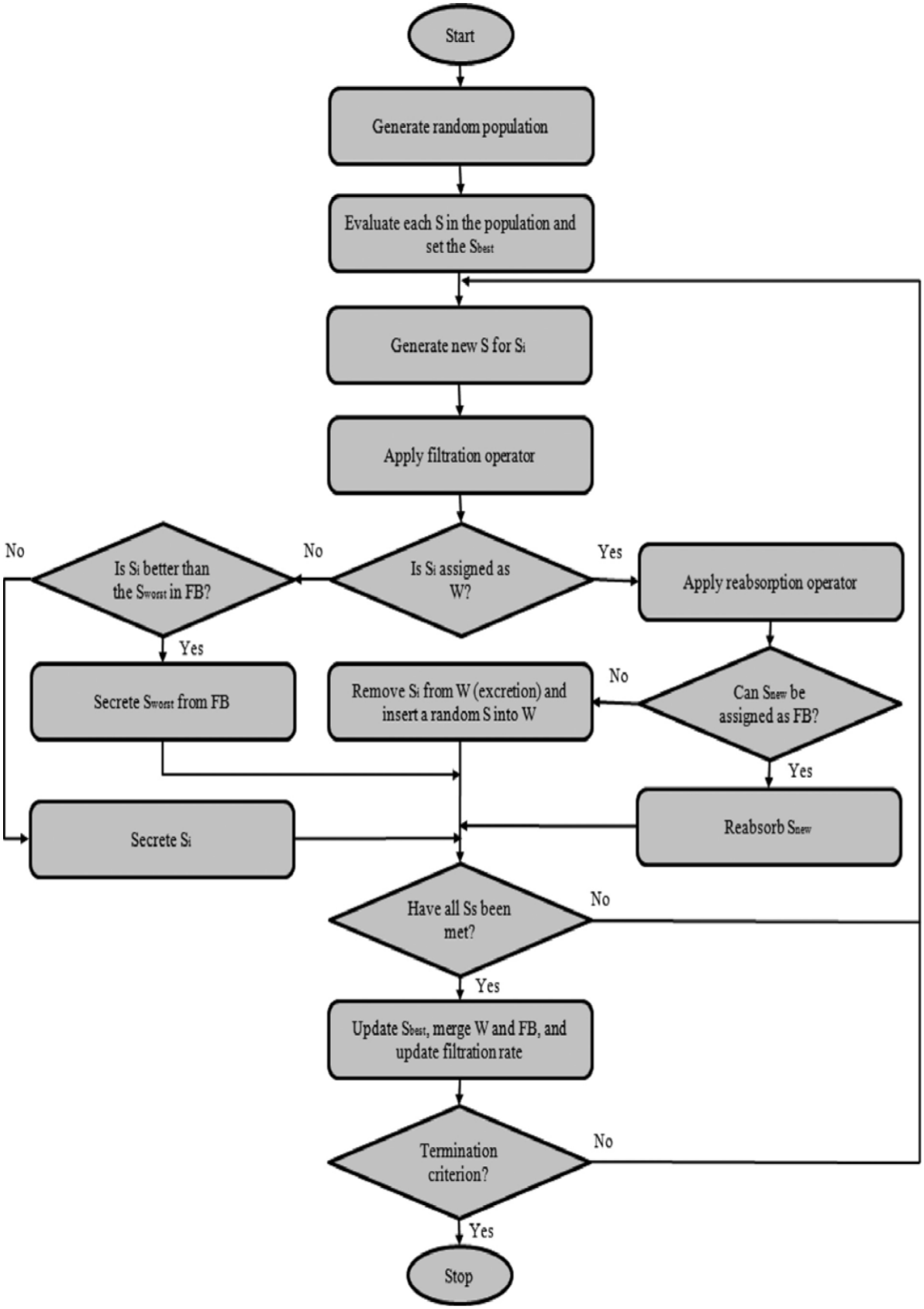
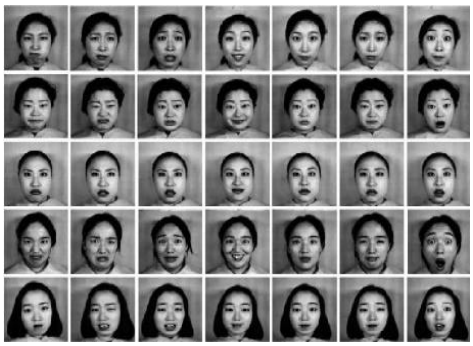


Fig. 4: Flowchart of KA [28].

(a)



(b)

Fig. 5: Datasets (a) FER2013[29] (b) JAFFE.

The expressions used to evaluate various performance parameters are given in (1) to (4) [30].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Table 3: The specific details of the experiment

Experiment specifications	Details
Metaheuristic algorithm	kidney
Population size	20
Iterations	250
Valid Padding Convolution	stride=1
Classifier	KNN(K=3) & MLP
Filter size	3x3 , 5x5 , 7x7, 9x9

This experiment, which was conducted based on Python software, was executed on the Google Colab platform, with the following specifications, which were used in the experiment. [31]

C. The Extraction of a Suboptimal Single Filter Using the KA Algorithm

Table 2, Table 3, Table 4, and Table 5 present the Confusion Matrix for the extraction of single filters of sizes 3x3, 5x5, 7x7, and 9x9 using the KA algorithm on the nearest neighbor classifiers. In Fig. 6, Fig. 7, Fig. 8, and Fig. 10, the optimization progress and extracted filters are shown.

Table 1: confusion matrix for extracting the suboptimal 3x3 single filter using the KA algorithm on KNN classifier (%)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	53.3	10.1	9.5	2.8	9.63	3	9.68
	disgust	9.4	72	4.6	2	2	3.5	2.5
	fear	15.15	14	60	1.5	17.82	13.2	6.2
	happy	2.5	2.6	2.5	85	2	6	4
	sad	10.7	1	7.4	2.2	55.5	1.1	16.16
	surprise	2	0	13	3	1.5	74.7	4.94
	neutral	6.95	0	3	4.5	11.55	1.5	56.52

Table 2: confusion matrix for extracting the suboptimal 5x5 single filter using the KA algorithm on KNN classifier (%)

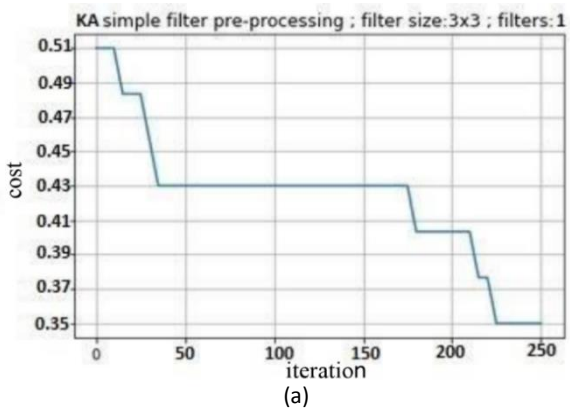
Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	57.66	9.6	8.43	2.2	8.63	1.6	8.88
	disgust	8.41	75	3.5	1.3	1.6	1.32	1.6
	fear	15.15	13.14	62.45	1.4	17.6	12.55	6.3
	happy	2.14	2	2.2	86.5	2.3	3.9	4.12
	sad	10.8	0	7.2	2.5	58.8	1.1	16.2
	surprise	1	0	12.12	1.5	1.5	77.9	1.99
	neutral	4.84	0.26	4.1	4.6	9.57	1.63	60.91

Table 4: confusion matrix for extracting the suboptimal 7x7 single filter using the KA algorithm on KNN classifier (%)

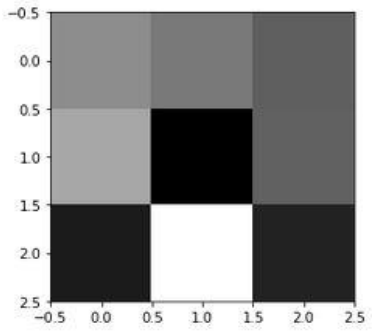
Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	52.3	11.4	9.5	1.8	8.38	2	8.18
	disgust	10.4	70	4.6	3.5	2	1.5	4
	fear	15.15	13	58	2	17.82	13.2	6.7
	happy	3.5	2.6	3.5	83	3	3.5	5
	sad	9.7	1	6.9	2.2	54	1.1	14.66
	surprise	3	1	12.5	4	2.5	77.2	6.44
	neutral	5.95	1	5	4.5	12.8	1.5	55.02

Table 5: confusion matrix for extracting the suboptimal 9x9 single filter using the KA algorithm on KNN classifier (%)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	49	12.4	10	2.3	9.38	2	6.68
	disgust	11.7	68	5.6	4.5	5	2.1	5.5
	fear	16.15	14	55.5	3.5	16.22	12.6	7.7
	happy	4	2.6	4.5	80.5	1.5	4.5	6
	sad	10.2	2	7.8	1.7	51	2.1	12.66
	surprise	3.3	0	11.6	5	2.5	75.2	8.94
	neutral	5.65	1	5	2.5	14.4	1.5	52.52

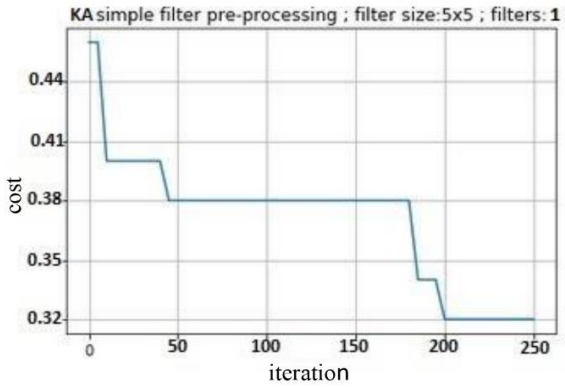


(a)

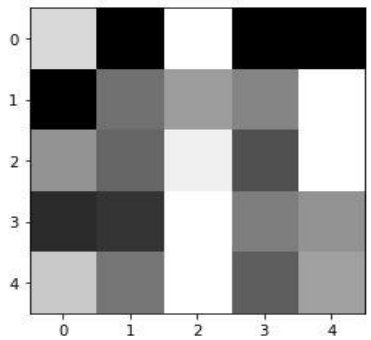


(b)

Fig. 6: (a) optimization progress, (b) adaptive filter 3x3.

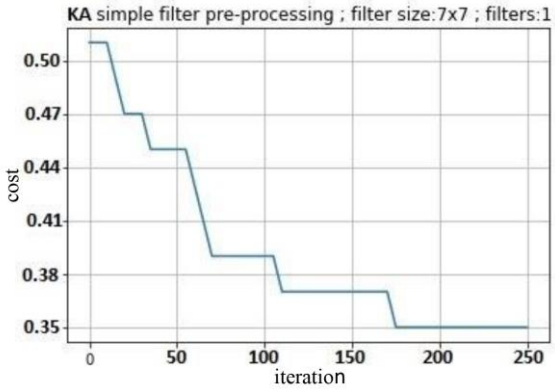


(a)

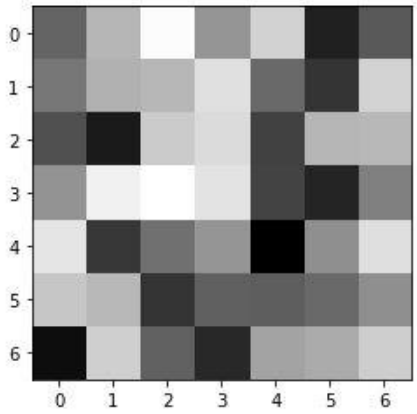


(b)

Fig. 7: (a) optimization progress, (b) adaptive filter 5x5.

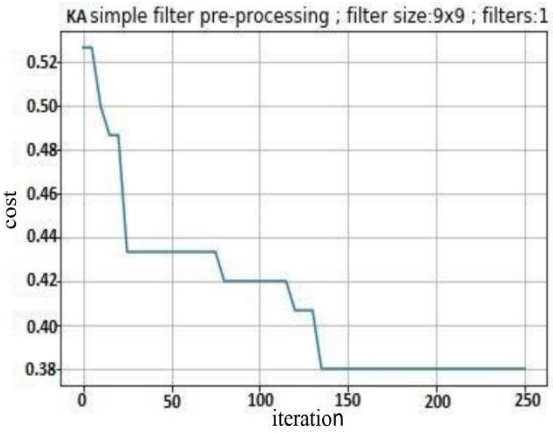


(a)

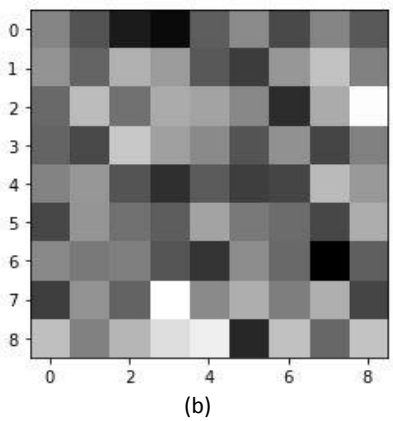


(b)

Fig. 8: (a) optimization progress, (b) adaptive filter 7x7.

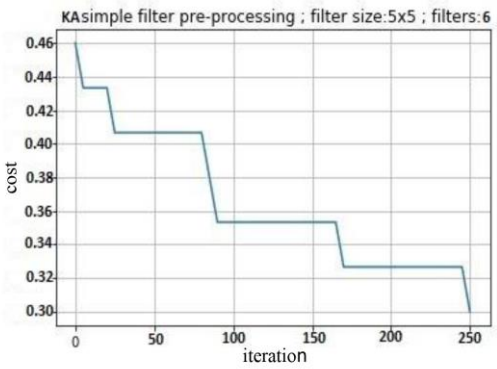


(a)

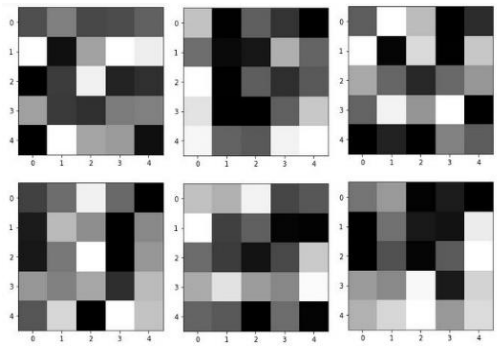


(b)

Fig. 9: (a) optimization progress, (b) adaptive filter 9x9.



(a)



(b)

Fig. 10: (a) optimization progress, (b) filter bank 5x5(knn).

Table 6 specifies the recognition rate and number of extracted features for each filter. The Confusion Matrix shows which individual filters have performed more successfully in specific classes. In each category, the numbers on the main diagonal indicate the accuracy rate of each filter.

Table 6: Comparing the accuracy results of facial expression detection and the number of features extracted by the overall algorithm for filters of different sizes

Filter Size	Accuracy	number of features
3×3	65.3%	900
5×5	68%	784
7×7	65%	676
9×9	62%	576

As can be seen in Table 6, the smaller the filter size, the greater the number of features extracted, emphasizing finer details. This filter, therefore, has a better local search ability. Conversely, as the filter size increases, the number of features decreases, resulting in more emphasis on general characteristics. As a result, a larger filter size is preferred for global searching. Therefore, these results will be used to develop suboptimal filterbanks for other experiments in order to achieve a more favorable outcome. Table 6 details the accuracy results of various facial expression recognition models and the number of

features extracted by the ka algorithm for optimized feature extraction.

The suboptimal adaptive filter with a size of 5×5 delivers the highest accuracy, achieving 68% and yielding a total of 784 features. Consequently, the next phase involves applying filter banks of 5×5 pixels to investigate their impact on facial expression recognition performance.

D. The Extraction of a Suboptimal 5x5 Filter Bank Using the KA Algorithm Via KNN

In the k-nearest neighbor classifier and KA algorithm, the performance accuracy was found to be 70%. Table 7 presents the confusion matrix of the optimized 5x5 filter bank (6 filters) on the knn classifier. The filter bank and optimization progress are depicted in Fig. 10. The number of features in this experiment is 4704.

Table 7: confusion matrix for extracting the suboptimal 5x5 filter bank using the KA algorithm on KNN classifier (%)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	59.66	8.6	8.43	1.5	9.23	1.1	7.39
	disgust	7.91	78	3.7	1	1	0.82	3.09
	fear	14.65	11.64	64.95	0.9	15.6	12.55	5.3
	happy	1.64	1.27	1.7	88	2	2.9	4.12
	sad	10.3	0	5.2	3	61.1	0.9	15.2
	surprise	2	0	13.12	1.3	3.25	80.1	1.24
	neutral	3.84	0.49	2.9	4.3	7.82	1.63	63.66

E. The Extraction of a Suboptimal 5x5 Filter Bank Using the KA Algorithm Via MLP

In the Multi-Layer Perceptron (MLP) neural network classifier and KA algorithm, the performance accuracy was found to be 78%. Table 7 presents the confusion matrix of the optimized 5x5 filter bank (6 filters) on the mlp classifier. The filter bank and optimization progress are depicted in Fig. 11. The number of features in this experiment is 4704. The Performance Parameters of the final model are presented in Table 7.

The accuracy of the proposed classification method and previous methods has been performed on the FER2013 database in Table 8. The best-reported accuracy is associated with the Ensemble ResMaskingNet method, combined with 6 other CNNs, achieving an accuracy of 82.76% [41]. In this article, a method is proposed for recognizing facial expressions using metaheuristic algorithms for optimizing feature extraction filters.

The method uses a continuous confrontation between optimized filter weights and nearest neighbor classifiers, as well as multi-layer perceptron neural networks to extract accurate features from the dataset and ensure accurate recognition of facial expressions. A stronger classifier is substituted in the proposed model by enhancing the weight structure of feature extraction filters. The computational cost for identifying the network model decreases. Ultimately, based on comprehensive datasets of various sizes, the proposed method is validated for its efficient performance in facial state recognition. It has been demonstrated that the proposed method offers distinct advantages in terms of precision of recognition, processing speed, and optimized filters. In this paper, an accuracy of 78% was achieved.

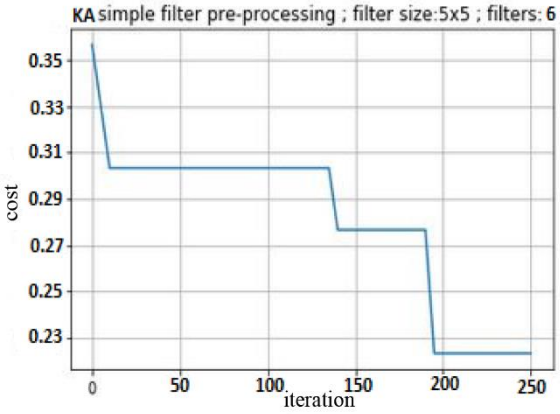
Table 9 compares the accuracy rates of different methods with the proposed approach, indicating that the proposed method has exhibited superior performance.

Table 8: confusion matrix for extracting the suboptimal 5x5 filter bank using the KA algorithm on MLP classifier (%)

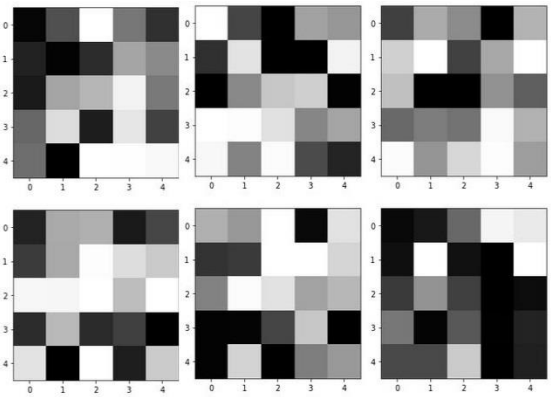
Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	67.76	7.23	6.23	0.5	5.8	0.1	4.23
	disgust	5.91	68.67	2.7	0.6	0	0.54	1.83
	fear	11.23	11.6	73.17	0.9	4.64	9.55	3.3
	happy	1.43	2	0.9	94	0.87	1.9	2.33
	sad	9.3	0.6	3.2	2	86.6	0.4	13.2
	surprise	3	2.25	9.12	0.3	0.5	85.67	1.24
	neutral	1.37	7.65	4.68	1.7	1.59	1.84	73.87

Table 9: Summary of results on FER2013

Method	Accuracy	Percision	Recall	F1-score
[16]CNN [16]	62.44	61.11	55.10	0.579
AlexNet	63.41	60.23	54.10	0.570
GoogleNet [20]	65.20	59.54	55.11	0.572
Human Accuracy [33]	65.5	62.33	60.22	0.612
VGG+SVM [19]	66.31	64.62	59.78	0.621
Conv+Inception [21]	66.40	65.14	60.32	0.626
Bag of Words [18]	67.40	66.50	62.41	0.643
Local Learning Bow [33]	67.48	66.74	60.78	0.636
Deep Emotion [34]	70.02	67.33	60.41	0.636
Attentional ConvNet [17]	70.02	68.38	65.32	0.668
EfficientNet	70.42	68.92	64.10	0.664
CNN+SVM [22]	71.20	69.08	65.52	0.672
ARM(ResNet-18) [23]	71.38	69.79	62.44	0.658
Inception [24]	71.60	68.19	66.14	0.671
Inception-v1	71.85	69.34	65.08	0.671
Ad-Corre [35]	72.3	69.68	67.14	0.683
ResNet [24]	72.40	70.30	66.50	0.683
SE-Net50 [36]	72.50	70.68	66.23	0.683
VGG [24]	72.70	70.92	67.38	0.691
Inception-v3	72.91	71.08	67.22	0.690
DenseNet-121	73.16	70.82	67.91	0.693
ResNet50 [36]	73.20	71.21	68.49	0.698
ResNet152	73.27	70.60	66.20	0.683
CNNs and BOVW + global SVM [37]	73.25	71.79	70.12	0.709
CBAM ResNet50	73.32	71.49	65.41	0.683
ResNet34v2	73.65	70.07	66.82	0.693
ResNet18 With Tricks [38]	73.7	71.33	67.27	0.692
Residual Masking Network [33]	74.14	71.60	73.20	0.723
LHC-NetC [39]	74.28	73.80	70.30	0.720
LHC-Net [39]	74.42	72.89	66.31	0.694
CNNs and BOVW + local SVM [37]	75.42	72.49	67.11	0.696
Segmentation VGG-19 [40]	75.97	73.38	68.52	0.708
Ensemble ResMaskingNet with 6 other CNNs [41]	76.82	73.17	68.23	0.706
Our Method	**78**	**77.62**	**72.08**	**0.747**



(a)

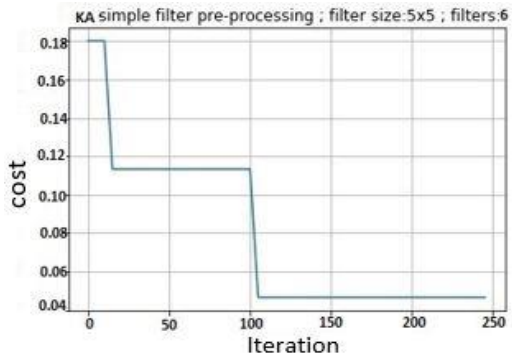


(b)

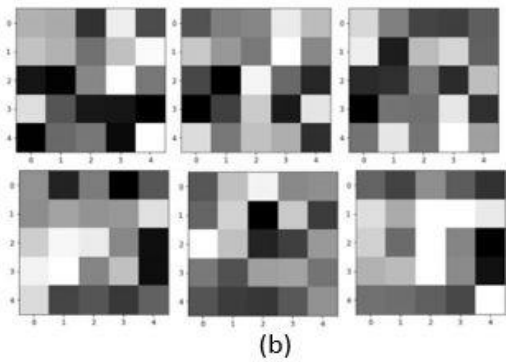
Fig. 11: (a) optimization progress, (b) filter bank 5x5(mlp) (fer2013).

F. The Final Proposed Model Was Evaluated on the JAFFE Dataset

It was shown that the proposed approach has outperformed other competitors on the JAFFE [42] database, which corresponds to a database of Japanese women, achieving an accuracy of 94.88 percent. presents the confusion matrix of the optimized 5x5 filter bank (6 filters) on the mlp classifier. The filter bank and optimization progress are depicted in Fig. 12.



(a)



(b)

Fig. 12: (a) optimization progress, (b) filter bank 5x5(mlp) (jaffe).

The accuracy of the proposed classification method has been compared with previous methods on the JAFFE

accuracy of 94.88% was achieved. In Table 10, we compare the accuracy rates of different methods using the proposed approach, demonstrating that the proposed method has demonstrated superior performance on the dataset JAFFE when compared with other methods.

Table 10: Summary of results on JAFFE

Reference	Model	Accuracy
Sajjanhar et al. [42]	Pre-trained InceptionV3	75.88
	Pre-trained VGG19	94.71
	Pre-trained VGG-Face	86.67
Bhatti et al. [43]	RELM	91.67
Minaee et al. [44]	Attentional CNN	92.8
Kartheek et al. [45]	SVM	66.2
Sun et al. [46]	PCANet	71.38
	LDANet	70.18
Goutam Kumar Sahoo et al. [47]	Pre-trained AlexNet	65.62
	Pre-trained SqueezeNet	57.8
	Pre-trained VGG19	84.4
Ghasemi et al.	***Our Method***	***94.88***

G. Comparative Analysis of Training Time of the Proposed Method

Based on the performance accuracy of the proposed method on two databases, the proposed method is shown to be superior to other methods. The training time of the proposed model is analyzed using AlexNet, SqueezeNet, and VGG19 models on the JAFFE dataset. Table 11 shows that the proposed model requires less training time than the existing model. The model is trained with a training and validation data split of 70–30%.

Table 11: The performance of trained networks on the Jaffe dataset in terms of training time

Model	Training time (Sec)
AlexNet	342
SqueezeNe	284
VGG1	920
Our Method	276

H. Analyses of Facial Expressions with Confusion Matrix Results

An evaluation of a classifier can be conducted by using the confusion matrix. Essentially, the confusion matrix represents the results of classification based on the actual information available and is extremely useful in evaluating the performance of a classifier for one or more classes in particular. This matrix indicates the recognition rate of correctly identifying samples by its main diagonal.

Table 8 shows the confusion matrix resulting from the classification on the 2013-FER dataset. In facial expression recognition, the detection of the state of happiness is generally easier compared to other states due to the presence of more pronounced facial features. Comparatively to other classes in the dataset, the recognition of the state of happiness performed better than other classes. It is important to note that the state of sadness exhibits many similarities with other emotional states such as anger and fear. The manifestations of emotions such as anger, fear, and sadness are somewhat similar, so it is possible for diagnoses to be made in error. The proposed framework, however, has shown higher performance in the states of sadness and fear when compared to other states in the 2013-FER dataset despite lower accuracy rates.

Table 12 presents the evaluation results of the proposed structure on the JAFFE [48] dataset. The similarity between these two emotions in this dataset makes it difficult to distinguish between the sad and fearful expressions. In addition, there have been some errors in the detection of the surprise expression due to the similarity between the surprise and fear expressions. Overall, the proposed framework performed well on the dataset.

Table 12: confusion matrix for extracting the suboptimal 5x5 filter bank using the KA algorithm on MLP classifier (%)(JAFFE)

Confusion Matrix		Actual Class						
		angry	disgust	fear	happy	sad	surprise	neutral
Predicted Class	angry	96	2	2	0	0	0	1
	disgust	3	97	1	0	1	0	0
	fear	0	0	95	1	8.26	7.23	0
	happy	0	0	0	96	0	0	1
	sad	0	0	0	0	90.74	0	2
	surprise	1	1	1	3	0	92.77	0
	neutral	0	0	0	0	0	0	96

Conclusion

This article presents a method to improve facial expression recognition using an optimized filter and the KA metaheuristic algorithm. The method incorporates four filters with sizes 3×3, 5×5, 7×7, and 9×9. Each filter affects the number of features used. As the size of the filters increases, the number of features decreases. The selection of appropriate facial expression features plays an important role in increasing the accuracy of facial expression recognition. Different sizes of filters were extracted to improve facial expression recognition accuracy, and it was found that a 5×5 filter size yields the highest accuracy. In addition to improving accuracy by utilizing suboptimal filters for facial expression recognition, the suboptimal filter has also contributed to enhancing algorithm precision. This is because, after applying the filter, images are obtained that are not intelligible to us. By utilizing the nearest neighbor classifier and selecting the suboptimal k value, as well as the KA metaheuristic algorithm, we have been able to extract the best features for facial expression recognition. The suboptimal size of a single filter was therefore determined, and facial expression recognition was performed using filter banks, which achieved better accuracy than a single filter. In the final stage, instead of using the k-nearest neighbors (kNN) classifier, a perceptron neural network was employed, which is a powerful classifier, to optimize facial expression recognition.

Based on the obtained results, it appears that this approach has several advantages over other methods, including the reduced number of parameters, the decreased amount of training time, the ease of implementation, and the availability of filters with optimized parameters. These reasons substantiate the superiority of this method over others. Furthermore, the proposed method has shown higher accuracy in recognizing facial expressions than other methods. Our algorithm has demonstrated superior performance in comparison to the other methods.

Author Contributions

Dr. Razavi was the supervisor and Dr.Mohammadzade and Taghipour adviser of the current research paper. They sketched the research framework and the roadmap. Also, they analyzed the results. E. Ghasemi searched in authentic journals to gather all relevant papers. Also, he collected the data and wrote the manuscript. Dr. Razavi, Dr. Mohammadzade, Dr. Taghipour and E. Ghasemi interpreted the results.

Acknowledgment

We would like to express our gratitude to the Research Institute of Communication and Information Technology and the Development Headquarters of Connectivity and Communication Technologies of the Presidential Vice President for Science, Technology and Knowledge-Based Economy, who supported Mr. Ehsan Ghasemi's thesis, and also to the research supervisor Dr. Shaghayegh Naderi.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

CNN	Convolutional Neural Network
ASD	Autism Spectrum Disorders
RLUE	rectified linear unit
ICML	International Conference on Machine Learning
ARM	Advanced RISC Machines
MTCNN	Multi-task Cascade Convolutional Network
FER	Face Expression Recognition
KNN	K-Nearest Neighbor
MLP	Multi Layer Perceptron
KA	Kidney Algorithm
FB	Filtering Better solution
W	Worse solution
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

References

[1] S. Minaee, M. Minaei, A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," Sensors, 21(9): 3046, 2021.

[2] S. Kaur, N. Kulkarni, "Emotion recognition-a review," Int. J. Appl. Eng. Res., 16(2): 103-110, 2021.

[3] E. Palagi, A. Celeghin, M. Tamietto, P. Winkielman, I. Norscia, "The neuroethology of spontaneous mimicry and emotional contagion in human and non-human animals," Neurosci. Biobehav. Rev., 111: 149-165, 2020.

[4] P. Singh, R. K. Mishra, S. Urolagin, V. Sharma, "Enhancing security by identifying facial check-in using deep convolutional neural network," in Proc. 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N): 1006-1010, 2021.

[5] J. He, "Algorithm composition and emotion recognition based on machine learning," Comput. Intell. Neurosci., 2022 .

[6] K. Richardson et al., "Robot enhanced therapy for children with autism (DREAM): A social model of autism," IEEE Technol. Soc. Mag., 37(1): 30-39, 2018.

[7] K. Osaka, R. Tanioka, F. Betriana, T. Tanioka, Y. Kai, R. C. Locsin, "Robot therapy program for patients with dementia: Its framework and effectiveness," Information Systems - Intelligent Information Processing, p. 87, 2021.

[8] D. Cruz-Sandoval, A. Morales-Tellez, E. B. Sandoval, J. Favela, "A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms," in Proc. ACM/IEEE International Conference on Human-Robot Interaction: 161-169, 2020.

[9] B. Szymona et al., "Robot-assisted autism therapy (RAAT). Criteria and types of experiments using anthropomorphic and zoomorphic robots. Review of the research," Sensors, 21(11): 3720, 2021.

[10] Z. Wang et al., "Masked face recognition dataset and application," arXiv:2003.09093v2, 2020.

[11] S. MacHiraju, S. Urolagin, R. K. Mishra, V. Sharma, "Face mask detection using keras, opencv and tensorflow by implementing mobilenetv2," in Proc. 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N): 1485-1489, 2021.

[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, 86(11):2278-2324, 1998.

[13] S. Fooladi, H. Farsi, S. Mohamadzadeh., "Detection and classification of skin cancer using deep learning, " J. Birjand Univ. Med. Sci., 26(1): 44-53, 2019.

[14] S. Albawi, T. A. Mohammed, S. Al-Zawi, "Understanding of a convolutional neural network," in Proc. International Conference on Engineering and Technology (ICET), 2017.

[15] B. Han, J. Sim, H. Adam, "BranchOut: Regularization for online ensemble tracking with convolutional neural networks," in Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[16] K. Liu, M. Zhang, Z. Pan, "Facial expression recognition with CNN Ensemble," in Proc. 2016 International Conference on Cyberworlds, CW 2016.

[17] S. Minaee, A. Abdolrashidi, "Deep-emotion: facial expression recognition using attentional convolutional network," arXiv:1902.01019v1, 2019.

[18] R. T. Ionescu, M. Popescu, C. Grozea, "Local learning to improve bag of visual words model for facial expression recognition," in Proc. ICML 2013 Workshop on Representation Learning, 2013.

[19] M. I. Georgescu, R. T. Ionescu, M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," IEEE Access, 7: 64827- 64836, 2019.

[20] P. Giannopoulos, I. Perikos, I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," Smart Innovation, Systems and Technologies, 85: 1-16, 2018.

[21] A. Mollahosseini, D. Chan, M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in Proc. IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

[22] Y. Tang, "Deep learning using linear support vector machines," arXiv, 1306.0239, 2013.

[23] J. Shi, S. Zhu, Z. Liang, "Learning to amend facial expression representation via de-albino and affinity," arXiv Prepr. arXiv2103.10189, 2021.

[24] C. Pramerdorfer, M. Kampel, "Facial expression recognition using convolutional neural networks: state of the art," arXiv Prepr. arXiv1612.02903, 2016.

[25] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Process. Lett., 23(10): 1499-1503, 2016.

[26] T. Shanmugaraja, R. Dhivaya Devi, M. Supriya, K. Murugan, T. Venkatesh, K. Siddarthraju, "A critical review on applications of adaptive filters," J. Crit. Rev., 7(6): 656-659, 2020.

[27] S. R. Ashokkumar, G. MohanBabu, S. Anupallavi, "A novel two-band equilateral wavelet filter bank method for an automated detection of seizure from EEG signals," Int. J. Imaging Syst. Technol., 30(4): 978-993, 2020.

[28] N. S. Jaddi, J. Alvankarian, S. Abdullah, "Kidney-inspired algorithm for optimization problems," Commun. Nonlinear Sci. Numer. Simul., 42: 358-369, 2017.

[29] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Netw., 64: 59-63, 2015.

[30] M. Taheri, M. Rastgarpour, A. Koochari, "A novel method for medical image segmentation based on convolutional neural networks with SGD optimization," J. Electr. Comput. Eng. Innov., 9 (1): 37-46, 2020.

[31] S. Fooladi, H. Farsi, S. Mohamadzadeh, "Segmenting the lesion area of brain tumor using convolutional neural networks and fuzzy k-means clustering," Int. J. Eng., 36(8): 1556-1568, 2023.

[32] H. Bayati, M. B. Dowlatshahi, M. Paniri, "Multi-label feature selection based on competitive swarm optimization," J. Soft Comput. Inf. Technol., 9(4): 56-69, 2020.

[33] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Netw., 64: 59-63, 2015.

[34] S. Minaee, M. Minaei, A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," Sensors, 21(9): 3046, 2021.

[35] A. P. Fard, M. H. Mahoor, "Ad-corre: Adaptive correlation-based loss for facial expression recognition in the wild," IEEE Access, 10: 26756-26768, 2022.

[36] A. Khanzada, C. Bai, F. T. Celepcikay, "Facial expression recognition with deep learning," arXiv Prepr. arXiv2004.11823, 2020.

[37] M. I. Georgescu, R. T. Ionescu, M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," IEEE Access, 7: 64827-64836, 2019.

[38] Y. Khaireddin, Z. Chen, "Facial emotion recognition: State of the art performance on FER2013," arXiv Prepr. arXiv2105.03588, 2021.

[39] R. Pecoraro, V. Basile, V. Bono, "Local multi-head channel self-attention for facial expression recognition," Information, 13(9): 419, 2022.

[40] S. Vignesh, M. Savithadevi, M. Sridevi, R. Sridhar, "A novel facial emotion recognition model using segmentation VGG-19 architecture," Int. J. Inf. Technol., 15(4): 1777-1787, 2023.

[41] L. Pham, T. H. Vu, T. A. Tran, "Facial expression recognition using residual masking network," in Proc. 25Th International Conference on Pattern Recognition (ICPR): 4513-4519, 2020.

[42] A. Sajjanhar, Z. Wu, Q. Wen, "Deep learning models for facial expression recognition," in Proc. 2018 International Conference on Digital Image Computing: Techniques and Applications (DICTA): 1-6, 2018.

[43] Y. K. Bhatti, A. Jamil, N. Nida, M. H. Yousaf, S. Viriri, S. A. Velastin, "Facial expression recognition of instructor using deep features and extreme learning machine," Comput. Intell. Neurosci., 2021: 1-17, 2021.

[44] S. Minaee, M. Minaei, A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," Sensors, 21(9): 3046, 2021.

[45] M. N. Kartheek, M. V. N. K. Prasad, R. Bhukya, "Windmill graph based feature descriptors for facial expression recognition," Optik (Stuttg)., 260: 169053, 2022.

[46] Z. Sun, H. Zhang, S. Ma, Z. Hu, "Combining filtered dictionary representation based deep subspace filter learning with a discriminative classification criterion for facial expression recognition," Artif. Intell. Rev., 55(8): 6547-6566, 2022.

[47] G. K. Sahoo, S. K. Das, P. Singh, "Performance comparison of facial emotion recognition: A transfer learning-based driver assistance

framework for in-vehicle applications," Circuits, Syst. Signal Process.: 1–28, 2023.

[48] M. J. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images," IEEE Trans. Pattern Anal. Mach. Intell., 21(12): 1357-1362, 1999.

Biographies

Ehsan Ghasemi received his AD.Sc. degree in Electrical Engineering from university of Islamic Republic of Iran broadcasting, in 2014 and his B.Sc. degree in Electrical Engineering from Islamic Azad University, Birjand, in 2018 and M.Sc. degree in Electrical Engineering from the Birjand University, in 2021 respectively. He is currently a Ph.D. student at Birjand University to receive a Ph.D. degree in Electronics Engineering. His research interests include Computer Vision, Pattern Recognition, optimization algorithms and Artificial Intelligence.

- Email: ehsanghasemi91@birjand.ac.ir
- ORCID: 0009-0002-6185-9049
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Seyyed Mohammad Razavi received his B.Sc. degree in Electrical Engineering from Amirkabir University of Technology, in 1994 and his M.Sc. and Ph.D. degree in Electrical Engineering from the Tarbiat Modares University, Iran, in 1996 and 2006 respectively. Now, he is a Full Professor in University of Birjand. His research interests include Computer Vision, Pattern Recognition and Artificial Intelligence.

- Email: smrazavi@birjand.ac.ir
- ORCID: 0000-0002-3493-7614
- Web of Science Researcher ID: AAF-7386-2021
- Scopus Author ID: 56214431100
- Homepage: https://cv.birjand.ac.ir/mrazavi/fa

Sajad Mohamadzadeh received the B.Sc. degree in Communication Engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. and Ph.D. degree in Communication Engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012 and 2016, respectively. Now, he works as Associte professor at department of electrical and computer engineering, University of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Deep Neural Network, Pattern recognition, Digital Signal Processing, Sparse Representation, and Deep Learning.

- Email: s.mohamadzadeh@birjand.ac.ir
- ORCID: 0000-0002-9096-8626
- Web of Science Researcher ID: NA
- Scopus Author ID: 57056477500
- Homepage: https://cv.birjand.ac.ir/mohamadzadeh/en

Mehran Taghipour-Gorjikolaie received the B.Sc. degree in Electrical Engineering from the University of Mazandaran, Iran, in 2008, and the M.Sc. (Hons.) and Ph.D. degrees in Electronic Engineering from the University of Birjand, Iran, in 2011 and 2016, respectively. He was a lecturer (assistant Professor) with the University of Birjand from 2016 to 2021, after that he was a research fellow in application AI, ML, and CV in healthcare and wellbeing at CVSSP, in the University of Surrey from 2021 to 2023, and now he is working as research fellow in London South Bank University. His research interests include the application of machine learning, artificial intelligence, computer vision and optimization algorithms.

- Email: mehran.taghipour-gorjikolaie@lsbu.ac.uk
- ORCID: 0000-0001-6132-8454
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
E. Ghasemi, S. M. Razavi, S. Mohamadzadeh, M. Taghipour-Gorjikolaie, "Facial expression recognition through suboptimal filter design using a metaheuristic kidney algorithm," J. Electr. Comput. Eng. Innovations, 12(2): 425-438, 2024.

DOI: 10.22061/jecei.2024.10512.717

URL: https://jecei.sru.ac.ir/article_2096.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Modeling and Simulation of DC Electric Railway System with Regenerative Braking: A Case Study of Isfahan Metro Line 1

P. Hamedani [1,*], ***S. S. Fazel*** [2], ***M. Shahbazi*** [3]

[1] *Department of Railway Engineering and Transportation Planning, University of Isfahan, Isfahan, Iran.*

[2] *School of Railway Engineering, Iran University of Science and Technology, Tehran, Iran.*

[3] *Department of Engineering, Durham University, Durham, U.K.*

Article Info

Article History:

Received 13 February 2024
Reviewed 15 March 2024
Revised 05 April 2024
Accepted 29 April 2024

Keywords:

DC electric railway
Regenerative braking
Power flow
Traction power systems
Isfahan metro
Train movement simulation

*Corresponding Author's Email Address:
p.hamedani@eng.ui.ac.ir

Abstract

Background and Objectives: Modeling and simulation of electric railway networks is an important issue due to their non-linear and variant nature. This problem becomes more serious with the enormous growth in public transportation tracks and the number of moving trains. Therefore, the main aim of this paper is to present a simple and applicable simulation method for DC electric railway systems.
Methods: A train movement simulator in a DC electric railway line is developed using Matlab software. A case study based on the practical parameters of Isfahan Metro Line 1 is performed. The simulator includes the train mechanical movement model and power supply system model. Regenerative braking and driving control modes with coasting control are applied in the simulation.
Results: The simulation results of the power network are presented for a single train traveling in both up and down directions. Results manifest the correctness and simplicity of the suggested method which facilitates the investigation of the DC electric railway networks.
Conclusion: According to the results, the train current is consistent with the electric power demand of the train. But the pantograph voltage has an opposite relationship with its electric power demand. In braking times, the excess power of the train is injected into the electrical network, and thus, overvoltage and undervoltage occur in the overhead contact line and the substation busbar. Therefore, at the maximum braking power of the train, the pantograph voltage reaches its maximum. The highest amount of fluctuation is related to the substation that is closest to the train. As the train moves away from the traction substations, the voltage fluctuations decrease and vice versa.

Introduction

In recent years, with the increase in population, the need to expand the public transportation network has increased. One of the most popular methods of urban transportation is using the urban metro system. Urban metro systems are usually fed through DC power supply systems. Two common electrification methods for metro trains are the third rail system and overhead contact system (OCS). The third rail system is usually used at the standard voltage level of 750 V. To increase the speed and capacity of the transportation network, the overhead contact network with 1500 VDC and 3000 VDC is utilized [1]-[4].

To design, and analyze the performance of traction stations and electric train electrification systems, as well as the technical-economic justification of electric rail projects, electrical modeling and simulation of the rail

power supply network is necessary [1]-[4]. In addition, the electrical simulation of the railway power supply network requires the extraction of movement graphs and speed profiles of trains. Determining the movement graphs and speed profile of trains can be very useful in calculating the train performance, determining the train travel time, the planning of train movement, increasing the line capacity, and also determining the amount of energy consumption [5]-[7]. So far, several studies have been conducted in the field of train movement simulation and movement profile extraction. Modeling the movement of a train on a track depends on several factors, and therefore the extraction of movement graphs requires complex analytical methods [5]. Train and track parameters such as train weight, aerodynamic resistance force, quality of brakes, gradient profile and curve profile of the movement path, friction coefficient, as well as speed and acceleration limitations, are among the factors that are effective in modeling the train movement. In [5], the main equations required for train movement calculations are presented. In [8]-[13], the speed profile of the trains has been optimized in the electric railway track to reduce energy consumption. The results were repeated on different railway lines and showed a reduction in travel time and energy consumption.

The electric railway network has a time-varying nonlinear model. But in general, the model of the DC railway network is simpler than the AC railway network. The DC electric railway network modeling has been studied in the literature [14]-[23]. Gauss-Seidel and Newton-Raphson are common methods in solving DC railway power flow [17]-[19]. In addition, the current injection method, point-Jacobi method, Zollenkopf's bifactorisation, and incomplete Cholesky conjugate gradient method, have been proposed and investigated [20]-[23]. Recently, the application of electric trains with regenerative braking has been significantly expanded in the urban transport network and subways. Various studies have been reported on the evaluation and management of electric energy consumption in electric railway networks, including regenerative trains [9], [19], [24]-[30]. However, consideration of the regenerative braking capability complicates the railway network modeling and the power flow solving.

This paper proposes a simplified simulation model for the DC electric railway network including the train mechanical movement model and power supply system model. Regenerative braking and driving control modes with coasting control are applied in the simulation. Based on the route data of Isfahan Metro Line 1, the simulation results of the power network are presented for a train traveling in both up and down directions. Results manifest the correctness and simplicity of the suggested method which facilitates the investigation of the DC electric railway networks.

Modeling of DC Electric Railway

A. Train Kinematics Modeling

For kinematics modeling of the trains, Newton's second law of motion can be utilized. The main equation of movement can be expressed as [31]:

$$F_T = TE - F_{grad} - F_{drag} = M_{eff}\alpha \quad (1)$$

where F_T is the total force [N]. TE, F_{grad}, and F_{drag} represent the traction effort [N], the gradient force [N], and the drag resistance force [N], respectively. The curvature resistance force is neglected in this paper. α is the acceleration of the train [m/s^2]. Moreover, M_{eff} is the effective mass of the train [Kg] and can be written as:

$$M_{eff} = (1+\lambda)\, M_t + M_l \quad (2)$$

M_t and M_l are the tare mass and payload mass of the train [Kg], respectively. Furthermore, λ is the rotary allowance. The drag resistance force can be calculated according to the Davis equation:

$$F_{drag} = a + b\,V + c\,V^2 \quad (3)$$

where F_{drag} represents the drag resistance force [N]. a [N], b [Nh/Km], and c [Nh^2/Km^2] are the Davis coefficients. V is the train speed [Km/h]. The gradient force can be written as:

$$F_{grad} = M\, g\, sin(\theta) \quad (4)$$

in which g (9.81 m/s^2) and θ [rad] are the gravity acceleration and the slope angle, respectively. $M=M_t+M_l$ [Kg] is the train mass.

As shown in Fig. 1, to model the train movement, four different operating modes can be considered including traction, cruising, coasting, and braking modes [31].

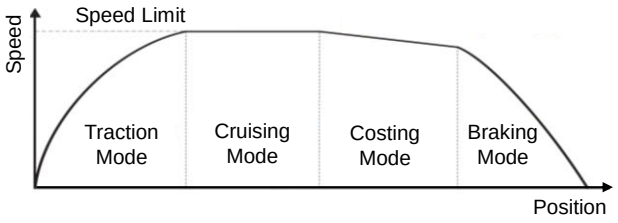


Fig. 1: Train operating modes.

The traction mode is the region in which the train accelerates. The cruising mode is the region in which the train runs at maximum speed according to the speed limits. The costing mode is the region in which the train runs with zero tractive effort in a downhill. The braking mode is the region in which the train decelerates.

In each operating mode, by solving the train movement specified in (1), the train position, speed, and acceleration can be extracted, as shown in Table 1. The tractive effort (TE) is the generated force by the traction motors for overcoming the resistance forces imposed to

the running train on the route. In the braking mode, the braking force (BR) is the generated by the braking system for stopping the train. Fig. 2 presents the typical tractive effort [KN] and braking force [KN] curves versus the speed [Km/h].

Table 1: Train Movement Equation in Different Modes

Operating Modes	Equations
Traction	$TE > Mg\,sin(\theta) + F_{drag}$ $\alpha = \frac{TE - Mg\,sin(\theta) - F_{drag}}{M_{eff}}$
Cruising	$TE = Mg\,sin(\theta) + F_{drag}$ $\alpha = 0$
Costing	$TE = 0$ $\alpha = \frac{-Mg\,sin(\theta) - F_{drag}}{M_{eff}}$
Braking	$BR < 0$ $\alpha = \frac{BR - Mg\,sin(\theta) - F_{drag}}{M_{eff}}$

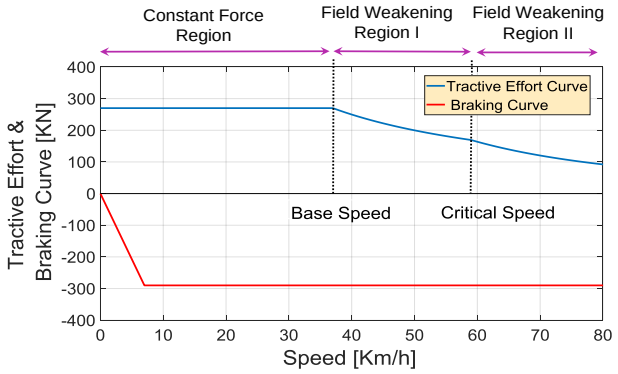


Fig. 2: Tractive effort and braking curve.

The tractive effort curve consists of three different regions including constant force region, field weakening region I, and field weakening region II. The constant force region is the area below the base speed. The flux-weakening region I starts from the base speed (due to the power limitation) and ends at the critical speed. The flux-weakening region II is the area above the critical speed (due to the motor limitation).

B. Train Power Modeling

The next step for simulating the train movement is to calculate the required mechanical power of the train (P_{me}):

$$P_{me} = TE \times V \tag{5}$$

The required electric power of the train (P_{el}) can be extracted with respect to the motoring and braking condition of the train [31], [32]. In the motoring operating condition, the mechanical power is positive while in the braking operating condition, the mechanical power is negative.

$$P_{el} = \begin{cases} \frac{P_{me}}{\eta} & if\, P_{me} \geq 0 \\ P_{me} \times \eta & if\, P_{me} < 0 \end{cases} \tag{6}$$

where η is the efficiency of the traction system from the pantograph to the wheel. P_{me} [W] and P_{el} [W] are the mechanical and electrical power of the train, respectively. The total power demand of the train (P_{train}) is the summation of the electrical power and the auxiliary power [31], [32]:

$$P_{train} = P_{el} + P_{aux} \tag{7}$$

P_{aux} [W] represents the auxiliary power that includes lighting, air conditioning, and other energy consumptions of the passengers.

C. DC Railway Power Network Modeling

The power network includes traction substations, overhead contact lines (OCLs), rails as return current circuits (RCRs), and trains. Fig. 3 shows the structure of the DC railway power network between two traction substations. Fig. 4 illustrates the equivalent circuit of the DC railway power network between two traction substations.

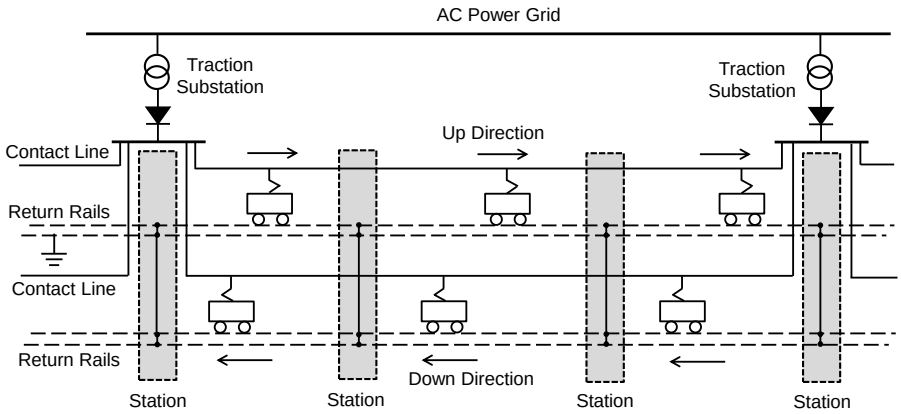


Fig. 3: Structure of the DC railway power network.

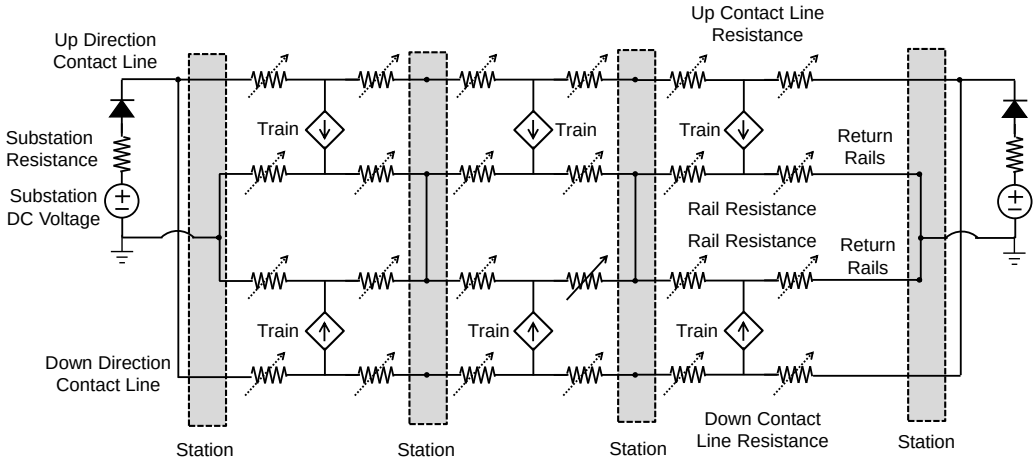


Fig. 4: Equivalent circuit of the DC railway power network between two traction substations.

In this work, the voltage regulation characteristic of the substation is assumed linear. Therefore, Thevenin's model of the traction substation can be utilized [19]. In Isfahan Metro, twelve-pulse rectifiers are used in the substations. The no-load voltage can be considered 1588 V [32]. The rated voltage and current can be considered 1500 V and 2100 A, respectively. As a result, the Thevenin's voltage is equal to 1500 V and the equivalent resistance of the traction substations can be calculated as:

$$R_{sub} = \frac{\Delta V}{\Delta I} = \frac{V_{noload} - V_{rated}}{I_{rated} - 0} = 0.0419\ \Omega \tag{8}$$

As shown in Fig. 4, in a DC railway power network, the OCL and RCR are modeled with resistance. The total resistance of the overhead contact line (R_c) and return current circuit between two passenger stations (R_r) can be written as:

$$R_c = L \times r_c \tag{9}$$

$$R_r = L \times \frac{r_r}{2} \tag{10}$$

where r_c [Ω/Km] and r_r [Ω/Km] are the resistance of the overhead contact line and rails per track. L is the length of the track between two consecutive passenger stations [Km]. The OCL and RCR resistance between the train and the previous passenger station depends on the length [18]:

$$R_c{}' = D \times r_c \tag{11}$$

$$R_r{}' = D \times \frac{r_r}{2} \tag{12}$$

where D is the traveled distance of the train from the previous stations [Km]. R_c' [Ω] and R_r' [Ω] are the OCL and RCR resistance between the train and the previous passenger station, respectively.

The OCL and RCR resistance between the train and the next passenger station can be expressed as [18]:

$$R_c{}'' = (L - D) \times r_c \tag{13}$$

$$R_r{}'' = (L - D) \times \frac{r_r}{2} \tag{14}$$

R_c'' [Ω] and R_r'' [Ω] are the OCL and RCR resistance between the train and the next passenger station, respectively. This method can be used for calculating the OCL and RCR resistances in both up and down tracks.

The trains are modeled as voltage-dependent current sources:

$$I_{train} = \frac{P_{train}}{V_{train}} \tag{15}$$

in which V_{train} [V] and I_{train} [A] are the pantograph voltage and current, respectively. P_{train} [W] can be calculated using (7). The direction of each current source depends on the motoring or braking condition of the train.

Results and Discussion

Based on the modeling strategy of the DC electric railway described in previous sections, a train movement simulator has been developed in this work. In this section, a case study based on the Isfahan Metro Line 1 is presented.

In Table 2, the train and line parameters required for the mechanical simulation of the rail system are represented. Line 1 of the Isfahan Metro is about 20.464 Km long. The initial design of this line had 21 passenger stations and 7 traction substations. In this work, the initial design of Line 1 is studied. In Table 3, the location of the passenger stations and traction substations of Isfahan Metro Line 1 are listed. The up track corresponds to the running direction from the GHO station to the DEMO station. The down track corresponds to the running direction from the DEMO station to the GHO station. The traction substations consist of twelve-pulse rectifiers.

Trains are fed with the OCL system. The rated voltage of the DC OCL is 1500 V.

Table 2: Train and Track Parameters of Isfahan Metro Line 1

Parameters	Value
Train Configuration	Tc-Mp-M-Mp-Tc
Train Tare Mass	176 [tons]
Passenger Mass in AW2 (6 Person/m^2)	78.19 [tons]
Max. Speed	80 [Km/h]
Base Speed	37 [Km/h]
Critical Speed	59 [Km/h]
Max. Tractive Effort	270 [KN]
Max. Braking Effort	290 [KN]
Auxiliary Power	380 [Kwatt]
Efficiency	90 [%]
Max. Acceleration	1 [m/s^2]
Max. Deceleration	1.1 [m/s^2]
Rotary Allowance	1.088
Davis Constants	a= 4226.816 b= 35.5866 c= 0.51516
Dwell Time	30 [sec]
RCR Resistance per track per Km (a Rail S49)	22.5 [mΩ/Km]
OCL Resistance per Km	55.5 [mΩ/Km]

Fig. 5 shows the gradient profile of Isfahan Metro Line 1 for the up track. According to the mechanical and dynamic equations of train movement that were introduced in section 2 and also using the route parameters of Isfahan Metro Line 1, the train movement graphs and train timetables are extracted. Because the results obtained in the mechanical simulator section are used as necessary input for the electrical simulator program of the electric railway line.

Table 3: Location of Passenger Stations and Traction Substations in The Up Track of Isfahan Metro Line 1

	Station Code	Passenger Station Name	Position (m)	Traction Substation
1	GHO	Ghods	0	TSS1
2	BAHA	Baharestan	1380	-
3	GOL	Golestan	2780	-
4	MOF	Shahid Mofateh	4106	TSS2
5	ALI	Shahid Alikhani	5404	-
6	JAB	Jaber	6538	-
7	KAV	Kaveh	7833	TSS3
8	CHA	Shahid Chamran	9131	-
9	BAHO	Shahid Bahonar	10037	-
10	SHO	Meydan Shohada	10873	TSS4
11	TAK	Takhti	11748	-
12	EMH	Emam Hossein	12472	-
13	ENG	Enghelab	13482	-
14	33POL	Si-o-se-pol	14599	TSS5
15	MIR	MIR	15056	-
16	SHA	Doctor Shariati	15748	-
17	AZA	Azadi	16467	-
18	DAN	Daneshgah	17312	TSS6
19	KAR	Kargar	18409	-
20	KUE	Kuy-e-Emam	19279	-
21	DEMO	Defa-e- Moghadas	20464	TSS7

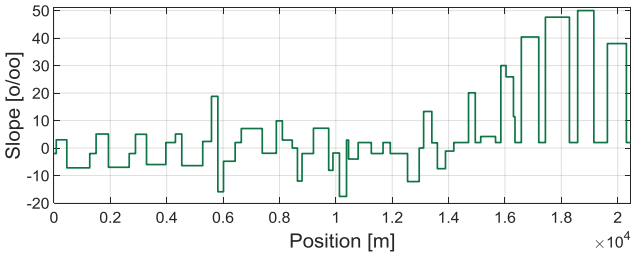


Fig. 5: The gradient profile of Isfahan Metro Line 1 for the up track.

According to the mechanical and dynamic equations of train movement that were introduced in section 2 and also using the route parameters of Isfahan Metro Line 1, the train movement graphs and train timetables are extracted. Because the results obtained in the mechanical simulator section are used as necessary input for the electrical simulator program of the electric railway line. In Fig. 6, the speed-distance profile of the first train of Isfahan Metro Line 1 is represented for the up direction. Furthermore, the speed limit of the train on the route is presented by the red dashed line. In Fig. 7, the distance-time graph of the first train of Isfahan Metro Line 1 is represented for the up direction. In Fig. 8, the timetable graph is presented for 14 trains on each track. A 2.5 min headway and a 1 min dwell time is considered. By using the train movement graphs, the electric power graphs of the trains are extracted in the double-track railway line. Fig. 9 shows the power demand-distance profile of the first train of Isfahan Metro Line 1 is represented for the up and down directions.

The electrical simulator program for train movement in both directions of Isfahan Metro Line 1 is obtained by Matlab software. For power flow analysis, the simulation time step has been considered 5 sec. The simulation results of a moving train on Isfahan Metro Line 1 are presented for both movement directions. In this program, the voltage and current values in the electric railway network are obtained by using the train timetable and the electric power curves.

The simulation results of the dc electric metro system for a moving train with regenerative braking are shown in Figs. 10-11. It is assumed that all traction substations can return the braking energy to the power grid. Therefore, in braking conditions, the braking energy of the train is used to provide the power of the auxiliary equipment, and its surplus is returned to the power grid.

In Fig. 10(a), the actual train power waveform, the pantograph voltage, and the train current are shown during the train movement from the GHO to the DEMO stations taking into account the regenerative braking. In Fig. 10(b), the actual train power waveform, the pantograph voltage, and the train current are shown during the train movement from the DEMO to the GHO stations taking into account the regenerative braking.

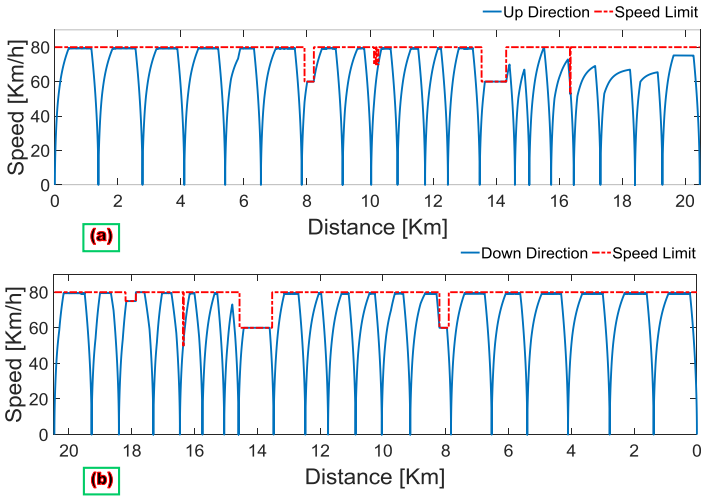


Fig. 6: The speed-distance profile of Isfahan Metro Line 1 for: a) up direction, b) down direction.

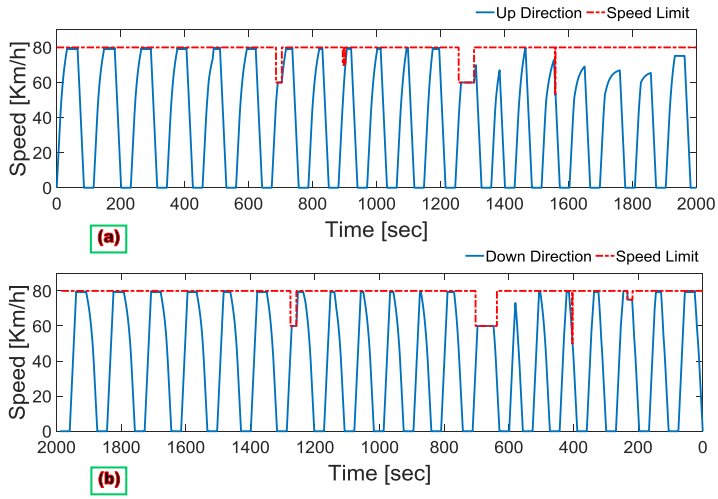


Fig. 7: The speed-time graph of Isfahan Metro Line 1 for: a) up direction, b) down direction.

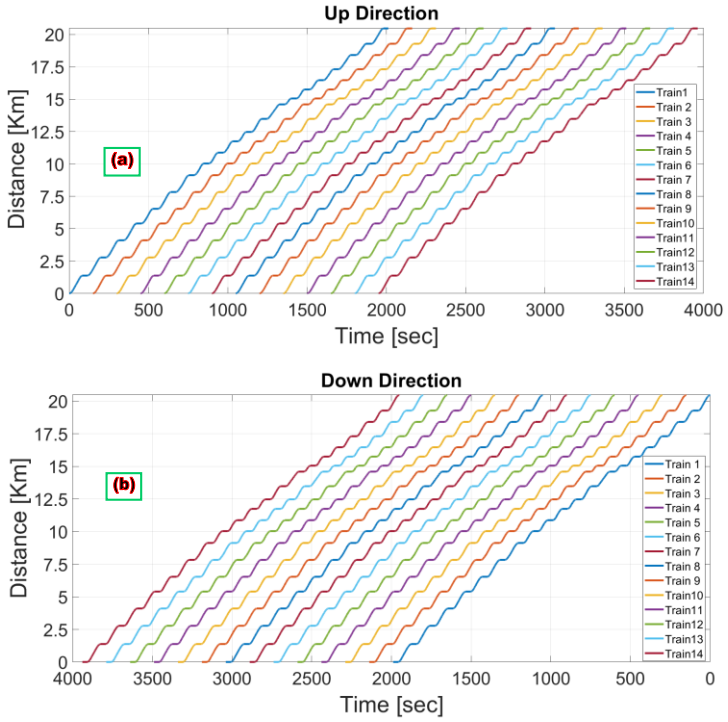


Fig. 8: The timetable graph of Isfahan Metro Line 1 for: a) up direction, b) down direction.

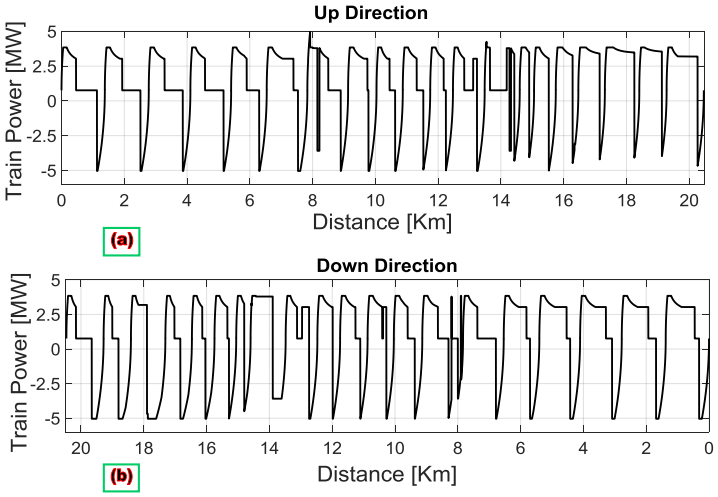


Fig. 9: The power demand-distance profile of Isfahan Metro Line 1 for: a) up direction, b) down direction.

As it is obvious, the train current is consistent with the electric power demand of the train, which shows the correctness of the power flow. But the pantograph voltage has an opposite relationship with its electric power demand. In this way, at the time of maximum power absorption from the electrical network, due to the increase in the current of the overhead network, the pantograph voltage drops and is at its minimum value.
Also, in braking times, the excess power of the train is injected into the electrical network, and thus, overvoltage and undervoltage occur in the overhead contact line and the substation busbar. Therefore, at the maximum braking power of the train, the pantograph voltage reaches its maximum. The obtained results show the correctness of the electric power flow program of the train

In Fig. 11, the voltage waveform of the traction substations is shown during the train movement from the GHO station to the DEMO station taking into account the regenerative braking. As it is clear, with the train movement on the railway track, the voltage of the busbar connected to the traction substation fluctuates. The highest amount of fluctuation is related to the substation that is closest to the train. As the train moves away from the traction substations, the voltage fluctuations decrease and vice versa.

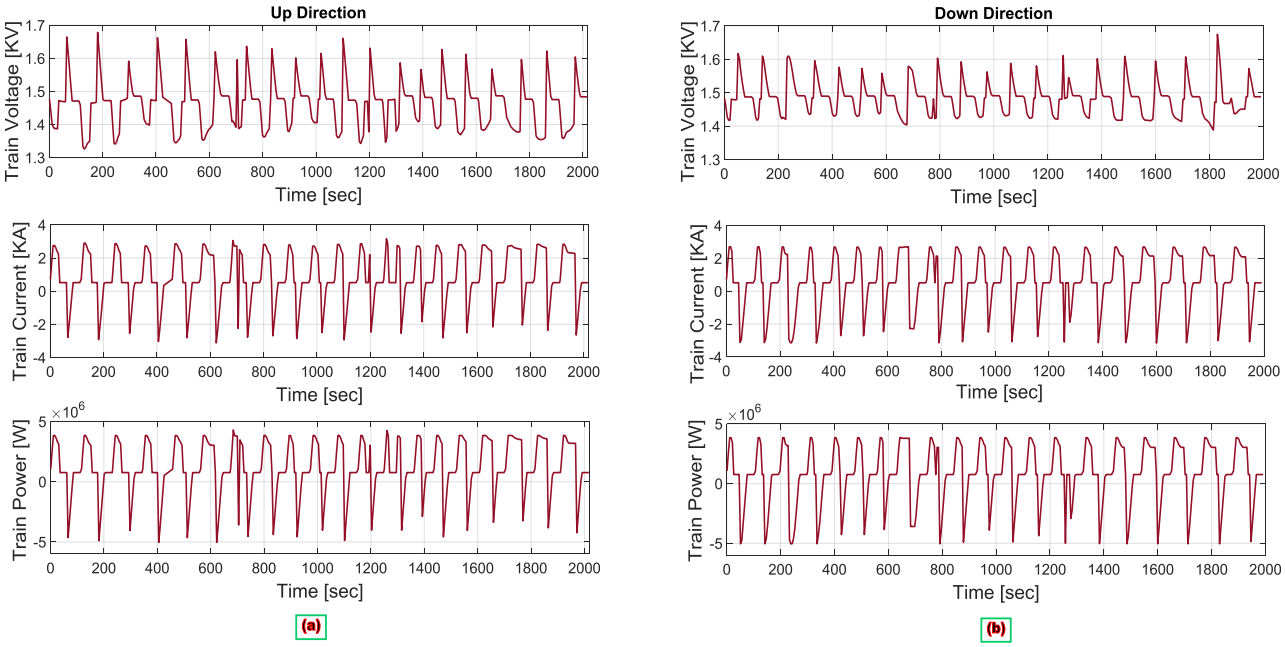


Fig. 10: The pantograph voltage, the pantograph current, and the train power demand during a single train movement on: a) up track, b) down track.

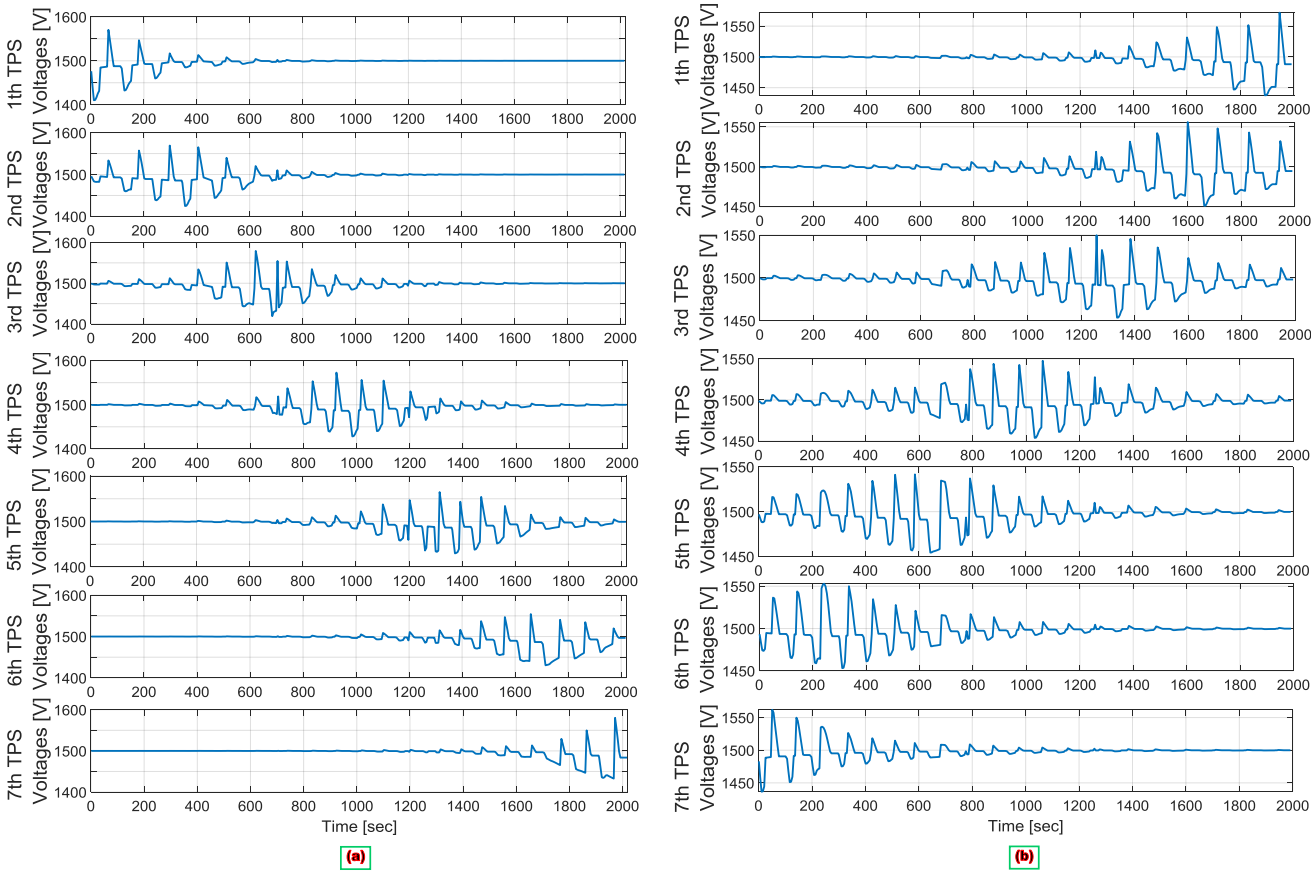


Fig. 11: The substations' voltages during a single train movement on: a) up track, b) down track.

Conclusion

The principal purpose of this paper is to suggest a DC electric railway network simulator including the train movement and traction power network. Simulation results based on the Isfahan Metro line are illustrated. Train current, voltage, and power assessment results manifest the simplicity and effectiveness of the proposed method for planning and operation investigations. The simulator can provide train current and voltage at any position in the track and additionally, the substation voltages by the train movement. The proposed model contains driving control modes with coasting control. Furthermore, the model includes regenerative braking on trains that cause overvoltage in the feeding line and substation busbars. The future work will focus on developing a multi-train simulator for the DC electric railway system including detailed energy consumption evaluation. Furthermore, over-voltage and under-voltage control strategies as well as rail potential assessment will be investigated.

Author Contributions

All authors contributed to the study conception and interpretation of results. P. Hamedani carried out the simulation results, wrote the manuscript, and supervised the project.

Acknowledgment

This work has been supported by the Center for International Scientific Studies & Collaborations (CISSC), Ministry of Science, Research, and Technology of Iran.

The authors gratefully acknowledge the Isfahan Regional Metro Company for providing the simulation parameters of Line 1.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

OCL	Overhead Contact Line
OCS	Overhead Contact System
RCR	Return Current Circuit

References

[1] M. Brenna, F. Foiadelli, D. Zaninelli, Electrical Railway Transportation Systems, Wiley-IEEE Press, 2018.

[2] A. A. Mohamed, A. Arshan Khan, A. T. Elsayed, M. A. Elshaer, Transportation Electrification: Breakthroughs in Electrified

Vehicles, Aircraft, Rolling Stock, and Watercraft, Wiley-IEEE Press, 2023.

[3] T. Fella, C. J. Goodman, P. Weston, (2010), "Validation of multi train simulation software," presented at the IET Conference on Railway Traction Systems (RTS), 2010.

[4] T. Kulworawanichpong, "Multi-train modeling and simulation integrated with traction power supply solver using simplified Newton–Raphson method," J. Mod. Transport, 23: 241-251, 2015.

[5] M. Chymera, C. J. Goodman, "The calculation of train performance," in Proc. IET 13th Professional Development Course on Electric Traction Systems: 1-13, 2014.

[6] Z. Wu, C. Gao, T. Tang, "An optimal train speed profile planning method for induction motor traction system," Energies, 14(16): 5153, 2021.

[7] D. Hu, Y. Yan, Z. Li, "Optimization methodology for coasting operating point of high-speed train for reducing power consumption," J. Cleaner Prod., 212: 438-446, 2019.

[8] Y. Arikan, E. Çam, "Optimizing of speed profile in electrical trains for energy saving with dynamic programming," presented at the International Symposium on Multidisciplinary Studies and Innovative Technologies, 2019.

[9] N. Zhao, C. Roberts, S. Hillmansen, Z. Tian, P. Weston, L. Chen, "An integrated metro operation optimization to minimize energy consumption," Transp. Res. Part C, 75: 168-182, 2017.

[10] Q. Lai, J. Liu, A. Haghani, L. Meng, Y. Wang, "Energy-efficient speed profile optimization for medium-speed maglev trains," Transp. Res. Part E Logist. Transp. Rev., 141, 2020.

[11] P. Martínez Fernández, I. Villalba Sanchís, V. Yepes, R. Insa Franco, "A review of modelling and optimisation methods applied to railways energy consumption, " J. Cleaner Producti., 222: 153-162, 2019.

[12] L. Yang, K. Li, Z. Gao, X. Li, "Optimizing trains movement on a railway network," Omega, 40(50): 619-633, 2012.

[13] V. Martinisa, M. Gallob, "Models and methods to optimize train speed profiles with and without energy recovery systems: A suburban test case," Procedia Social Behav. Sci., 87: 222-233, 2013.

[14] K. Bih-Yuan, L. Jen-Sen, "Solution of DC power flow for nongrounded traction systems using chain-rule reduction of ladder circuit Jacobian matrices," ASME/IEEE Jt Railr Conf., 2002.

[15] T. Yii-Shen, W. Ruay-Nan, C. Nanming, "Electric network solutions of DC transit systems with inverting substations," IEEE Trans. Veh. Technol, 47(4): 1405-1412, 1998.

[16] C. T. Tse, K.L. Chan, S. L. Ho, S. C. Chow, W. Y. Lo, "Tracking techniques for DC traction loadflow," in Proc. International Conference on Developments in Mass Transit Systems: 286-290, 1998.

[17] T. Kulworawanichpong, "Simplified Newton–Raphson power-flow solution method", Int. J. Electr. Power Energy Syst., 32(6): 551-558, 2010.

[18] H. Alnuman, D. Gladwin, M. Foster, "Electrical modelling of a DC railway system with multiple trains," Energies, 11(11): 3211, 2018.

[19] Z. Tian, S. Hillmansen, C. Roberts, P. Weston, N. Zhao, L. Chen, M. Chen, "Energy evaluation of the power network of a DC railway system with regenerating trains," IET Electr. Syst. Transp., 6(2): 41-49, 2016.

[20] C. L. Pires, S. I. Nabeta, J. R. Cardoso, "ICCG method applied to solve DC traction load flow including earthing models," IET Electr. Power Appl., 1(2): 193-198, 2007.

[21] Y. Cai, M. R. Irving, S. H. Case, "Iterative techniques for the solution of complex DC-rail-traction systems including regenerative braking," IEE Proc. Gener. Transm. Distrib., 142(5): 445-452, 1995.

[22] Y. Cai, M.R. Irving, S.H. Case, "Modelling and numerical solution of multibranched DC rail traction power systems," IEE Proc., Electr. Power Appl., 142(5): 323-328, 1995.

[23] R. A. Jabr, I. Džafić, "Solution of DC railway traction power flow systems including limited network receptivity," IEEE Trans. Power Sys., 33(1): 962-969, 2018.

[24] M. Popescu, A. Bitoleanu, "A review of the energy efficiency improvement in DC railway systems," Energies, 12: 1092, 2019.

[25] Z. Tian, P. Weston, N. Zhao, S. Hilmansen, C. Roberts, L. Chen, M. Chen, "System energy optimisation strategies for metros with regeneration," Transp. Res. Part C Emerg. Technol., 75: 120-135, 2017.

[26] S. Lin, D. Huang, A. Wang, Y. Huang, L. Zhao, R. Luo, G. Lu, "Research on the regeneration braking energy feedback system of urban rail transit," IEEE Trans. Veh. Technol., 68(8): 7329-7339, 2019.

[27] M. Davoodi, H. Jafari Kaleybar, M. Brenna, D. Zaninelli, "Energy management systems for smart electric railway networks: A methodological review, " Sustainability, 15(16): 12204, 2023.

[28] X. Shen, H. Wei, T. T. Lie, "Management and Utilization of Urban Rail Transit Regenerative Braking Energy Based on the Bypass DC Loop," IEEE Trans. Transp. Electrif., 7(3): 1699-1711, 2021.

[29] S. Park, SR. Salkuti, "Optimal energy management of railroad electrical systems with renewable energy and energy storage systems," Sustainability, 11(22): 6293, 2019.

[30] F. Lin, S. Liu, Z. Yang, Y. Zhao, Z. Yang, H. Sun, "Multi-Train energy saving for maximum usage of regenerative energy by dwell time optimization in urban rail transit using genetic algorithm," Energies, 9(3): 208, 2016.

[31] Z. Tian, N. Zhao, S. Hillmansen, S. Su, C. Wen, "Traction power substation load analysis with various train operating styles and substation fault modes," Energies, 13(11): 2788, 2020.

[32] Z. Tian, "System energy optimisation strategies for DC railway traction power networks," Ph.D. dissertation, Dept. of Electronic, Electrical and Systems Eng., University of Birmingham, UK, 2017.

Biographies

Pegah Hamedani was born in Isfahan, Iran, in 1985. She received B.Sc. and M.Sc. degrees from University of Isfahan, Iran, in 2007 and 2009, respectively, and the Ph.D. degree from Iran University of Science and Technology, Tehran, in 2016, all in Electrical Engineering. Her research interests include power electronics, control of electrical motor drives, supply system of the electric railway (AC and DC), linear motors & MAGLEVs, and analysis of overhead contact systems. She is currently an Assistant Professor with the Department of Railway Engineering and Transportation Planning, University of Isfahan, Isfahan, Iran. Dr. Hamedani was the recipient of the IEEE 11th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC'20) best paper award in 2020.

- Email: p.hamedani@eng.ui.ac.ir
- ORCID: 0000-0002-5456-1255
- Web of Science Researcher ID: AAN-2662-2021
- Scopus Author ID: 37118674000
- Homepage: https://engold.ui.ac.ir/~p.hamedani/

Seyed Saeed Fazel was born in Iran, in 1966. He received the M.Sc. degree in Electrical Engineering from the Iran University of Science and Technology, Tehran, Iran, in 1993, and the Ph.D. degree in Electrical Engineering from the Berlin University of Technology, Germany, in 2007. He spent four years (1994–1998) as an Engineer with Jahad Daneshgahi Elm Va Sanat (JDEVS). Since 1998, he has been an Assistant Professor with the Iran University of Science and Technology. His research interests include power electronics, medium voltage converter topologies, and electrical machines.

- Email: fazel@iust.ac.ir
- ORCID: 0000-0002-7220-0146
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: http://www.iust.ac.ir/content/3123/Dr.fazel

Mahmoud Shahbazi received the B.Sc. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2005, the M.Sc. degree in Electrical Engineering from the Amirkabir University of Technology, Tehran, Iran, in 2007, and the Ph.D. degrees in Electrical Engineering from the Université de Lorraine, Nancy, France and the Sharif University of Technology, Tehran, Iran, in 2012. He is currently an Associate Professor in Electrical Engineering with the Department of Engineering, Durham University, Durham, U.K. His research interests include renewable energy integration, microgrids, power electronic converters and health monitoring, and fault-tolerant operation in these systems.

- Email: mahmoud.shahbazi@durham.ac.uk
- ORCID: 0000-0002-6057-3228
- Web of Science Researcher ID: NA
- Scopus Author ID: 37091679200
- Homepage: https://www.durham.ac.uk/staff/mahmoud-shahbazi/

How to cite this paper:
P. Hamedani, S. S. Fazel, M. Shahbazi, "Modeling and simulation of DC electric railway system with regenerative braking: A case study of isfahan metro line 1," J. Electr. Comput. Eng. Innovations, 12(2): 439-448, 2024.

DOI: 10.22061/jecei.2024.10665.728

URL: https://jecei.sru.ac.ir/article_2098.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Review Paper

A Survey Study on Intrusion Detection System in Wireless Sensor Network: Challenges and Considerations

M. Hosseini Shirvani [*]***, A. Akbarifar***

Department of Computer Engineering, Sari Branch, Islamic Azad University, Sari, Iran.

Article Info

Article History:
Received 12 February 2024
Reviewed 21 March 2024
Revised 17 May 2024
Accepted 26 May 2024

Keywords:
Intrusion detection system
Security architecture
Anomaly based detection
Misuse based detection
Specification based detection

[*]Corresponding Author's Email Address:
mirsaeid_hosseini@iausari.ac.ir

Abstract

Background and Objectives: Wireless sensor networks (WSNs) are ad-hoc technologies that have various applications in different industries such as in healthcare systems, environment and military surveillance, manufacturing, and IoT context in general. Expanding the scope of sensor network applications has led researchers to develop solutions to provide sustainable communications and networks for distributed environments, as well as how to secure these methods with limited resources.

Methods: The lack of infrastructure space and the vulnerable nature of these networks make it difficult to design security models and algorithms for them. So, to run the sensor network in safe mode, any type of attack must be detected before any security breach is materialized. According to the importance of the network and also the nature of the sensor networks along with the critical challenge of energy consumption, solutions and defensive lines such as intrusion prevention and intrusion detection systems will be selected.

Results: This paper surveys subjectively the intrusion and anomaly detection system in WSNs to determine potentials and challenges for further processing. Therefore, designing an efficient and optimal intrusion detection solution applicable to wireless sensor networks, IoT, and other ad-hoc networks has been a major challenge that will help the researcher to design or choose the best approach for their future research.

Conclusion: This research also paves the way of interested researchers to find existing challenges and shortcomings for further processing.

Introduction

Wireless sensor networks (WSNs) have more characteristic features such as limitations on energy resources, bandwidth, and storage memory [1]. Regarding the limited computation conditions, security approaches in traditional networks have not been useful in WSNs. It has been obvious that the limited features of WSN imply that it does not use IP protocol for network operations. Hereupon, the design of the novel and effective detection approach in WSN has been a big challenge for researchers. Even Though WSN has significant features in terms of operations such as low installation costs and lack of care for network operations, at the physical defensive line, there has been no gateway, router, or switch to monitor the information flows. On the other side, limitations on energy sources pose a great challenge to the security of these networks [2], [3]. Hence, the security architecture of proposed networks has been a big concern, especially for applications where non-functional requirements such as availability, integrity, confidentiality, reliability etc., have prime importance [2]. Numerous researchers have investigated security attacks as well as intrusion detection systems (IDS) in the WSN context [4]-[9]. Therefore, in order to run

the wireless sensor network in safe mode, any unauthorized access or manipulation of node information, traffic and transit interactions must be actually detected at the correct high rate. Along with the protection of WSN, detection and prevention mechanisms such as cryptography algorithms and IDS should be considered. It has been regarded that the other prevention systems such as intrusion prevention systems (IPS) and honeypots have the requirement to the effective algorithms in order to reduce the power source and other limited features of WSNs. Security in the highly valuable network has been a major requirement for the researchers in the sensors era. For instance, in the healthy control system usage, the patient records should not be accessed by third parties. On the other hand, it has been so important to utilize security mechanisms in military-based features such as battlefield surveillance, minefields and etc. importance of the military usage will appear when the lack of space in the network will casualties' friend armies, the importance of these networks will increase [10]-[12]. Most of the detection mechanisms in traditional networks haven't been able to be directly utilized in the WSN [13], [14]. It has been obvious that traditional approaches have been integrated for wired and IP-based networks and these solutions have not been directly applicable to the sensor networks. Hence, the researchers should consider non-functional requirements such as lack of infrastructure, dynamic topology changes, easiest physical access, extended routing protocols, and limited computation sources [15], [16]. The designing of IDS in WSN has the following requirements that researchers should consider:

- Lack of infrastructure,
- Dynamic Network topology changes,
- Physical facile access,
- Different routing protocols,
- Resources limitation.

To realize the above functional requirements, research on real-time IDS has been constantly increasing. Due to the nature of the problem and the requirement for detection, limited research on detection and meta-discovery algorithms to optimize intrusion detection systems has been presented. Limited and also valuable research has been presented on heuristic and meta-heuristic algorithms to optimize intrusion detection systems. An ant colony optimization (ACO) algorithm based on an Ad-hoc On-Demand Distance Vector (AODV) protocol has been applied for the detection of Blackhole attacks. The authors have applied Grover quantum meta-heuristic algorithms to optimize attack path detection [15]. The authors' proposed approach has been capable of improving some fundamental network parameters such as throughput, end-to-end delay, and packet delivery ratio in comparison with other approaches [15]. Binitha and Sathya have conducted extensive research on the optimization of bio-inspired algorithms [16]. They proposed an overview of evolutionary algorithms, genetic algorithms, genetic programming, evolutionary strategy, and also in the swarm intelligence (SI) category, they have also discussed on particle swarm optimization (PSO) algorithm, Ant colony optimization (ACO), bacteria foraging algorithm (BFA), Glowworm Swarm Optimization (GSO), shuffled frog-leaping algorithm (SFLA), the intelligent water drops algorithm (IWDA), Ford-Fulkerson algorithm (FFA), Feasible Solution Algorithm (FSA). In their surveyed paper, a complete and separate review of the mentioned algorithms in the form of a table with a description of operators, application areas, and control parameters has been investigated. Fu et al. have described the anomaly-based detection framework for hierarchical networks by adapting their framework to the risk theory and negotiation selection algorithm in AIS [17]. They have provided a framework for misbehavior detection utilizing the advantages of artificial intelligence and fuzzy theory to encounter the resource constraints in the typical sensor network.

They have also compared their method with the Watchdog approach. Their method has been able to detect the correct and high rate as well as the incorrect detection rate with the low rate [17]. An approach exploring the adaptability of Bio-inspired methods and their application in the field of computer networks has been published in [7]. According to this research, strategies that mimic the system as much as possible run the risk of inheriting behavioral characteristics as well as environmental constraints. This process will eventually confront the phenomenon of the evolution of nature and the limitations of the physical world [18]. As mentioned earlier concerning to the importance of several applications and industries along with regarding to the limitations of WSN, this survey study reviews published literature and proposes subjective classifications and comparisons of papers.

Then, the challenges & potentials along with research gaps are outlined to guide the future direction for improving existing schemes to bridge available gaps. Therefore, the remainder of this paper is structured as follows. Section 2 stipulates the importance of security in the WSN context. Section 3 introduces intrusion detection systems (IDSs) along with our subjective classification. Section 4 provides applications of machine learning (ML) techniques in IDS schemes. An informative example is brought in section 5 to show the effect of attacks on WSNs. Section 6 pays on intrusion detection systems in new IoT and AI technologies. Section 7 concludes the paper.

The Importance of Maintaining Security in Networks with High Information Sensitivity

Due to the cheap and simple installation capability, wireless sensor networks have been applied to various branches of science and technology. Nowadays, the increasing applications of these networks in important areas such as segmentation of information collection about human behavior and activities can be divided into health control systems, battlefield monitoring, and identification, as well as highway traffic and IoT applications. In the field of physical monitoring and environmental phenomena, it can mention areas of application, such as the ocean, wildlife, earthquake, pollution, forest fires, and water quality. In the field of monitoring industrial sites, applications such as building safety, the performance of production devices have been available. On the other hand, the security in WSN is very sensitive, important, and also very essential functional requirement parameter that the researcher should fully consider it. For example, patient health records should not be disclosed to a third party. On the other hand, securing the WSN has been very important in tactical (such as military) applications. The importance of security in the network has been multiplied when the safe space in the network leads to casualties in the friendly forces on the battlefield.

Hence, counter-attack approaches in wired and wireless networks have been included three main following components:

• Prevention (Maximum defense against attacks): this step has been intended to prevent any attack from taking place before it occurs. Therefore, the proposed technique must defend against the target of the attack.

• Detection (Awareness of the attack Presence): if the adversary devises and exploits a strategy to bypass the techniques applied by the deterrent, the defense mechanism against the attack will fail. At this time, security solutions have been immediately transferred to the attack detection phase and especially the identification phase.

• Mitigation (performing serious action against the attack): in the final step, the mitigation phase has been neutralizing any attacks before the occurrences by clearing infected groups, and then it secures the network.

In network security terminology, the intrusion has an unwanted and unauthorized activity that appears active or passive. It should be considered that in the sensor network, these attacks have been divided into the following categories:

• Active attacks (Sending the malicious packets, delete packets and worm attacks),

• Passive attacks (Eavesdropping and information gathering).

In a secure and sensitive system, if the first line of defense, i.e., intrusion prevention, has not worked properly, the second line of defense, the intrusion detection system, will play a vital role. IDS has been detected for each misbehavior and malisons activates or behavior that has been created by the members. In each security architecture plan, the IDS has been provided all of the following information for other tracking systems:

• Intruder detection,

• Intruder location detection (solitary or regional in WSN),

• Intrusion time

• Intrusion mechanism (active or passive),

• Intrusion type (Worm based attack, flooding, etc.),

• Permeable layer (physical, data link, or network vulnerable layers).

This information has been very effective in the third line of defense and clearly in the mitigation and reaction phase and compensates and corrects the results of the attacks according to the information that has been obtained from the intruder attack. Hence, IDS has very critical for network security architecture. Overall, the researchers and developers should consider IDS as the main requirement for critical systems. It has been obvious that the limited life cycle of the sensor has been playing a vital role in security decisions. These decisions have been more strategic because while achieving functional and non-functional requirements, it has been created an impact on the network platform. Therefore, the requirement to pay attention to the needs of the network has been recommended to researchers before any action in the requirements assessment phase [19]-[26].

Intrusion Detection System (IDS) and Subjective Classification

In a system or a network, each type of unauthorized and unproved activity has been referred to as intrusion. An IDS has been a set of tools, methods, and resources to help identify, access, and report intrusion. Intrusion detection has been generally part of the system's global protection that has been configured around the system and has not an individual criterion for protection. In general, the first line of defense has been the IPS which included solutions such as encryption, authentication, access control, secure routing, etc. Infiltration and compromise with a node lead them to the emergence of confidential information such as security keys for the adversary. Hence, IDS has been designed to detect secure system resources before an Intruder attack to detect intrusions. From a security point of view, the intrusion detection system has been always in the second line of defense [13]. Hence, functional requirements on WSN have been as follows:

• Low false positive rate: This rate has been estimated by calculating the percentage of normal changes that have been detected as misbehavior and abnormality.

• High True Positive Rate: This rate has been estimated by calculating the percentage of anomalies that have been detected.

Fig. 1 illustrates our subjective classification framework, derived from literature, of IDS utilization in the WSN context. In forthcoming subsections, each branch of the subjective framework is elaborated.

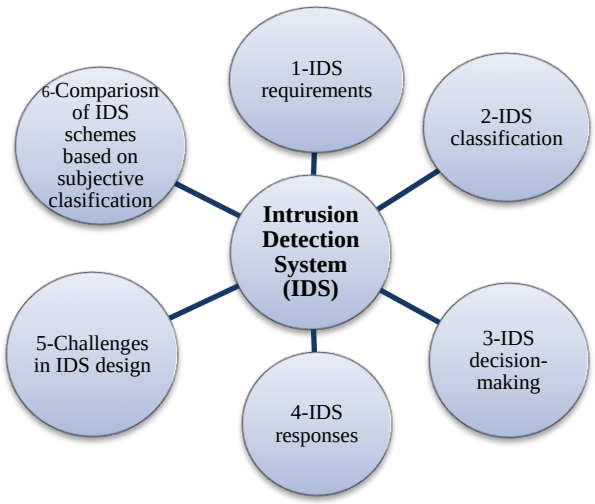


Fig. 1: The subjective classification framework for IDS in WSN context.

Intrusion Detection System Requirements

In the IDS designing process, the following non-functional requirements should be considered [14]:

• Do not add new vulnerabilities to the system,

• Limited requirements for system resources and failure to degrade system performance by introducing overheads,

• Continuous execution and transparent presence for users and the system,

• Utilizing the standards for cooperation,

• Reliability and also minimum rate of false-positive (FP) and false-negative (FN) in the detection phase.

Intrusion Detection System Classification

According to Fig. 2, IDS has been classified into the type of intruder locations, type of intrusion, detection methodology, data source examined, data collection site processing, infrastructure, and scope of application. Attempts have been made to briefly explain each of these sections based on previous research [27]-[30].

Type of Intruder Locations Placement

Generally, intruder locations placement in a network has classified as follows:

• External intruder: it has been obvious that outside the network, with various attacks, intruders have been tried to gain unauthorized access to the network.

• Internal intruder: in this category, a node has been tricked and used to place on the network.

On the other hand, on the ADHOC networks, the internal attacks have been utilized of two following nodes:

• Selfish node: This node has been utilized network resources, but it does not cooperate and has not directly damaged other nodes.

• Malicious node: This node aimed to harm and flood other nodes by creating a denial-of-service attacks (DOS) quite the same normal type of DOS attacks in IP-based networks.

An IDS has been able to detect external and internal intrusion. Researchers should consider that internal intrusion detection and also discovery have been more difficult than the external type of it. It has been obvious that internal intrusions have the key parameters that have been needed to thwart the precautionary measures taken by the authentication mechanism [36]-[38].

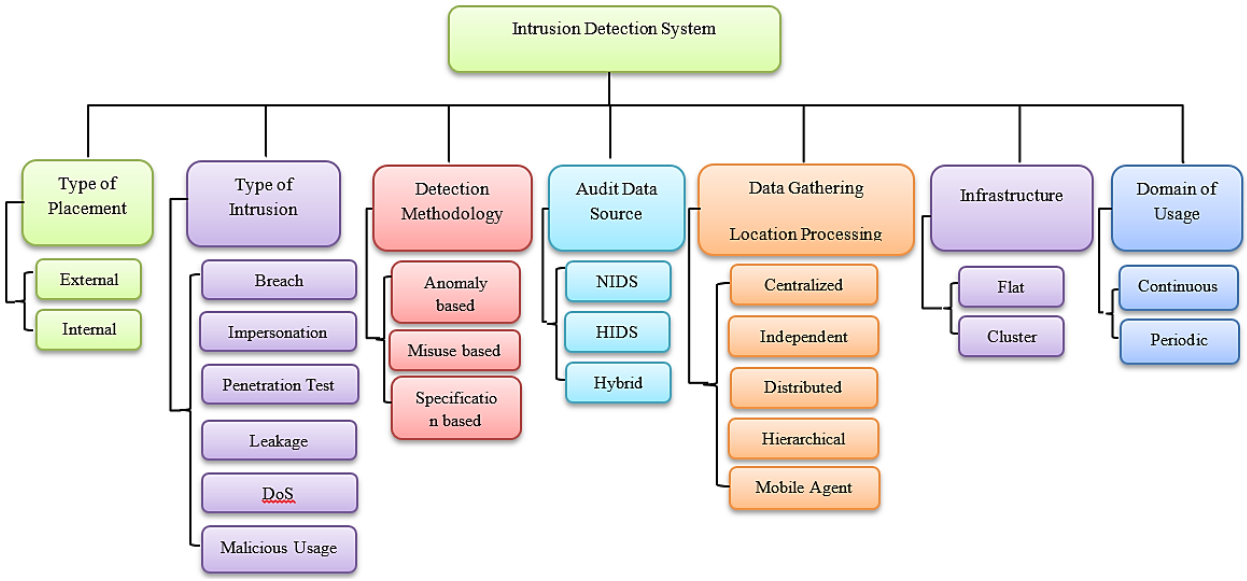


Fig. 2: Intrusion Detection System Classification based on type of intrusions.

Type of Intrusions

The intrusion in a network has been occurred in the following various forms:

• Attempted break-in: Create a search for unauthorized access to the network,

• Masquerade: Acting and using a fake identity to obtain unauthorized access to the network,

• Penetration testing: Obtaining unauthorized access to the network,

• Leakage: Unintentional information leakage from a network,

• Denial of service: flooding and blocking network resources,

• Malicious application: Aimed at intentionally hitting and damaging network resources.

Although IDS may have been provided more detailed detection solutions for the above attacks, system administrators always want a complete defensive system with the ability to detect all intrusions.

Detection Methodology

From the perspective of security architecture and critical Non-functional requirements such as performance, IDS have been classified into three following categories:

• Anomaly-based detection,

• Misuse-based detection,

• Specification-based detection.

In an anomaly-based detection strategy, the irregular solution has been based on statistical behavioral modeling.

The normal operating behavior of the members has been described for the system and the definite amount of deviation from the normal behavior has been expressed as the flag of irregularity. Disadvantages of this method included the fact that normal profiles must be updated at regular intervals due to rapid changes in network behavior. This model detects intrusion accurately and stably with small and limited values (FP, FN), under conditions where the network has been statistically considered in terms of behavioral pattern.

One of the advantages of this method has its use to detected unknown attacks or attacks encountered [13]-[15]. Based on the processing nature that has been created in the behavioral model, Anomaly detection has been divided into three following categories:

• Statistical based detection,

• knowledge-based diagnosis,

• Machine learning.

The Fig. 3 describes the classification of irregular intrusion detection systems based on their detection algorithms.

In statistically-based IDS, network traffic has been captured and a profile has been generated that represents random and sudden behavior. A reference profile has been also created when the network is in safe condition without attacks presence. After the network has been monitored, profiles have been generated at regular intervals. Hence, by comparing the reference profile, a rank has been generated. If this rating exceeds a certain threshold, the IDS assigns the flag of abuse to it.

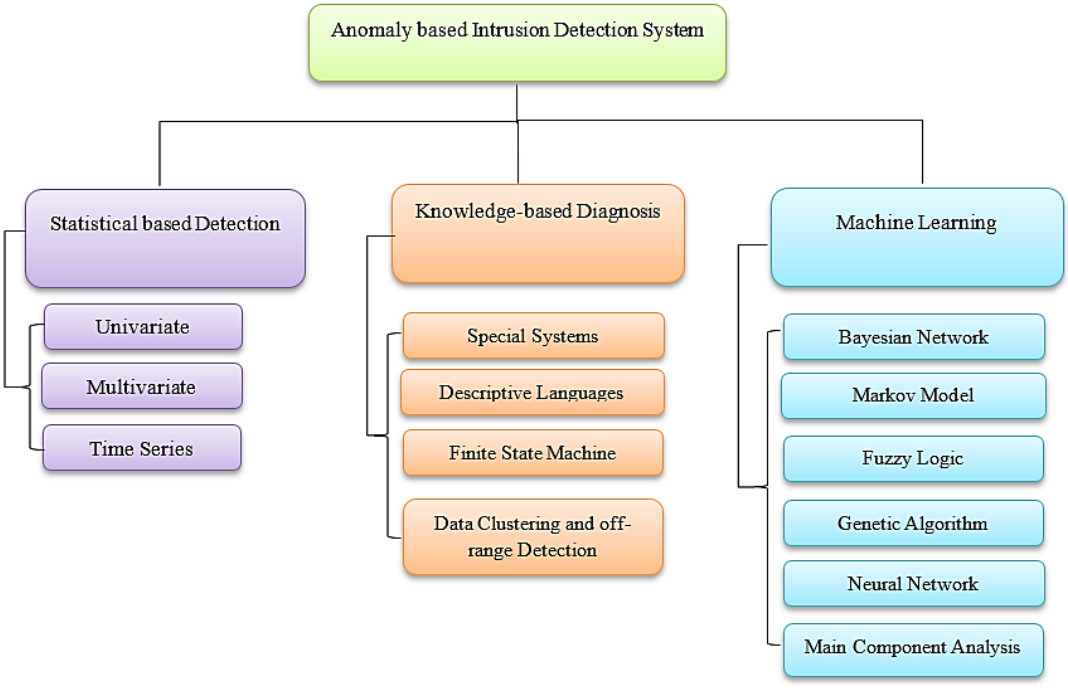


Fig. 3: Classification of Irregular Intrusion Detection System based on Detection Algorithms.

According to Fig. 3, this method has been categorized into the following sections:

- Univariate: The parameters have been modeled as separate Gaussian random variables.
- Multivariate: Correlation between two or more criteria has been considered.
- Time series model: An interrupt timer has been used during an event counter that records the order and time between the arrival of observations as well as their values in a report.

The following described an example of a detection methodology for detecting packet deletion attacks:

The percentage of FP transmissions from node m has been the rate at which packets has sent by node n among packets that sent from node M to node m with T specified time. This process has been calculated by the following equation. Table 1, Has been described the parameters that have been utilized in (1).

$$FP_m = \frac{Forwarded\ Packets}{Packets\ to\ be\ forwarded} = \frac{\#(m,M) - \#([m],M)}{\#(M,m) - \#(M,[m])} \quad (1)$$

In (1), If the Dominator that called Packets to be forwarded hasn't equaled to 0 and also if the value of the $FP_m = 0$, then the event has been recognized as "unconditional package deletion" and m has been also identified as an intruder. If the Dominator of (1), hasn't equaled to 0 and if $FP_m \leq specified\ threshold\ of\ satisfaction$ and if the condition of (2) has been met, then the event has been recognized as non-random package deletion and m has been also recognized as an intruder.

$$0 < FP_m < TF_P < 1 \quad (2)$$

Table 1: Abbreviation that utilized to detecting packet deletion attacks

Abbreviation	Parameter's role
m	Supervised node
M	Nodes to be monitored
$\#(m,M)$	The number of out-band packets from node $\boldsymbol{m}$ where $\boldsymbol{M}$ is the next step.
$\#([m],M)$	Output packets of the source node that called m where $\boldsymbol{M}$ is the next step.
$\#(M,m)$	Output packets of $\boldsymbol{M}$ where m is the next destination.
$\#(M,[m])$	Output packets of $\boldsymbol{M}$ where $\boldsymbol{m}$ is the final destination
FP_m	Percentage of sending packets from $\boldsymbol{m}$

In machine learning anomaly-based detection systems, an explicit or implicit model of the analyzed patterns has been generated. These models have been updated at regular intervals to improve IDS performance based on previous results. To optimize the IDS, the following solutions have been adopted.

- Markov model based on Markov transfer theory,
- Bayesian networks based on possible relationships between variables of interest,
- Fuzzy logic based on approximation and uncertainty,
- Genetic algorithm,
- Principal component analysis (PCA) based on dimensional technique.

In misused-based Detection, signatures, and identifiers (profiles) have been generated from previously known attacks and have been used as a reference for diagnosing future attacks. For instance, an example, of a typical ID and signature would look like the following example:

- Three unsuccessful login attempts in five minutes have been created by a brute force attack.

However, the advantage of this type of detection has been the ability to correctly and effectively detect known attacks. On the other hand, the disadvantage of this method has been that if the attack has a new type and has not already in the profiles, the misuse-based detection has not been able to detect it. These systems have been very similar to antivirus systems that often detect all or all known patterns of attacks. The researcher should be considered that the solution mentioned has been used in closed and non-public structures. Therefore, despite the high detection rate, software architects and developers have been required to utilize this method in their security architectures and networks. On the other hand, the following requirement has been proposed to monitor network anomalies:

- The requirement of interrupt: indicated a latency between the arrivals of two consecutive messages that must be within a certain range.
- The requirement of confidentiality: the passing message must be sent through the middle nodes.
- The requirement of integrity: the sender's main message should not be distracted when it reaches the recipient.
- The requirement of delay: packets must be resent after a specified waiting period.
- The requirement of iteration: Identical messages have been measured from a single node as well as a specific number.
- The requirement of radio's confidentiality rate: messages must have originated only from neighboring nodes.

- Noise rule: The number of collisions for packet transfer must be lesser than the threshold value.

In specification-based Detection, a set of specifications and constraints described the correct operation of a defined program or protocol. Then the implementation of the program has been monitored by considering the defined specifications and limitations [13]. This solution has been provided the ability to detect previously unknown attacks with low FP rates. There have been significant differences between IDS types. Anomaly-based IDS has been tried to detected anomaly behaviors, but misuse detection tried to recognize abnormal behaviors. Specification-based IDS techniques combined the benefits of abnormal detection and abuse detection through the manual development of features and constraints to determine system behaviors. Intrusion-based detection techniques have been similar to irregular detection strategies. In each, the attacks have been detected by deviating from the normal profile. Because feature-based detection techniques have been based on the extension of features and constraints manually, they have a limited false alarm rate compared to the high false alarm rate in anomaly detection. The cost of obtaining a limited false alarm is that it will take a long time to develop the details of the features and restrictions [39]-[42].

Audit Data Source

IDS has been categorized into the following parameters based on the audit data source and depending on the location of the analyzed data:

- A network-based intrusion detection system (NIDS): Actively or passively this system has listened to network communications, then records packets and evaluates packets. The mentioned system has the ability to analyze the entire packets transfer capacity and IP addresses or ports.
- A host-based intrusion detection system (HIDS): HIDS has been able to detect intrusions such as changing critical system files on the host side, repeated attempts to miss-access the host, unreasonable allocation of memory to a particular process, and input-output activities. The HIDS has been performed the detection operation by real-time monitoring of the host system or by checking the log file on the host side.
- Hybrid-based intrusion detection system: The Hybrid system has been Consist of NIDS and HIDS components in an efficient method utilizing mobile agents. Mobile agents referred to each host and check the system file log. Meanwhile, the central agent has been nationwide examining the entire network traffic for the presence of anomalies [59].

Estimation the Location of the Collected Data

Based on the location of the collected data, IDS has been divided into the following categories:

- centralized,
- Stand-alone IDS and independent,
- Cooperative and distributed,
- Hierarchical,
- Mobile agent-based IDS.

In the **centralized** IDS, a centralized computer has been observed all network activity and it has been detected intrusions by analyzing network surveillance activities and data.

In the **Stand-alone** IDS, the system has been executed separately on each node. Network members have been unaware of intrusions that have been occurred around them because a Stand-alone IDS does not allow the node to cooperate or transmit information to each other. They have only worked if they have been alone and independent.

The **Cooperative and distributed** IDS has been proposed for flat infrastructure networks. In this scenario, each node has been executed as an IDS agent which has been participated in the detection of penetration testing and the global response to the network. If a node has been detected an intrusion without evidence or without result, it has been able to independently issue a network warning regarding an attack.

The **Hierarchical** IDS has been proposed for multi-layer network infrastructure such as cluster structure. Cluster heads have been required to monitor their member nodes and also participate in global intrusion detection decision-making operations.

In the **Mobile agent-based** IDS, each mobile agent has been assigned to the selected node in order to create a specific task of the IDS and the intrusion detection has been done in cooperation with these nodes. After a specified period of time or after a specific time has elapsed for the task to be performed, agents have been moved to other predefined nodes to increased network life and also the efficiency of the IDS. Mobile agents have characteristics such as mobility, self-control, and compatibility.

In the **mobile AD-HOC networks**, IDS has been divided into the following parameters:

At the **Flat** infrastructure, all of the nodes have been considered with the same capability and have the ability to participate in routing functions. This facility has suitable for civilian applications such as a conference network or a classroom.

In the **Cluster** infrastructure, Nodes have not been considered as same. Nodes have been subdivided into clusters at a given transfer rate, and then nodes have

been selected a node as a cluster head to centralize routing information for that cluster. Generally, headers have been consisted of many powerful devices with backup batteries to achieve greater transmission range. Hence, headers have been the virtual backbone of the network. Depending on the routing protocol, the middle gateway has been replayed the packets among the headers. This type of infrastructure will be very suitable for military applications with hierarchical commands [67].

Decision Making on the Intrusion Detection System

There have been two following mechanisms for decision making on an IDS:

- Stand-alone decision making,
- Cooperative decision-making.

An IDS deduces four non-zero probability decisions as a result of the decision-making process in an event.

- Intrusive and abnormal (False Negative): There has been an intrusion into the system, but the IDS has been failed to detect it and it recognizes the event as abnormal.
- Not intrusive and abnormal (False Positive): There has been no intrusion into the system, but the IDS concluded that a normal event is abnormal.
- Not intrusive and not abnormal (True Negative): There has been no intrusion in the system and the IDS has been concluded the event as abnormal.
- Intrusive and abnormal (True Positive): There has been an intrusion on the system and the IDS has been detected the event as abnormal [71], [72].

Intrusion Response

It has been obvious that the IDS has not met the prevention criteria at the time of the attack and leaves this process to the IPS section [106]. The intrusion detection system has been worked reactively in comparison to the active operation of the IPS. Whenever a production intrusion warning has been issued by an IDS, the following has been raised based on the characteristics of the system:

- Create an audition or review record
- All of the network members, system administrators, and base stations should be notified of intrusion. If possible, the location and identity of the adversary should be stated in the alert message.
- If possible, a reduction strategy should be considered to stop infiltration. For instance, a modified auto action should be generated by the collaboration activity of network members, especially the event neighbor.
- There has been no trusted source and decisions must be made by a colleague.

Challenges and Regulations of Intrusion Detection Systems Design in Sensor Networks

The proliferation of sensor networks has been leading researchers to develop and expand solutions to provide sustainable communications and networks for distributed environments, as well as how to secure these methods with limited resources. The lack of stable infrastructure space such as gates, routers, base stations, etc., makes it very difficult to design security models and algorithms for the sensor network. Limited bandwidth, throughput, battery source has scarce resources that should be considered considerably in the network architecture design phase [108]. To create an intrusion detection system in the sensor network, the system must contain the following requirements:

- Localize auditing:

The IDS in WSN must work with the main data as well as cross-sectional inspection because in WSN there has been no central point that can collect the global data examined. This is separate from the base station.

- Resource constraint:

The IDS must utilize the minimum resources for each network. Communications between two nodes should not saturate the available bandwidth to detect intrusion.

- Lack of trust in the elements:

Unlike wired networks, sensor network security has been easily compromised. Hence, the IDS should not trust every node and element in the network.

- Distributed:

Data collection and analysis must be in multiple situations. In addition, the distributed solution has been applied to implement the correlation detection and warning algorithm.

- Securely:

IDS should be resistant and withstand attacks.

A Comprehensive Comparison Among the Proposed IDS in Literature

In the hierarchical structure, cluster-based IDS, as well as clustering algorithms have been consumed significant network energy to form clusters.

Agent-based IDS has been reduced network load and latency. On the other hand, it has been led to high energy loss in the associated nodes. The cost of communication between the agents and the coordinator has been made it possible to create congestion and bottlenecks in the network.

Rule-based IDS have been Easy to configure and also executed. They have been required to constantly update the rules and regulations to counter new attacks.

Data mining-based IDS can detect new attacks. Unfortunately, these systems have been required to have high computational complexity as well as high power

consumption for their data samples. There has been also a requirement for efficient analysis tools to analyze large amounts of data as well as memory space to store data.

On the Game theory-based IDS, the detection rate has been set by the network security manager utilizing changing the parameters. The disadvantage of this system has been the incompatibility as well as human intervention for sustainable operation, because a wide variety of intrusion detection algorithms have been available, selective intrusion detection solutions must be embedded for the desired features, requirements, and applications based on network hazards [114], [115].

Security has been a functional requirement that required optimal and correct detection of the adversary and satisfaction in accurately determining the exact duration of the attack. The following have been suggestions for specific applications in IDS in the sensor network architecture:

- For itinerant applications, where the sensor nodes have been in motion, it would be appropriate to utilized distributed IDS methods due to their scalability, robustness, and speed.
- For static applications, in a situation where there has been a centralized processing unit in the base station or data in the sink, the utilization of centralized solutions has been appropriate due to their robustness and ability to detection of a wide range of attacks.
- For cluster-based applications, utilizing of hierarchical intrusion detection system would be appropriate.

Various IDS for WSN have been described in Table 2, and include the required network architecture, detection technique, and features of each method.

Related Research on IDS by Incorporation Machine Learning (ML) Techniques

Since the handling of WSNs' challenges has high complexity, various researchers have proposed the utilization of machine learning (ML) on IDS [119], [120]. The ML algorithms can manage huge data with optimum speed and accuracy. These algorithms have been utilized to design the accurate models that specifically were designed for the classification, clustering, and also prediction processes. The ML techniques played a vital role in the IDS for WSNs when it has been utilized in support vector machines (SVMs), Gaussian naive Bayes, and Random Forest logic regression algorithms. The ML can be subjectively categorized as follows:

- Supervised learning,
- Unsupervised learning,
- Reinforcement learning.

It is obvious that firstly all of the ML algorithms have been labeled as training data that specify an input, output data, and some system parameters. The Fig. 4 depicts the ML classification. The supervised learning has been used as regression and also classification model. An unsupervised learning has been used to classify sample sets as well as groupings. In reinforcement learning, agents have been prepared for the learning process by interacting with the environment. It should consider that the combination of supervised learning and unsupervised learning has been considered semi-supervised and as a hybrid algorithm inherited all the main functions of the mentioned field.

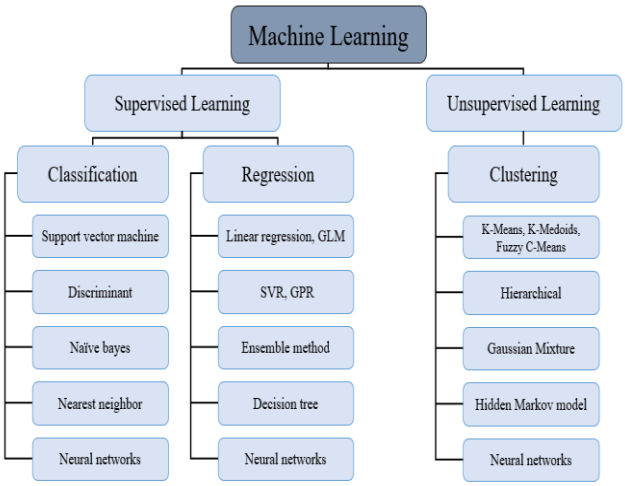


Fig. 4: Classification of ML algorithms [20].

On the other side, researchers have proposed a multi-core ML-based IDS [121].

In this structure, a prototype has been considered a hierarchical intrusion detection model. The sequence and attractiveness of multi-core functions promise to reduce detection time and high detection rates. Fig. 5 illustrates a multi Kernel-Extreme learning machine (MK-ELM) algorithm.

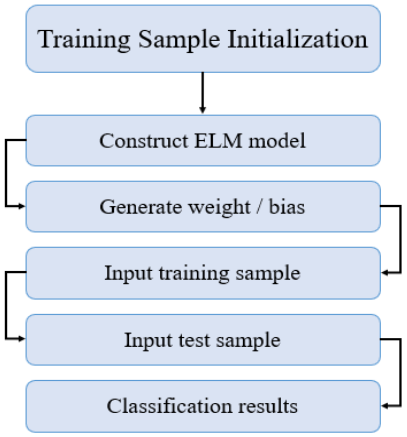


Fig. 5: MK-ELM algorithm Diagram [31].

Table 2: Comparison between researched IDS in the last 20 years

Proposed IDS	Architecture	Detection methodology	Distinctive feature
[6]	Distributed	Rule based	Scalable, powerful and accelerates detection.
[14]	Centralized	Anomaly based	Scalable and Reliable detection on black hole utilizing meta-heuristic and quantum speedup.
[8]	Hierarchical	Automatic scout	Relying on the diffusion nature of sensor node communications and the use of high node diffusion densities.
[9]	Hierarchical	Rule based	Monitor nodes and routing tables.
[22]	Hierarchical	Rule based	Energy storage, extension of network life, inability to add nodes to the network.
[23]	Hierarchical	Rule based	Combine existing solutions in order to achieve more complete solutions.
[37]	Hierarchical	Specification based	Achieve optimal performance or centralized design.
[56]	Distributed	Rule based	Detection of selective transmission attacks and black holes based on the presence of an intruder.
[60]	Centralized	Anomaly based	Only able to detect wormhole attacks.
[61]	Hierarchical	Anomaly based	Focus on collected data to maintain node or connection security.
[70]	Stand-alone	Anomaly based	Local detection, lack of node notification of attack.
[90], [91]	Hierarchical	Game theory	Monitor only at one time and one cluster.
[98]	Stand-alone	Rule based	Detection of anomalies in all network layers.
[101]	Distributed	Anomaly based	Extract sensor network stability by information from neighboring nodes.
[109]	Distributed	Rule based	Minimize processing overhead when detecting abnormalities within the network.
[112]	Centralized	Statistics based	Using heuristic ranking algorithms to determine undesirable nodes in the network.
[113]	Hierarchical	Statistics based	Use a cluster-based hierarchical secure management protocol to identify malicious and selfish nodes.
[116]	Hierarchical	Anomaly based	Efficient detection that utilizing fuzzy C-means clustering
[117]	Hierarchical	Anomaly and Misuse based	Optimal detection accuracy with the ability to Determine the type of attacks
[118]	Distributed	Anomaly based	Covering the important non-functional requirements such as reliability, efficiency, scalability, Interoperability, low overhead and etc.
[127]	Hierarchical	Anomaly based	Efficient energy consumption.
[128]	Hierarchical	Anomaly based	High accuracy in detection phase with low overhead.
[129]	Distributed	Anomaly and Misuse based	Low False Positive rate and high accuracy with low complexity.
[130]	Hierarchical	Anomaly and Misuse based	High accuracy in detection phase by a Composition between SVM classifier and signature-based approach with low overhead

Other researchers have proposed various reviews of the application of game theory in sensor network security [32]. According to the various applications of common game theory approaches for the security era, the design method has been divided into the following categories:

- Denial of service (DOS) prevention,
- Intrusion detection,
- Upgrade security level,
- Coexistence with destructive nodes.

The theory analyzes a myriad of possible scenarios before creating an operation. Hence the decision-making process has the modeling ability. The summary and also the main axes of the researches on the game theory has been described as the following categories:

First category:

- There are No-Cooperative game [33],
- Cooperative game [34],
- Repeated game to prevent DOS attack [35].

Second category:

- There has been no cooperative and also Markov games models for intrusion detection [36]-[40].
- There have been Auction theory and coalitional game theories to strengthen security [41], [42].
- There has only signaling game approach to coexist with malicious nodes [43]. In [45], an overview of IoT intrusion detection systems has been provided. They have been described the IoT architecture as shown on the Fig. 6.

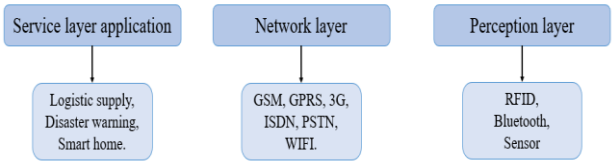


Fig. 6: IoT architecture [45].

In the same field, researchers have proposed a rule-based IDS based on the proposed event processing model (EPM) to solve the problem of real-time intrusion detection in the IoT [46]. Obviously, this model has been based on the EPM in which the rules that have been stored in the rules pattern repository (RPR) and then considered as a reference. The mentioned approach in relation to the existing IDS has consumed more CPU resources while consuming less memory and minimizing processing time. On the other side, anomaly IDS for WSN has been proposed. The steps of this IDS have been as follows:

- Local audit phase: evaluation of the packets to validate reputable neighbors.
- Rule application: this step has been worked on promiscuous mode.
- In the third step: routing attacks have been detected by validating the collected data.

It is obvious that the proposed mechanism is able to just detect routing attacks which means the weakness of this approach. From our viewpoint, this solution should not fully be considered in WSN because of its disadvantages and its weakness. In other anomaly-based IDS, a soft processing and computing system approach was proposed [47]. The main purpose of this research is to increase the performance of the system and identify any event strongly. The authors have proposed and also run famous algorithms such as PSO, LBP, LDA, PCA, SVM, and GSM. In their approach, the detection rate has been increased by decreasing the number of features. Another paper has fully described an agent-based IDS approach [48]. This approach has been used for several factors as well as classification to detect intrusion. The authors used the mobile agents to detect intrusion with below elements:

- Collector agent,
- Misuse detection,
- Anomaly detection based on SVM classification.

From the point of view of reading and criticizing their writing, it is obvious that their proposed system has fewer parameters to describe the attacks [47]. This research design could have a more useful diagnosis by creating more complex diagnostic parameters as well as using statistical anomaly detection and creating attack signatures. It is also clear that the defects have been obtained and the solution for the article under discussion at the end of this research project has been evaluated and simulated. Another important research on the IDS era has been proposed based on GA K-Means [49]. In this research, the false positive (FP) rate has been decreased and a high detection rate has been obtained. It should consider that this approach was suitable for dynamic topology changes. Clearly, this measure has been considered a critical and also non-functional requirement in IDS security architecture design. This approach has been able to detect new attacks without pattern and also allows intrusion and traffic analysis.

In a hierarchical model [50], researchers have described an IDS for blackhole attack detection in WSN based on simulation on NS2 software. In this approach, the sensor node and base station (BS) have exchanged control packets. Each control packet has consisted of a node identifier and the number of packages that have been sent to the cluster headers (CH). Obviously, the BS has been monitored to detect a black hole attack. The solution presented in [50] consumes less energy to detect intrusion. As a critique of this article, it can be acknowledged that although the effects of the attacks have been significantly reduced, there has been no guarantee that other blackhole attacks will be identified in their security architecture plan. This means that the researcher doesn't fully consider the non-functional and functional requirements in the requirement engineering and security architecture phase.

On the other side, an optimized IDS has been proposed for Sybil detection [51]. Firstly, an approach focused on sending data packet query confirmation has been

implemented. It should consider that the CH has been saved from the table. This tale is used to store the identity and location of other nodes. This process is somewhat similar to the address resolution protocol (ARP). Secondly, all of the legal nodes have responded to the eclipse with their true identities and real coordinates. This has been where the Sybil node was detected. The results of their research indicated that in the mentioned system, the destructive node and specifically the Sybil attack can be accurately detected and the energy efficiency has been improved.

Another promising solution was proposed to detect wormhole attacks and flooding by simulation in NS2 [52]. In this design, the abnormal behavior of the nodes has been monitored by the energy prediction algorithm. The attack can be assessed on a scale of both real and predictable scenarios. Although the plan's approach minimizes energy consumption, the plan only detects wormhole attacks and flooding.

By reviewing on further research, a Man in the middle IDS (MITM-IDS) has been proposed to isolate attacks and reconfigure attack nodes [44]. Their simulation results show 89.147% efficiency in detecting MITM attacks. In this plan, a penetration detection system based on deep learning techniques has been introduced in order to deal with a popular attack. The strength of this model has been the rapid detection of malicious behavior due to the less complexity. Obviously, ARP spoofing and poisoning can be considered similar to the attack of inserting a malicious node in the sensor network and attacking the non-functional requirements for confidentiality, authentication, and availability, which are the most important security system design parameters such as intrusion detection.

The implementation and application of this system have led to a wide range of solutions in sensor network security. Of course, the limitations of the WSN should not be forgotten in this regard. Authors in [53] have identified various vulnerabilities and security issues across the sensor network. In this research structure, unique classes have been discussed as follows:

- Inner work style.
- Interrelated convention stack.
- Organize provisioning, oversight and transmitting issues.

In the mentioned plan, the calculations of the proposed conventions and oversight have been collected and evaluated. The current issues in the field of research and establishment of IDS in IoT and WSN have been expressed and a qualitative evaluation of approaches has been done [54].

A complete subjective classification of the research that has been done so far is depicted in Fig. 7. To make the best decision to choose the right system more than 70 papers have been studied from 2006 until 2022.

The Fig. 7 has been represented a complete classification of the research that has been conducted on the establishment of IDS.

It has been worth noting that KilerBee has a framework for exploiting ZigBee vulnerabilities. Numerous researchers have examined the challenges as well as future paths ahead for IDS in WSN [105], [115]-[120].

In [122], authors' findings on IoT IDS retrieval and attacks in different layers have been summarized in Fig. 8. Although the explanations of the researchers of the above article have been extensive, structurally, the IDS architecture, as well as the engineering of IDS requirements, have not been addressed. So, if corrected with appropriate words, the same non-functional and functional requirements have been addressed. Mentioning this point will make their research more productive, adding that functional requirements have a factor in the proper functioning of the system, and non-functional requirements, which have no tangible, will meet the unknowing needs of the user. This is the main part of the requirement engineering process and also a critical point for software architecture.

Researchers have proposed a fuzzy logic-based approach to prevent the intrusion on WSN utilizing WSN-DS dataset [123]. Their system has 3 phases:

- Feature extraction,
- Membership Computation,
- Apply fuzzy rules.

On the other side, they utilized of 3 following colors in their simulation.

- Orange: Probability of node destruction,
- Green: the node is secure in the network,
- Red: Destructive node.

Their Proposed FZMAI approach have been consist of some primary parameters as follows:

- Packet transmission rate to the base station,
- Energy Consumption,
- Signal Strength,
- Packet Delivery ratio (PDR),
- Received packet.

Their results have shown a 98.29% improvement in accuracy assessment.

Compared to other fuzzy methods, their approach has been more efficient than others. The advantage of their proposed system is that the malicious node has been prevented from entering the system and thus intrusion has been prevented. Their FZMAI model has been presented in Fig. 9.

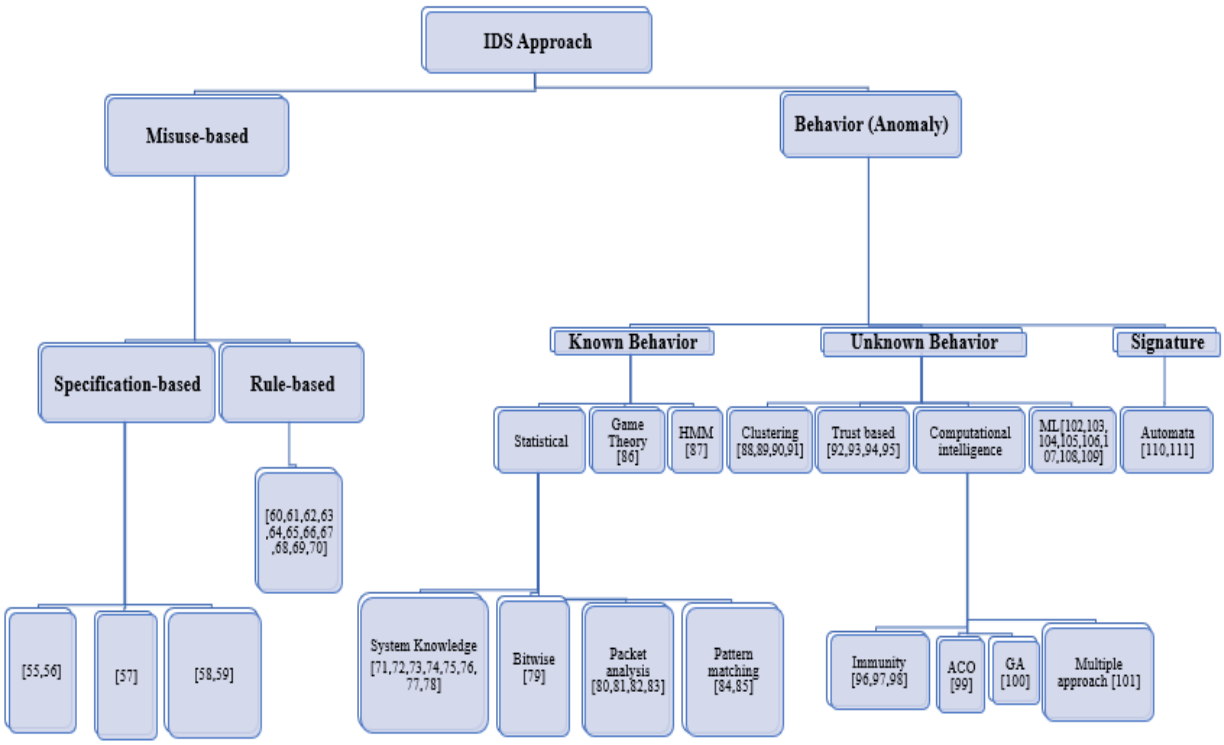


Fig. 7: The IDS classification and related researches.

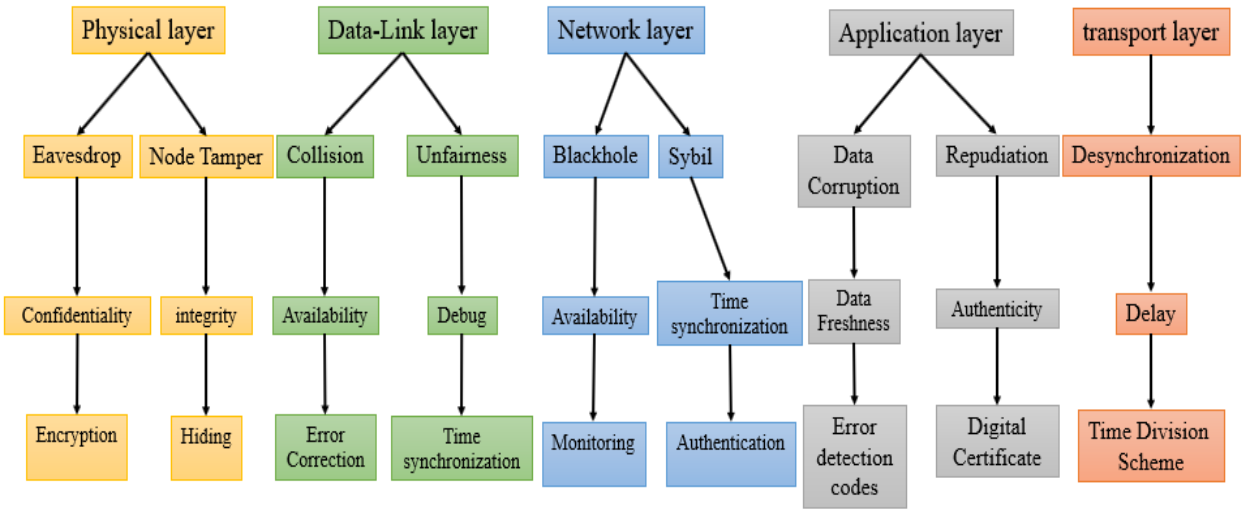


Fig. 8: Different threats in different network layers.

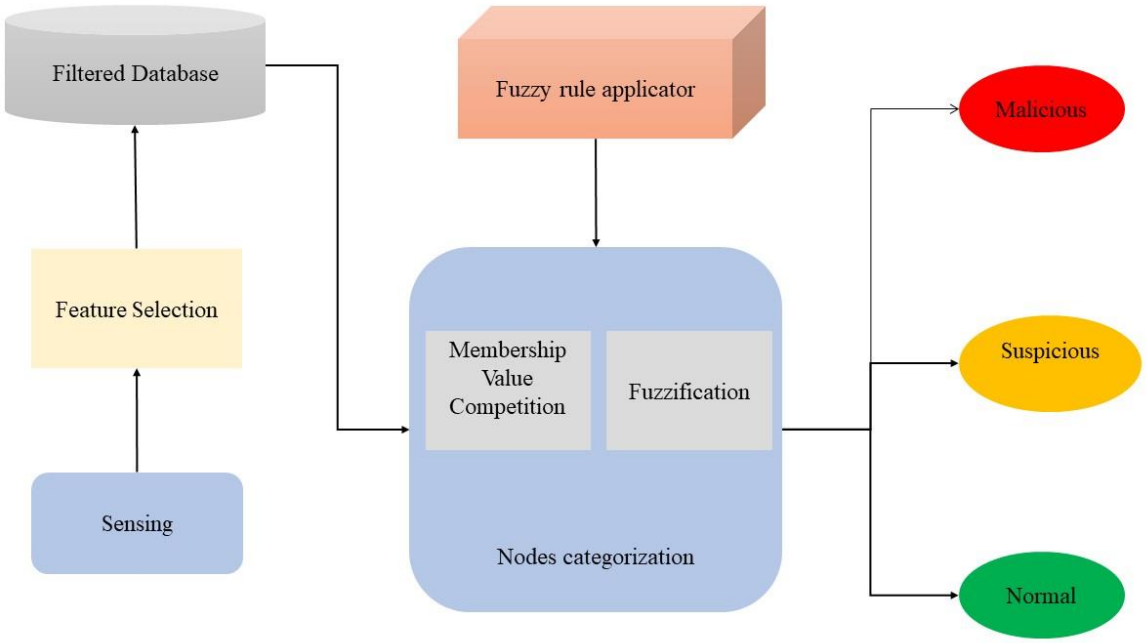


Fig. 9: The FZMAI model [123].

As a critique of their research, it should be noted that their system has not provided transparency in assessing the throughput, latency, and congestion rates in the network. Obviously, in the calculations of the membership function and fuzzy logic, the rate of FP and TP has not been given. According to the proposed algorithm, they have obtained a high FP rate due to the convergence of nodes and data volume.

In [124], researchers have proposed a mechanism for designing sink hole attack detection in the context of hierarchical networks. According to them, HWSN has been the first step in detecting 3 following attacks:

- Sink hole message modification node (SDP),
- Sink hole message delay node (SDL),
- Sink hole message dropping node (SDP).

In their approach, HWSN has been divided into several unconnected clusters, each cluster having a powerful final sensor called a cluster-head, which has been responsible for detecting malicious activity in the infrastructure of the cluster. The simulation and the results obtained from the NS2 simulator show the fact that the detection rate of this approach has been 95% and the FP rate has been 1.25%.

In [125], researchers have proposed an advanced model of IDS based on KNN utilizing the AOA optimization algorithm in WSN. This process has led to an improvement in the face of DoS attacks. to increase the accuracy of this model, a parallel strategy has been used to strengthen the relationship between the population, and also a Levy Flight strategy has been used to adjust the optimization values. The PL-AOA algorithm performs well in benchmark function testing and effectively ensures the improvement of KNN classification operations. The aforementioned model achieved 97% accuracy and almost 10% improvement over the original KNN during the DOS attack. In [126], researchers have studied the solution of secure node detection based on ANN in WSN. Their results have indicated that the optimized solution based on the biological neural network strengthens the diagnosis in WSN. On the other side, Insecure nodes negatively have affected network performance and will naturally interfere with system behavior. Regression analysis for both methods has detected changes when all nodes have been safe and also in insecure status. Diagnosis based on packet delivery ratio and energy consumption can be effectively implemented in the ANN.

Review in literature shows that more than 30% of articles have been based on detecting routing attacks. These attacks have included Blackhole, wormhole, Sybil, Sinkhole, and Selective forwarding. Table 3 has been shown the IDS that have been embedded for 15 unique scenarios. The sensor network has been less abstract than the IoT, but always has computational limitations on processing and energy consumption. On the other hand, reasoning about these systems has been much simpler because of their homogeneity and the possibility of behavioral analysis. It has been obvious that the WSN has been called an IoT system sub-category. To make the best decision to choose the right system more than 70 papers have been studied from 2006 until 2022. The papers have

been collected based on how the attack structure has been positioned and the solutions used in Table 3. Table 3 through Table 8 and also Fig. 10 through Fig. 14 are dedicated for easier reference as well as classification and comparison of different areas and a comprehensive collection of previous research comparison based on prominent parameters.

Table 3: Frequency of research that has been conducted about IDS placement in WSN and IoT

Detection method	IDS placement		
	NIDS	HIDS	CIDS
Rule based	[65], [112], [113]	--	[60], [61]
Signature-based	[64]	--	[62]
Anomaly-based	[113], [84], [75], [104], [102], [103], [67]	[51], [79]	[108], [114], [109], [82], [59]
Statistical-based	[81], [73], [74]	[76], [85], [87]	[69], [94]
Stateful-based	[111]	--	--
Clustering-based	[88], [95], [89]	--	--
CI-based	[96]	[97]	[99], [101], [100], [98]
Specification-based	[55], [57]	--	[56], [83]
Trust-based	[92]-[94]	--	--
Autonomous-based	[110], [111]	--	--
Game Theory-based	[86]	--	--
Misuse-based	[66]	--	[107]

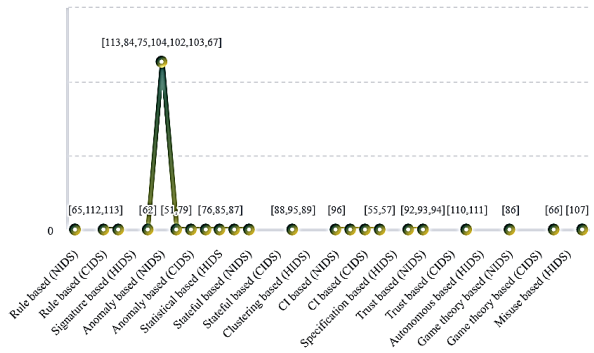


Fig. 10: The scope of research conducted in the field of deployment and type of IDS in WSN and IoT in previous years.

Table 4: Various security scenarios and IDS detection in WSN and IoT

Attack structure	IDS placement		
	NIDS	HIDS	CIDS
Scanning	[66]	--	[100]
Web exploit	--	--	[100]
Routing attack	[93]-[95], [88], [113], [55], [74], [73], [64], [67], [83]	--	[99], [58], [107], [56], [109], [59]
Rank	--	--	--
Information leakage	[66]	--	[98], [99], [101]
Replay	[57], [110]	[76]	[101], [98]
Spoofing	[57], [110]	[76]	[101], [98]
Packet Drop	[57], [110]	--	[101], [108], [98]
Flooding	[65], [75], [86], [81]	[72], [85]	[77], [101], [60], [61], [71], [78], [98]
Worm	--	[79]	--
Injection	--	[79]	--
Anomaly Behavior	[96], [103], [104]	--	[82]

Table 5: Extent of research on the spatial environmental conditions of IDS in different networks

Network Type	IDS placement		
	NIDS	HIDS	CIDS
WSN	[78], [88], [75], [104], [90], [92]	[87]	[99], [58], [107], [101], [59]
IoT	[63], [111], [97], [110], [86]	[79], [85], [97]	[108], [114], [69], [100], [71], [70]
IPV6	[73]	--	[62]
Smart grid	--	--	[92]
Smart City	[34]	--	--
Smart home	[96]	--	--
ZigBee	[102]	--	--
ICS	[84]	--	--
Bluetooth	[65], [66]	--	--
Relay Comm	--	[76]	--
RPL	[55], [74], [103], [67]	--	[49], [69], [75]
6Low pan	[81], [93]	[72]	[44], [45]
Clustered	[95]	--	--
Healthcare	[94]	--	--
BACnet	[57]	--	--

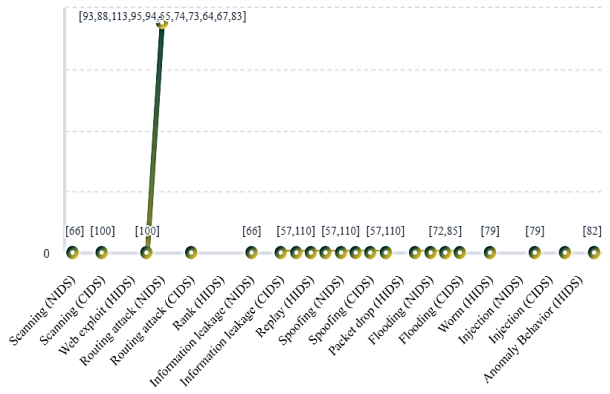


Fig. 11: The scope of research conducted in the field of Various attacks scenarios and IDS detection in WSN & IoT.

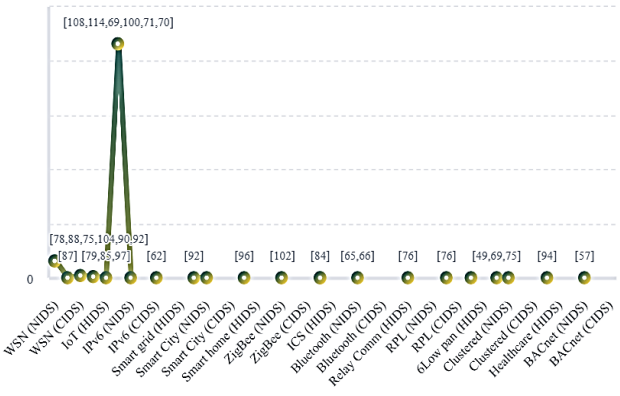


Fig. 12: The extent of the research done around the IDS in the action space and the infrastructure of various networks.

Table 6: IDS simulator in WSN and IoT

IDS Simulator	IDS placement		
	NIDS	HIDS	CIDS
NS2	[75]	--	--
Omnet++	[92]	--	
OpenSim	[96]	--	--
Tossim	[64], [86]	--	--
Avrora	[102]	--	--
R	[68]	--	--
Matlab	[88], [94], [98]	[87]	[108], [106], [91], [100]
Qualent	--	[72]	--
Contiki/Cooja	[93], [74], [73], [67]	--	[106], [56], [109], [82]
Not Specified	[81]	--	--

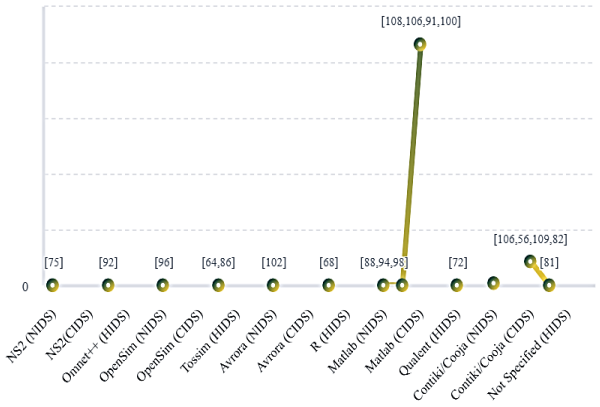


Fig. 13: The extent of the research conducted around the utilizing of various IDS simulators in WSN and IoT.

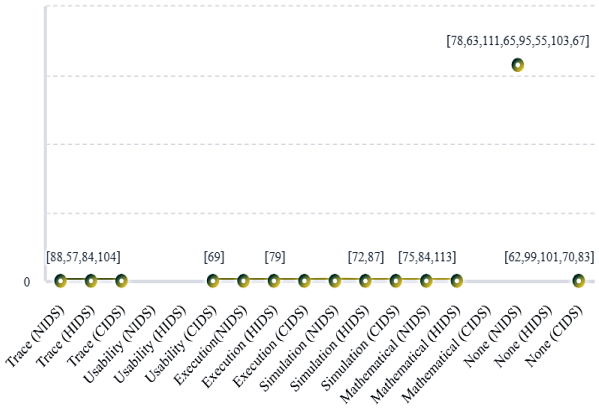


Fig. 14: The extent of previous research in the field of IoT assessment and utilizing of different types of IDS.

Table 7: Evaluation on IoT & IDSs

Method	IDS placement		
	NIDS	HIDS	CIDS
Trace	[88], [57], [84], [104]	[85]	[98], [107]
Usability	--	--	[69]
Execution	[57], [66], [110]	[79]	[58], [69], [100], [60], [61], [71]
Simulation	[96], [81], [88], [64], [86], [102], [92]	[72], [87]	[93], [68], [75], [94], [74], [73], [67]
Mathematical	[75], [84], [113]	[76], [85], [113]	--
None	[78], [63], [111], [65], [95], [55], [103], [67]	--	[62], [99], [101], [70], [83]

Also, Table 8 provides a summary and useful information on the NIDS in WSN and IOT that researchers have done for the last 10 years.

Table 8: Summary of research that have been conducted in NIDS on WSN and IOT

Paper	Detection Approach	IDS Placement	Envolve	Evaluation	Detection
[80]	Anomaly-based	Distributed	WSN, AD-HOC, IoT	О, Б, ж, £	$B^1, B^2, B^3, B^4, B^5, B^6$
[96]	Immune-based	Centralized	Smart-home	о	$Internal^2$
[78]	Anomaly-based	Distributed	WSN	χ	--
[81]	Statistical-based	Centralized	6Low PAN	о	DDoS
[63]	Rule-based	Centralized	Test Bed	χ	--
[88]	Cluster-based	--	WSN	0, Б	Sybil
[55]	Specification-based	Distributed	RPL	χ	Topology
[89]	Clustering-based	--	IoT	£	--
[102]	Anomaly-based	Distributed	ZigBee	о	Killer Bee
[83]	Anomaly-based	--	ICS	Б, £	--
[64]	Signature-based	Centralized	WSN	о	$Routing^3$
[86]	Game-Theory	Centralized	IoT	о	DoS
[65]	Rule-based	Centralized	Bluetooth	χ	B^4
[92]	Trust-based	Distributed	WSN	о	B^5
[66]	Misused-based	Centralized	Bluetooth	ж	B^6
[103]	Anomaly-based	Distributed	RPL	χ	$Internal^2$
[104]	Anomaly-based	Distributed	WSN	Б	$Internal^2$
[67]	Hybrid-based	Distributed	RPL	о	$Routing^3$
[73]	Statistical-based	Distributed	IPv6	о	$Routing^3$
[74]	Statistical-based	Distributed	RPL	о	$Routing^3$
[93]	Trust-based	Distributed	6Low PAN	о	Sink hole
[110]	Automata-based	Centralized	IoT	ж	DDoS
[105]	Hybrid-based	Centralized	Smart City	о	Routing
[111]	Automata-based	Centralized	HTTP	χ	$Protocol^1$
[94]	Trust-based	Distributed	Healthcare	о	$Routing^3$
[75]	Anomaly-based	Centralized	WSN	£, о	Energy-DoS
[95]	Cluster-based	Centralized	Clustered	χ	$Routing^3$
[57]	Specification-based	Centralized	BAC Net	Б, ж	$Protocol^1$
[114], [115]	Fuzzy logic-based	Distributed	WSN	--	
[111], [115]	Agent approach	Distributed	WSN	--	
[105], [115]	Rule-based	Distributed	WSN	--	
[112], [115]	ANN	Centralized	WSN	--	
[101], [115]	CVM	Centralized	WSN	WSN-DS	
[65], [115]	Fuzzy logic-based	Centralized	WSN	KDD Cup99	
[83], [115]	ANN	Distributed	WSN	--	
[75], [115]	Random forest	Centralized	WSN	KDD Cup99	
[100], [115]	K-means & SVM	Distributed	WSN	KDD Cup99	
[115], [116]	SVM	Distributed	WSN	KDD Cup99	
[115], [117]	Trust-based	Distributed	WSN	KDD Cup99	
[95], [115]	Trust-based	Distributed	WSN	--	
[115], [118]	Trust-based	Distributed	WSN	--	

Legend

Activity	Abbreviations	Operation
Detection range	B	Custom Attack List
	1	Spoofing, MITM, Drop Packet, Replay attack
	2	Unusual activity by a component in the framework
	3	Routing Attacks such as Worm, Sink & Black Holes.
	4	Resource Drain DoS, Spoofing
	5	Collision attacks, Selective forward & hello Flooding
	6	Reconnaissance, DoS, theft and leakage attacks
Evaluation style	ж	Execution evaluation performed
	Б	Trace evaluation performed
	£	Mathematical evaluation performed
	о	Simulation performed
	χ	Without any Evaluation

Attacks Against WSN (Passive & Active Attacks)

The passive attack has been limited to sniffing the exchanged traffic. This type of attack has been easier to realize and difficult to detect because they do not involve any alteration of the data. Since the attacker does not make any modification to the exchanged information. The intention of the attacker can be the find out the confidential information or the knowledge of the significant nodes in the network (cluster head node), by analyzing routing information, to prepare an active attack.

In active attacks, an attacker tries to remove or modify the messages transmitted on the network. The attacker can also inject his traffic or replay old messages to disturb the operation of the network or to cause a DOS. In WSN, among the most known active attacks, it can quote Tampering, Blackhole, Selective forwarding, Sybil, Hello flooding, Jamming, Blackmail, Exhaustion, Wormhole, and identity replication attacks [16].

A wormhole attack has been a routing scenario that will happen on the network layer. In this scenario, the attackers have been required to import at least 2 malicious nodes. These two nodes have been classified via a low latency link directly. This Direct link called tunnel caused the conflict and also aberration in routing protocols. A malicious node takes the packets in a part of the network and then will forward them via its malicious tunnel. The wormhole scenario has been running when the other node has been in the discovery phase. Note that, in this scenario, there has not been any negotiation between the sensors [15], [16]. In the Blackhole scenario, attackers have been required to import at least a malicious node into the network. On the other mean, this node will modify the routing table for malicious goals, once it takes the incoming traffic then there hasn't been any retransmission for sensed data. In the Sybil scenario, attackers can use the identities of the others nodes to take part in the distributed algorithm such as the election [18].

The traditional routing protocols faced many problems due to dynamic behavior and resource constraints. These Attacks can occur when the malicious node present in the network has been intended to attack directly the data traffic and intentionally drops, delay or alter the data traffic passing through it.

Blackhole Attack has been a very dangerous active attack on the MANETs and WSN. It has been formed during the week routing infrastructure when a malicious node joins the network this problem arises. Detection systems for ad hoc networks have been extremely difficult due to the lack of a central controller, bandwidth limitations, and dynamic topology in mobile ad hoc networks. Routing protocols have been a great guide to authors in evaluating connection quality and estimating destination info. In this paper we simulate Blackhole and Wormhole attacks in a cluster-based network with NS-2 simulator and also a safe cluster-based network for comparison phase, then we have been exploited some primary requirements such as throughput, end to end delay packet delivery ratio, normalized routing load, receive the packet with AWK language and then we have been plotted them on the following figures. To validate those results, we have been running them on Debian, Ubuntu, and Kali Linux operation systems, separately. For the network basic parameters and presents a comparison ability with another approach, we have been used other researcher measures that have been shown in Table 9, [14]-[18].

Table 9: Simulation parameters in a WSN

Parameters	Settings
Number of nodes	21
Network area	500*500 (m^2)
Routing protocol	AODV
Maximum packet in IFQ	50 (**ms**)
End time of simulation	10.0

In the following, we have been traced the primary parameters of simulation based on AGT's level trace.

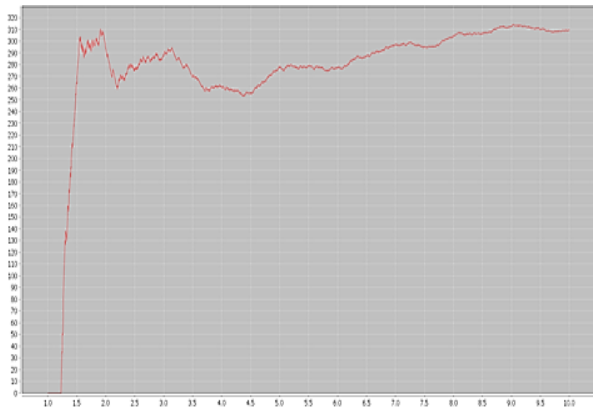


Fig. 15: Throughput in clear scenario.

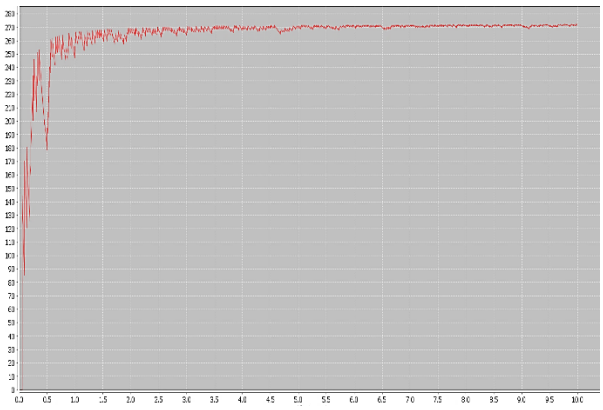


Fig. 16: Throughput in Wormhole scenario.

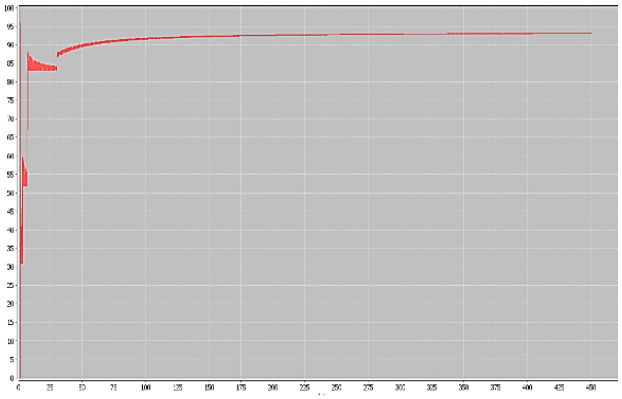


Fig. 17: Throughput in clear scenario.

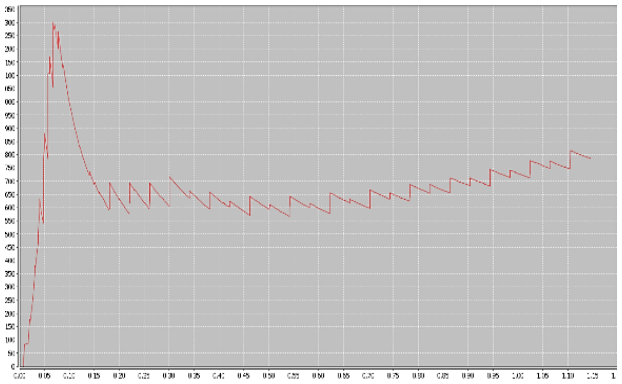


Fig. 18: Throughput in node scan scenario.

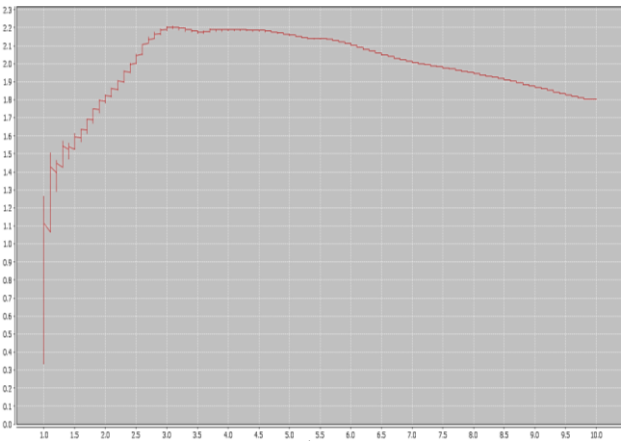


Fig. 19: End to end delay in the clear scenario.

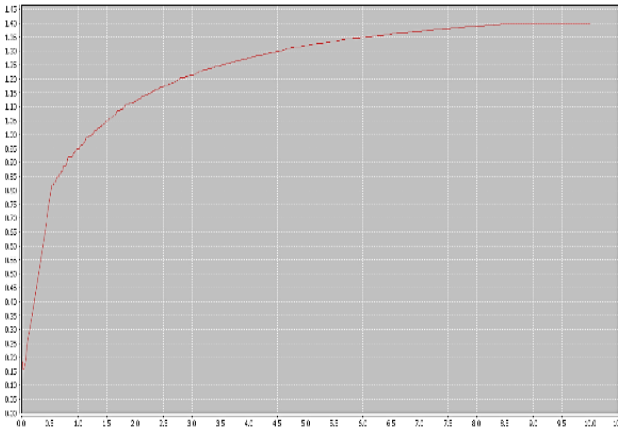


Fig. 20: End to end delay in Wormhole scenario.

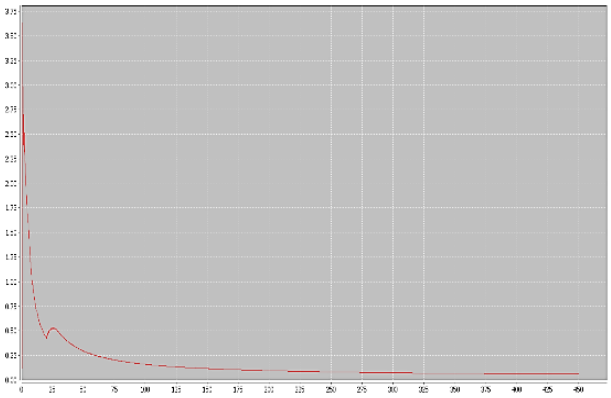


Fig. 21: End to end delay in the clear scenario.

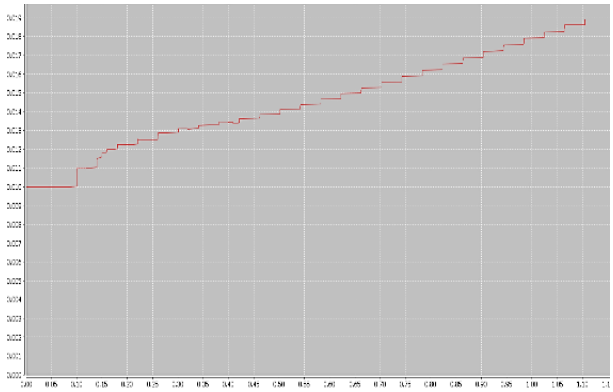


Fig. 22: End to end delay in nodes scan scenario.

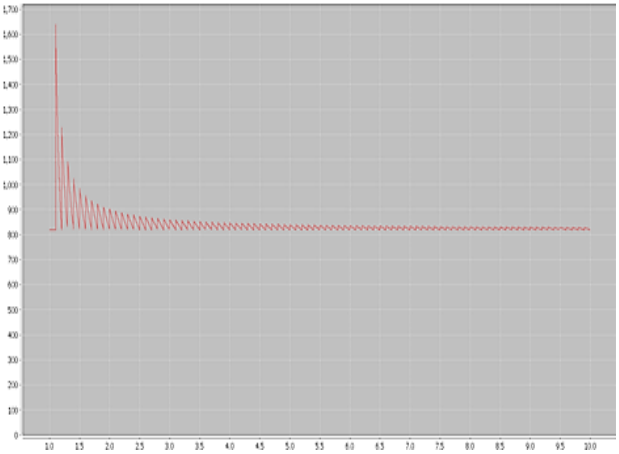


Fig. 23: Sent packet in clear scenario.

Fig. 24: Sent packet in Wormhole scenario.

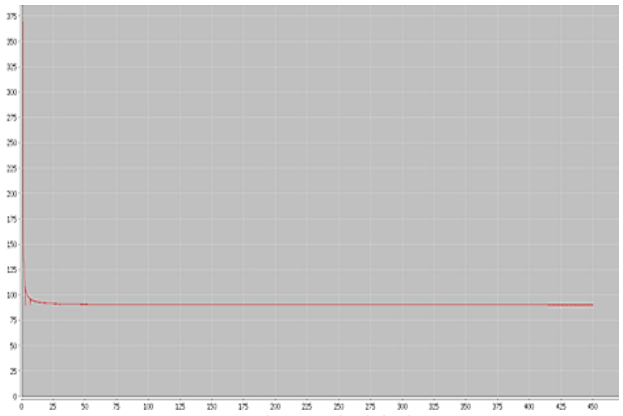


Fig. 25: Sent packet in Black hole scenario.

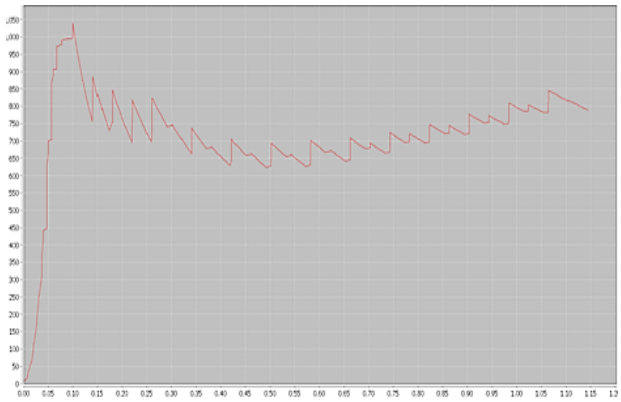


Fig. 26: Sent packet in nodes scan scenario.

Table 10: Evaluation of the simulation scenarios

Network scenario	AVG of THR	RPKT	PDR	NRL	AVG of DLY
Clear scenario	2174.19	5399	539900	0.61	8996.44
Wormhole scenario	1354.91	6489	648900	0.05	6667
Black hole scenario	2604.68	65504	6550400	0.03	308981
Malicious node scenario	680.42	32745	409312	0.03	66009.1
Scan scenario	2875.85	1375	1556.82	0.03	18.8681

According to Table 10, the average throughput in a clear scenario has been equaled to 2174.19 kbps. Figs. 15-18 described that the throughput ratio has increased in total time. On the other hand, in the wormhole attack, it showed that after 1 unit of the simulation time, we had stable traffic in the time. In the scanning scenario, the throughput has been increased. So according to the clear scenario and malicious scenarios we can predict the attacks. Table 10 and Figs. 19-22, described that the average clear scenario has been about 8996.44 kbps. After half of the simulation time, the end-to-end delay ratio has been decreased. In the wormhole scenario, this rate has been increased and in the Black hole, it has been strongly decreased. We have the same scenario with sent packet parameters. Now, based on the information of Table 10 and also Figs. 23-26, we can detect the attacks on base station units with the Splunk or other SIEM monitoring management's software.

State-of-the-Art Intrusion Detection Techniques for New Technologies

Nowadays, researchers encounter buzzwords: deep learning [131], Internet of Things (IoT) [132], [133], Safeguard in WSN, cybersecurity attacks [134], security tenets [134] etc. Unfortunately, besides developing the aforementioned technologies, cybersecurity attacks have been extended. So, different safeguard techniques were recently introduced in the literature to obviate the troubles.

In this vein, Mohamed Saied et al. [135] recently presented a survey study on how to enhance intrusion detection systems in the IoT domain by incorporating artificial intelligence approaches. To obviate problems such as blackhole and sinkhole attacks in healthcare wireless sensor networks, an efficient IDS system was recently presented by Webber et al. in 2023 [136]. This method tries to guarantee the security of sensitive data packets in the early stages of the healthcare system. One of the recently applicable technologies is Internet of Vehicles which is used in smart transportation. It is also subjected to different wireless attacks. To mitigate the problem in such networks, an intrusion detection framework was proposed by Selim Korium et al. in 2024 [137].

The authors utilized machine learning techniques such as feature selection methods to detect security attacks and have suitable countermeasures. Tseng and Change in 2023 [131] proposed a deep learning neural network-based ensemble binary detection model for recognizing the multi-level class intrusion detection attacks. A defense mechanism for intrusion detection and prevention models was proposed by incorporating the ensemble learning model [138]. A computational intelligence-based on particle swarm optimization (PSO) algorithm was recently proposed in the literature to solve intrusion detection problems in IoT/5G/Wi-Fi wireless scenarios [139].

Table 11 is dedicated to comparing state-of-the-art technologies that utilize intrusion detection methods to solve the intrusion challenges. This table compares literature based on their innovations, subject challenges, how to solve them, utilized methods, and future direction. It paves the way for further processing to obviate existing shortcomings.

Table 11: Comparison of literature in terms of prominent factors

Author(s) /Ref. /Year	Innovation	Subject Challenge	Used method/solution	Future Direction
Tseng & Change [131], 2023;	It uses data enhancement and ensemble model to improve detection accuracy.	Limited number of attack detection in distributed wireless systems	It utilizes machine learning techniques such as ensemble binary model and convert it to multi-class detection model.	Presenting a comprehensive ensemble model with the most accuracy prediction model to detect different kinds of attacks
Jayanayudu & Sudhir [133], 2023;	It engages meta-heuristic algorithm to make a balance between malicious detection and electricity consumption.	Malicious detection	It makes a safeguard against wireless attacks in IoT WSN environment.	Presenting an efficient multi-objective algorithm which make tread-off between conflicting objectives
Webber et al. [136], 2023;	Presenting an innovated intrusion detection model with high accuracy and minimum data loss	Keeping network quality of service (QoS) level along with malicious detection in early stages	It uses Minkowski K-means clustering method to define meaningful similarity.	Presenting an ensemble model which engages the advantages of existing classifiers to decrease detection error
Korium et al. [137], 2023;	It present low-execution time malicious detectors by utilizing ML methods.	Malicious detection with low time complexity	It uses different datasets and use feature selection methods to detect malic behaviors as soon as possible.	Presenting a comprehensive detection model which can efficiently works of complicated datasets
Ntizikira et al. [138], 2024;	It present a detection system framework which detects malic behaviors in different IoT applications' attacks.	Intrusion detection problem in diverse IoT applications	It utilizes different heuristic to improve detection accuracy.	Presenting a light weight deep learning algorithm which makes a trade-off between time complexity and classification accuracy
Sivagami et al. [139], 2023;	It present network intrusion detection system (NIDS) by utilizing PSO algorithm.	Identifying malicious activity	It utilizes hybridizing ML algorithm and PSO optimizer to improve presented system's performance.	Presenting an ensemble of deep learning algorithms to improve detection accuracy

Results and Discussion

Since the security tenets are a vital non-functional requirement for mission-critical applications in the WSN context, a review of the intrusion detection system in such vulnerable networks is necessary. The reason why the current survey study was conducted is to investigate solutions presented in the literature for finding existing challenges and potential solutions for further processing. The subjective classification has been done and the literatures have been analyzed based on the proposed classifications. The existing challenges and prominent considerations were proposed for improving the current schemes in the WSN context. On the other hand, since these kinds of networks have limited resources, especially in power provisions, the energy-efficient IDS systems are favorable. To this end, ML techniques are beneficial. As a result of the low energy consumption requirements in the sensor network, the use of a hierarchical model will be useful. This means that the network must be divided into clusters, and each of them will have a cluster head. Accordingly, energy consumption will be minimized by avoiding the requirement for all nodes to send data to the base station. Also, intrusion detection algorithms with high energy consumption have been implemented only on the cluster head leading to energy storage and ultimately increasing the life of the network. Intrusion detection energy consumption has been an important point from a security point of view. The WSN consumes a lot of energy by sensing events, processing the information that has been collected, and transmitting the resulting data. Therefore, an IDS has been required to use as little energy as possible to store the energy necessary for operation in the WSN. This research also paves the

way of interested researchers to find existing challenges and shortcomings for further processing.

Author Contributions

Dr. Mirsaeid Hosseini Shirvani was the supervisor of the current research plan. He sketched the research framework and the roadmap. Also, he analyzed the results and tabulated the outcome derived from excerpted literatures. In this line, Amir Akbarifar searched in authentic journals to gather all relevant papers. In addition to, he prepared the blueprint of the research plan. He and his supervisor cooperatively summed up the work.

Acknowledgment

This work is completely self-supporting, thereby no any financial agency's role is available.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviation

WSN	Wireless Sensor Network
AVG	Average
DLY	Delay
THR	Throughput
PKT	Packet
PDR	Packet Delivery Ratio
RPKT	Received Packets
NRL	Normalized Routing Load

References

[1] F. Liu, J. Xu, L. Zhang et al., "DESNN algorithm for communication network intrusion detection," Wireless Pers. Commun., 126: 1705-1720, 2022.

[2] F. Junior, N. Silva, A. Guelfi, S. Takeo Kofuji, "IoT6Sec: reliability model for Internet of Things security focused on anomalous measurements identification with energy analysis," Wireless Netw., 25: 1533-1556, 2019.

[3] T. Do-Dac, K. Ho-Van, "Energy harvesting cognitive radio networks: security analysis for Nakagami-m fading," Wireless Netw., 27: 1561-1572, 2021.

[4] U. Ghugar, J. Pradhan, "NL-IDS: Trust based intrusion detection system for network layer in wireless sensor networks," in Proc. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC): 512-516, 2018.

[5] B. J. Santhosh Kumar, S. Sinha, "An intrusion detection and prevention system against DOS attacks for internet-integrated WSN," in Proc. 2022 7th International Conference on Communication and Electronics Systems (ICCES): 793-797, 2022.

[6] A. P. da Silva, M. Martins, B. Rocha, A. Loureiro, L. Ruiz, H. C. Wong, "Decentralized intrusion detection in wireless sensor networks," in Proc. 1st ACM International Workshop on Quality of Service and Security in Wireless and Mobile Networks (Q2SWinet '05), ACM Press: 16-23, 2005.

[7] M. Meisel, V. Pappas, L. Zhang, "A taxonomy of biologically inspired research in computer networking," Comput. Network, 54(6): 901-916, 2010.

[8] K. Chaitanya, A. Ghosh, "Analysis of Denial-of-Service attacks on Wireless Sensor networks using simulation, IT Security for the Next Generation - European Cup 2011: 1-13, 2010.

[9] R. C. Chen, C. F. Hsieh, Y. F. Huang, "A new method for intrusion detection on hierarchical wireless sensor networks", in Proc. ACM ICUIMC-09, 2009.

[10] S. Otoum, B. Kantarci, H. T. Mouftah "A novel ensemble method for advanced intrusion detection in wireless sensor networks," in Proc. 2020 IEEE International Conference on Communications (ICC 2020): 1-6, 2020.

[11] M. Aloqaily, S. Otoum, I. Al Ridhawi, Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities," Ad Hoc Networks, 90: 101842, 2019.

[12] S. Otoum, B. Kantarci, H. T. Mouftah, "On the feasibility of deep learning in sensor network intrusion detection, "IEEE Networking Lett., 1(2): 68-71, 2019.

[13] I. Butun, , S. D. Morgera, R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," IEEE Commun. Surv. Tutorials, 16(1): 266-282, 2014.

[14] M. Hosseini Shirvani, A. Akbarifar, "Anomaly-based detection of blackhole attacks in WSN and MANET utilizing quantum-metaheuristic algorithms," J.Commun. Eng., 9(1): 77-92, 2020.

[15] Y. Maleh, A. Ezzati, "A review of security attacks and intrusion detection schemes in wireless sensor networks," arXiv preprint arXiv:1401.1982, 2014.

[16] S. Binitha, S. S. Sathya, "A survey of bio inspired optimization algorithms," Int. J. Soft Comput. Eng., 2(2): 137-151, 2012.

[17] R. Fu et al., "Biologically inspired anomaly detection for hierarchical wireless sensor networks," J. Networks 7(8): 1214-1219, 2012.

[18] M. Hosseini Shirvani, A. Akbarifar, "A comparative study on anonymizing networks: TOR, I2P, and riffle networks comparison, J. Electr. Comput. Eng. Innovations (JECEI), 10(2): 259-272, 2022.

[19] E. Baraneetharan, "Role of machine learning algorithms intrusion detection in WSNs: a survey," J. Inf. Technol., 2(03): 161-173, 2020.

[20] I. Butun, S. D. Morgera, R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," IEEE Commun. Surv. Tutorials, 16(1): 266-282, 2013.

[21] X. Tan, S. Su, Z. Huang, X. Guo, Z. Zuo, X. Sun, L. Li, "Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm," Sensors, 19(1): 203, 2019.

[22] C. C. Su, K. M. Chang, Y. H. Kuo, M. F. Horng, "The new intrusion prevention and detection approaches for clustering-based sensor networks," in Proc. IEEE Wireless Communications and Networking Conference, 2005.

[23] A. A. Strikos, "A full approach for intrusion detection in wireless sensor networks," School of Information and Communication Technology, KTH, Stockholm, Sweden 16453, 2007.

[24] Y. Otoum, D. Liu, , A. Nayak, "DL-IDS: a deep learning–based intrusion detection framework for securing IoT," Trans. Emerg. Telecommun. Technol., 33(3): e3803, 2022.

[25] N. Moustafa, B.Turnbull, K. K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features

for protecting network traffic of internet of things," IEEE Internet Things J., 6(3): 4815-4830, 2018.

[26] M. Hosseini Shirvani, A. Akbarifar, A. Nazokkar, "Reliability non-functional requirement evaluation in mission-critical systems with an architectural strategy for future systems," Int. J. Comput. Appl., 46(4): 227-251, 2024.

[27] A. Aldweesh, A. Derhab, A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," Knowledge-Based Syst., 189: 105124 2020.

[28] G. Marín, P. Casas, G. Capdehourat, "Rawpower: Deep learning based anomaly detection from raw network traffic measurements," in Proc. the ACM SIGCOMM 2018 Conference on Posters and Demos: 75-77, 2018.

[29] R. Sharma, V. A. Athavale, "Survey of intrusion detection techniques and architectures in wireless sensor networks," Int. J. Adv. Networking Appl., 10(4): 3925-3937, 2019.

[30] K. Haseeb, I. Ud Din, A. Almogren, N. Islam, "An energy efficient and secure IoT-based WSN framework: An application to smart agriculture," Sensors, 20(7): 2081, 2020.

[31] W. Zhang, D. Han, K. C. Li, F. Massetto, "Wireless sensor network intrusion detection system based on MK-ELM," Soft Comput., 24(16): 12361-12374, 2020.

[32] S. Shen, G. Yue, Q. Cao, F. Yu, "A survey of game theory in wireless sensor networks security," J. Networks, 6(3): 521, 2011.

[33] G. Zhang, T. Wang, G. Wang, A. Liu, W. Jia, "Detection of hidden data attacks combined fog computing and trust evaluation method in sensor-cloud system," Concurrency Comput. Pract. Exper., 33(7): 1-1, 2021.

[34] A. Agah, K. Basu, , S. K. Das, "Preventing DoS attack in sensor networks: a game theoretic approach," in Proc. IEEE International Conference on Communications (ICC 2005), 5: 3218-3222, 2005.

[35] L. Yang, D. Mu, X. Cai, "Preventing dropping packets attack in sensor networks: A game theory approach," Wuhan Univ. J. Nat. Sci., 13(5): 631-635, 2008.

[36] M. A. Khan, S. U. Jan, J. Ahmad., S. S. Jamal, A. A. Shah, N. Pitropakis, W. J. Buchanan, "A Deep learning-based intrusion detection system for mqtt enabled iot," Sensors, 21: 7016, 2021.

[37] S. Rajasegarar, C. Leckie, M. Palaniswami, J. C. Bezdek, "Distributed anomaly detection in wireless sensor networks", in Proc. 10th IEEE Singapore International Conference on Communication systems, 2006.

[38] V. Kelli, V. Argyriou, T. Lagkas, G. Fragulis, E. Grigoriou, , & P. Sarigiannidis, P. IDS for industrial applications: a federated learning approach with active personalization. Sensors, 21(20), 6743 (2021)

[39] M. A Almaiah, "A new scheme for detecting malicious attacks in wireless sensor networks based on blockchain technology," in Artificial Intelligence and Blockchain for Future Cybersecurity Applications: 217-234. Springer, Cham 2021.

[40] M. Khudadad, Z. Huang, "Novel intrusion detection methods for security of wireless sensor network, "J. Fundam. Appl. Sci., 10(2S): 173-189, 2018.

[41] I. Almomani, A. Alromi, "Integrating software engineering processes in the development of efficient intrusion detection systems in wireless sensor networks," Sensors, 20(5): 1375, 2020.

[42] S. Subbiah , K. S. M. Anbananthen, S. Thangaraj, S. Kannan, D. Chelliah, "Intrusion detection technique in wireless sensor network using grid search random forest with Boruta feature selection algorithm," J. Commun. Networks, 24(2): 264-273, 2022.

[43] W. Wang, M. Chatterjee, K. Kwiat, "Coexistence with malicious nodes: A game theoretic approach," in Proc. 2009 International Conference on Game Theory for Networks: 277-286, 2009.

[44] T. Sherasiya, H. Upadhyay, H. B. Patel, "A survey: Intrusion detection system for internet of things," Int. J. Comput. Sci. Eng. (IJCSE), 5(2): 91-98, 2016.

[45] X. Jia, Q. Feng,T. Fan, Q. Lei, "Rfid Technology And Its Applications In Internet Of Things (Iot)," In Proc. IEEE 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet): 1282-1285, 2012.

[46] C. Jun, C. Chi, "Design of complex event-processing IDS in internet of things," in Proc. IEEE 2014 sixth International Conference on Measuring Technology and Mechatronics Automation: 226-229, 2014.

[47] A. Alsadhan, N. Khan, "A proposed optimized and efficient intrusion detection system for wireless sensor network," Int. J. Comput. Inf. Eng., 7(12): 1621-1624, 2013.

[48] Y. El Mourabit, A. Toumanari, A. Bouirden, H. Zougagh, R. Latif, "Intrusion detection system in Wireless Sensor Network based on mobile agent," in Proc. Second World Conference on Complex Systems (WCCS): 248-251, 2014.

[49] G. Sandhya, A. Julian, "Intrusion detection in wireless sensor network using genetic K-means algorithm," in Proc. 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies: 1791-1794, 2014.

[50] S. Athmani, D. E. Boubiche, A. Bilami, "Hierarchical energy efficient intrusion detection system for black hole attacks in WSNs," in Proc. 2013 World Congress on Computer and Information Technology (WCCIT): 1-5, 2013.

[51] A. B. Karuppiah, J. Dalfiah, K. Yuvashri, S. Rajaram, S. Pathan, "A novel energy-efficient sybil node detection algorithm for intrusion detection system in wireless sensor networks," in proc. 2014 3rd International Conference on Eco-friendly Computing and Communication Systems: 95-98, 2014.

[52] D. Harini, N. Balakrishnan, A. P. Renold, "Distributed detection of flooding and gray hole attacks in wireless sensor network," in Proc. 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM): 178-184, 2015.

[53] F. Fazlic, S. A.Hashemi, A. Aletic, A. Abd Almisreb, S. M. Norzeli, N. M. Din, "A survey on security in wireless sensor network," Southeast Eur. J. Soft Comput., 8(1), 2019.

[54] L. Arnaboldi, C. Morisset, "A review of intrusion detection systems and their evaluation in the IoT," arXiv preprint arXiv:2105.08096, 2021.

[55] A. Le, J. Loo, Y. Luo, A. Lasebae, "Specification-based IDS for securing RPL from topology attacks," in Proc. 2011 IFIP Wireless Days (WD): 1-3, 2011.

[56] I. Krontiris, T. Dimitriou F. C. Freiling, "Towards intrusion detection in wireless sensor networks," in Proc. 13th European Wireless Conference, 2007.

[57] H. Esquivel-Vargas, M. Caselli, A. Peter, "Automatic deployment of specification-based intrusion detection in the BACnet protocol," in Proc. the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy: 25-36, 2017.

[58] H. Bostani, M. Sheikhan, "Hybrid of anomaly-based and specification-based ids for internet of things using unsupervised opf based on mapreduce approach," Comput. Commun., 98: 52-71 2017.

[59] Y. Otoum, D. Liu, A. Nayak, "DL-IDS: a deep learning–based intrusion detection framework for securing IoT," Trans. Emerg. Telecommun. Technol., 33(3): 3803, 2022.

[60] E. Ngai, J. Liu, M. Lyu, "On the intruder detection for sinkhole attack in wireless sensor networks," in Proc. 2006 IEEE International Conference on Communications , 2006.

[61] S. S. Doumit, D. P. Agrawal, “Self-organized criticality and stochastic learning based intrusion detection system for wireless sensor networks,” in Proc. IEEE Military Communications Conference (MILCOM’03), 2003.

[62] J. P. Amaral, L. M. Oliveira, J. J. P. C. Rodrigues, G. Han, L. Shu, “Policy and network-based intrusion detection system for IPv6-enabled wireless sensor networks,” in Proc. IEEE International Conference on Communications (ICC): 1796-1801, 2014.

[63] J. M. R. Danda, C. Hota, “Attack identification framework for IoT devices,” in Advances in Intelligent Systems and Computing, 434: 505-513, Information Systems Design and Intelligent Applications, Springer, 2016.

[64] H. Sedjelmaci, S. M. Senouci, "Efficient and lightweight intrusion detection based on nodes' behaviors in wireless sensor networks," in Proc. Global Information Infrastructure Symposium - GIIS 2013: 1-6, 2013.

[65] K. M. J. Haataja, New efficient intrusion detection and prevention system for Bluetooth networks,” in Proc. the 1st International Conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications, 2010.

[66] T. OConnor, D. Reeves, “Bluetooth network-based misuse detection,” in Proc. Annual Computer Security Applications Conference (ACSAC): 377-391, 2008.

[67] L. Wallgren, S. Raza, T. Voigt, “Routing attacks and countermeasures in the RPL-Based internet of things,” Int. J. Distrib. Sens. Netw., 9(8), 2013.

[68] V. Garcia-Font, C. Garrigues, H. Rifà-Pous, “Attack classification schema for smart city WSNs,” Sensors, 17(4):771, 2017.

[69] T. K. Buennemeyer, T. M. Nelson, L. M. Clagett, J. P. Dunning, R. C. Marchany, J. G. Tront, "Mobile device profiling and intrusion detection using smart batteries," in Proc. 41st Annual Hawaii International Conference on System Sciences (HICSS 2008) :296-296, 2008.

[70] I. Onat, A. Miri, “A real-time node-based traffic anomaly detection algorithm for wireless sensor networks,” in Proc. 2005 Systems Communications, 2005.

[71] D. Midi, A. Rullo, A. Mudgerikar, E. Bertino, “Kalis—A system for knowledge-driven adaptable intrusion detection for the Internet of Things,” in Proc. 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS): 656-666, 2017.

[72] T. Lee, C. Wen, L. Chang, H. Chiang, M. Hsieh, “A lightweight intrusion detection schemebased on energy consumption analysis in 6LowPAN,” in Advanced Technologies, Embedded and Multimedia for Human-centric Computing, 1205-1213 2014.

[73] S. Raza, L. Wallgren, T. Voigt, “SVELTE: Real-time intrusion detection in the Internet of Things,” Ad Hoc Networks, 11(8): 2661-2674, 2013.

[74] T. Matsunaga, K. Toyoda, L. Sasase, “Low false alarm rate RPL network monitoring system by considering timing inconstancy between the rank measurements,” in Proc. 2014 11th International Symposium on Wireless Communications Systems (ISWCS): 427-431, 2014.

[75] G. Han, J. Jiang, W. Shen, L. Shu, J. Rodrigues, “IDSEP: a novel intrusion detection scheme based on energy prediction in cluster-based wireless sensor networks,” IET Inf. Secur., 7(2):97-105, 2013.

[76] S. Wu Kim, “Physical integrity check in cooperative relay communications,” IEEE Trans. Wireless Commun., 14(11): 6401-6413, 2015.

[77] R. Ahmad, I. Alsmadi, W. Alhamdani, L. Tawalbeh, “A comprehensive deep learning benchmark for IoT IDS,” Comput. Secur., 114: 102588, 2022.

[78] L. Besson, P. Leleu, "A distributed intrusion detection system for ad-hoc wireless sensor networks: The AWISSENET distributed intrusion detection system," in Proc. 2009 16th International Conference on Systems, Signals and Image Processing: 1-3, 2009.

[79] D. Summerville, K. M Zach, Y. Chen, “Ultra-lightweight deep packet anomaly detection for internet of things devices,” in Proc. 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC): 1-8, 2015.

[80] M. Talukder, S. Sharmin, S. Uddin et al., “MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs,” Int. J. Inf. Secur., 2024.

[81] E. Cho, J. Kim, C. Hong, “Attack model and detection scheme for botnet on 6LoWPAN,” in Proc. Asia-Pacific Network Operations and Management Symposium, 515-518, 2009.

[82] D. Shreenivas, S. Raza, T. Voigt, “Intrusion detection in the RPL-connected 6LoWPAN Networks,” in Proc. 3rd ACM International Workshop on IoT Privacy, Trust, and Security: 31-38 2017.

[83] Z. Banković, José M. Moya, A. Araujo, J. Goyeneche, “Intrusion detection in sensor networks using clustering and immune systems,” in Proc. 10th international conference on Intelligent data engineering and automated learning (IDEAL'09): 408-415, 2009.

[84] D. Hadžiosmanović, R. Sommer, E. Zambon, P. Hartel, “Through the eye of the PLC: semantic security monitoring for industrial processes,” in Proc. 30th Annual Computer Security Applications Conference: 126-135, 2014.

[85] D. Oh, D. Kim, W. Ro, “A malicious pattern detection engine for embedded security systems in the Internet of Things,” Sensors, 14(12): 24188-24211, 2014.

[86] H. Sedjelmaci, S. Mohammed Senouci, M. Al-Bahri, “A lightweight anomaly detection technique for low-resource IoT devices: A game-theoretic methodology,” in Proc. 2016 IEEE International Conference on Communications (ICC): 1-6, 2016.

[87] X. Song, G. Chen, X. Li, “A weak hidden Markov Model based intrusion detection method for wireless sensor networks,” in Proc. 2010 International Conference on Intelligent Computing and Integrated Systems: 887-889, 2010.

[88] L. Deng, D. Li, X. Yao, D. Cox, H. Wang, “Mobile network intrusion detection for IoT system based on transfer learning algorithm,” Cluster Comput., 1-16, 2018.

[89] L. Liu, B. Xu, X. Zhang, X. Wu, “An intrusion detection method for internet of things based on suppressed fuzzy clustering,” EURASIP J. Wireless Commun. Networking, 113, 2018.

[90] A. Agah, S. K. Das, K. Basu, M. Asadi, “Intrusion detection in sensor networks: A non-cooperative game approach,” in Proc. 3rd IEEE International Symposium on Network Computing and Applications (NCA’04): 343-346, 2004.

[91] A. Agah, S. K. Das, “Preventing DoS attacks in wireless sensor networks: A repeated game theory approach,” Int. J. Network Secur., 5(2): 145-153, 2007.

[92] K. Yadav, A. Srinivasan, “iTrust: an integrated trust framework for wireless sensor networks,” in Proc. the 2010 ACM Symposium on Applied Computing: 1466-1471, 2010.

[93] C. Cervantes, D. Poplade, M. Nogueira, A. Santos, “Detection of sinkhole attacks for supporting secure routing on 6LoWPAN for Internet of Things,” in Proc. 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM): 606-611, 2015.

[94] A. Khan, Z. Herrmann, “A trust based distributed intrusion detection mechanism for internet of things,” in Proc. 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA): 1169-1176, 2017.

[95] T. Jiang, G. Wang, H. Yu, "A dynamic intrusion detection scheme for cluster-based wireless sensor networks," in Proc. World Automation Congress 2012: 259-261, 2012.

[96] A. Arrington, L. Barnett, R. Rufus, A. Esterline, "Behavioral modeling intrusion detection system (bmids) using internet of things (iot) behavior-based anomaly detection via immunity-inspired algorithms," in Proc. 2016 25th International Conference on Computer Communication and Networks (ICCCN): 1-6, 2016.

[97] C. Liu, J. Yang, R. Chen, Y. Zhang, J. Zeng, "Research on immunity-based intrusion detection technology for the internet of things," in Proc. 2011 Seventh International Conference on Natural Computation, 1: 212-216, 2011.

[98] V. Bhuse, A. Gupta, "Anomaly intrusion detection in wireless sensor networks," J. High Speed Networks, 15(1): 33-51, 2006.

[99] H. Arolkar, S. Sheth, V. Tamhane, "Ant colony based approach for intrusion detection on cluster heads in WSN," in Proc. ICCCS: 523-526, 2011.

[100] A. Hassanzadeh, R. Stoleru, "Towards optimal monitoring in cooperative ids for resource constrained wireless networks," in Proc. 20th International Conference on Computer Communications and Networks (ICCCN): 1-8, 2011.

[101] I. Onat, A. Miri, "An intrusion detection system for wireless sensor networks," in Proc. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2005.

[102] B. Stelte, G Rodosek, "Thwarting attacks on zigbee-removal of the killerbee stinger," in Proc. the 9th International Conference on Network and Service Management (CNSM): 219-226, 2013.

[103] N. Kumar, E. Nigussie, R. K. Kanth, S. Virtanen, J. Isoaho, "Distributed internal anomaly detection system for internet-of-things," in Proc. 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC): 319-320, 2016.

[104] T. Luo, S. Nagarajan, "Distributed anomaly detection using autoencoder neural networks in wsn for iot," in Proc. 2018 IEEE International Conference on Communications (ICC): 1-6, 2018.

[105] O. Can, O. K. Sahingoz, "A survey of intrusion detection systems in wireless sensor networks," in Proc. 2015 6th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO): 1-6, 2015.

[106] A. Amouri, V. Alaparthy, S. Morgera, "Cross layer-based intrusion detection based on network behavior for IoT," in Proc. 2018 IEEE 19th Wireless and Microwave Technology Conference (WAMICON): 1-4, 2018.

[107] L. Coppolino, S. D. Antonio, A. Garofalo, L. Romano, "Applying data mining techniques to intrusion detection in wireless sensor networks," in Proc. 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing: 247-254, 2013.

[108] N. Abhishek, T. Lim, B. Sikdar, A. Tandon, "An intrusion detection system for detecting compromised gateways in clustered IoT networks," in Proc. 2018 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR): 1-6, 2018.

[109] S. Rajasegarar, C. Leckie, M. Palaniswami, J. C. Bezdek, "Quarter sphere based distributed anomaly detection in wireless sensor networks," in Proc. IEEE ICC '07, 2007.

[110] S. Misra, P. Venkata, H. Agarwal, A. Saxena, M. S. Obaidat, "A learning automata based solution for preventing distributed denial of service in Internet of things," in Proc. 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing: 114-122, 2011.

[111] Y. Fu, Z. Yan, J. Cao, O. Koné, X. Cao, "An automata based intrusion detection method for internet of things," Mobile Inf. Syst., 2017.

[112] C. Wang, T. Feng, J. Kim, G. Wang, W. Zhang, "Catching packet droppers and modifiers in wireless sensor networks," IEEE Trans. Parallel Distrib. Syst., 23(5): 835-843, 2012.

[113] F. Bao, R. Chen, M. J. Chang, J. H. Cho, "Hierarchical trust management for wireless sensor networks and its applications to trustbased routing and intrusion detection," IEEE Trans. Netw. Serv. Manag., 9(2): 169-183, 2012.

[114] S. Etalle, "Network monitoring of industrial control systems: The lessons of security matters, in Proc. ACM Workshop on Cyber-Physical Systems Security & Privacy: 1–1, 2019.

[115] N. Kaur, P. Rattan, "A critical review of intrusion detection systems in WSN: challenges & future directions," Ann. Rom. Soc. Cell Biol., 25(4): 3020-3028, 2021.

[116] T. Wang, Z. Liang, C. Zhao, "A detection method for routing attacks of wireless sensor network based on fuzzy C-means clustering," in Proc. 6th International Conference on Fuzzy Systems and Knowledge Discovery: 445-449, 2009.

[117] K. Q. Yan, S. C. Wang, C. W. Liu, "A hybrid intrusion detection system of cluster-based wireless sensor networks," in Proc. International Multi Conference of Engineers and Computer Scientists: 1, 2009.

[118] N. Chitradevi, V. Palanisamy, K. Baskaran, S. Prabeela, "Efficient distributed clustering-based anomaly detection algorithm for sensor stream in clustered wireless sensor networks," Eur. J. Sci. Res., 54(4):484-498, 2011.

[119] M. Chauhan, M. Agarwal, "Study of various intrusion detection systems: A Survey," in Smart and Sustainable Intelligent Systems: 355-372, 2021.

[120] Z. Ahmad, A. Shahid Khan, , C. Shiang, J. Abdullah, F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," Trans. Emerg. Telecommun. Technol., 32(1): e4150, 2021.

[121] A. A. Anitha, L. Arockiam, "A review on intrusion detection systems to secure IoT networks, "Int. J. Comput. Networks Appl., 9(1): 38-50, 2022.

[122] Y. Farooq, H. Beenish, M. Fahad, "Intrusion detection system in wireless sensor networks-a comprehensive survey," in Proc. 2019 Second International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT): 1-6, 2019.

[123] N. Singh, D. Virmani, X. Z Gao,"A fuzzy logic-based method to avert intrusions in wireless sensor networks using WSN-DS dataset," Int. J. Comput. Intell. Appl., 19(03): 2050018, 2020.

[124] M. Wazid, A. K. Das, S. Kumari, M. K. Khan, "Design of sinkhole node detection mechanism for hierarchical wireless sensor networks," Secur. Commun. Networks, 9(17): 4596-4614, 2016.

[125] G. Liu, H. Zhao, F. Fan, , Liu, G, Q. Xu, , S. Nazir, "An enhanced intrusion detection model based on improved kNN in WSNs," Sensors, 22(4): 1407, 2022.

[126] B. Hasan, S. Alani, M. A. Saad, "Secured node detection technique based on artificial neural network for wireless sensor network," Int. J. Electr. Comput. Eng., 11(1): 2088-8708, 2021.

[127] W. Huai-bin, Y. Zheng, W. Chun-dong, "Intrusion detection for wireless sensor networks based on multi-agent and refined clustering," in Proc. International Conference on Communications and Mobile Computing: 450-454, 2009.

[128] S. Siripanadorn, W. Hattagam, N. Teaumroong, "Anomaly detection in wireless sensor networks using self-organizing map and wavelets," Int. J. Common. 4(3):74-83, 2010.

[129] M. Sa, M. R. Nayak, A. K. Rath, "A simple agent based model for detecting abnormal event patterns in a distributed wireless sensor networks," Int. J. Comput. Sci. Secur., 4(6):580-588, 2011.

[130] H. Sedjelmaci, M. Feham, "Novel hybrid intrusion detection system for clustered wireless sensor network," Int. J. Netw. Secur. Appl., 3(4):1-14, 2011.

[131] C. Henry Tseng, Y. Chang, "EBDM: Ensemble binary detection models for multi-class wireless intrusion detection based on deep neural network," Comput. Secur., 133: 103419, 2023.

[132] M. Hosseini Shirvani, M. Masdari, "A survey study on trust-based security in internet of things: challenges and issues, Internet Things, 100640, 2022.

[133] D. Jayanayudu, A. C. Sudhir, "Shuffled frog leap and ant lion optimization for intrusion detection in iot-based WSN," in Proc. Fourth International Conference on Computer and Communication Technologies, 606, 2023.

[134] M. Hosseini Shirvani, A. M. Rahmani, A. Sahafi, "An iterative mathematical decision model for cloud migration: A cost and security risk approach," Softw. Pract. Exper., 2017: 1-37, 2017.

[135] M. Saied, S. Guirguis, M. Madbouly, "Review of artificial intelligence for enhancing intrusion detection in the internet of things," Eng. Appl. Artif. Intell., 127(A): 107231, 2024.

[136] L. Webber, A. Arafa, A. Mehbodniya, S. Karupusamy, B. Sha, A. Dahiya, P. Kanani, "An efficient intrusion detection framework for mitigating blackhole and sinkhole attacks in healthcare wireless sensor networks," Comput. Electr. Eng., 111: 108964, 2023.

[137] M. S. Korium, M. Saber, A. Beattie, A. Narayanan, S. Sahoo, P. H. J. Nardelli, "Intrusion detection system for cyberattacks in the Internet of Vehicles environment," Ad Hoc Networks, 153: 103330, 2024.

[138] E. Ntizikira, L. Wang, J. Chen, K. Saleem, "Honey-block: Edge assisted ensemble learning model for intrusion detection and prevention using defense mechanism in IoT," Comput. Commun., 214: 1-17, 2024.

[139] V. Sivagaminathan, M. Sharm, S. Henge, "Intrusion detection systems for wireless sensor networks using computational intelligence techniques," Cybersecurity, 6: 27, 2023.

Biographies

Mirsaeid Hosseini Shirvani received his B.Sc., M.Sc., and Ph.D. all in Computer Software Engineering Systems at Universities in Tehran, Iran. He has been teaching miscellaneous computer courses in several universities in Mazandaran province of IRAN since 2001. He also published several papers in authentic and worldwide well-reputed journals. Currently, he serves as a Professor in the Computer Engineering Department at IAU (Sari-Branch). He was in Stanford's top 2% of the most cited scientist across the world in two consecutive years 2021 and 2022. His research interests are in the areas of cloud computing, fog computing, IoT, distributed systems, parallel processing, machine learning, and evolutionary computation.

- Email: mirsaeid_hosseini@iausari.ac.ir & mirsaeid_hosseini@yahoo.com
- ORCID: 0000-0001-9396-5765
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Amir Akbarifar received his B.Sc. and M.Sc. in Computer Software Engineering Systems in Islamic Azad University in IRAN. Currently, He is a Ph.D. candidate in Computer Engineering Department at IAU (Sari-Branch). He has been teaching computer courses in Gorgan Islamic Azad University in Golestan province of IRAN since 2022. He has published numerous articles in prestigious magazines. He is an information Security Analyst and his fields of study includes: Security, Software Architecture, Artificial Intelligence, Quantum Computing, and Cryptography.

- Email: Amir.Akbarifar@gmail.com
- ORCID: 0000-0003-3227-2847
- Web of Science Researcher ID: AFC-8068-2022
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
M. Hosseini Shirvani, A. Akbarifar, "A survey study on intrusion detection system in wireless sensor network: Challenges and considerations," J. Electr. Comput. Eng. Innovations, 12(2): 449-474, 2024.

DOI: 10.22061/jecei.2024.10148.683

URL: https://jecei.sru.ac.ir/article_2116.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Enhancing High-Performance Computing: A Comprehensive Study on Dual-Doped Source/Drain Reconfigurable Field Effect Transistor

Z. Ahangari *

Department of Electronic, Yadegar- e- Imam Khomeini (RAH) Shahr-e-Rey Branch, Islamic Azad University, Tehran, Iran.

Article Info

Article History:
Received 15 March 2024
Reviewed 20 April 2024
Revised 01 June 2024
Accepted 26 June 2024

Keywords:
Reconfigurable field effect transistor
Multi-Doped source/drain
Gate workfunction
Logic gate

*Corresponding Author's Email Address:
z.ahangari@gmail.com

Abstract

Background and Objectives: In this study, a reconfigurable field-effect transistor has been developed utilizing a multi-doped source-drain region, enabling operation in both n-mode and p-mode through a simple adjustment of electrode bias. In contrast to traditional reconfigurable transistors that rely on Schottky barrier source/drain with identical Schottky barrier height, the suggested device utilizes a straightforward fabrication process that involves physically multi-doped source and drain. The proposed structure incorporates a bilayer of n^+ and p^+ in the source and drain regions.
Methods: The device simulator Silvaco (ATLAS) is utilized to conduct the numerical simulations.
Results: The transistor exhibits consistent transfer characteristics in both modes of operation. The influence of key design parameters on device performance has been analyzed. A notable aspect of this transistor is the integration of an XNOR logic gate within a single device, rendering it suitable for high-performance computing circuits. The findings indicate that on-state currents of 142 µA/µm and 57.2 µA/µm, along with on/off current ratio of 8.68×10^7 and 3.5×10^7, have been attained for n-mode and p-mode operation, respectively.
Conclusion: A single-transistor XNOR gate design offers potential advantages for future computing circuits due to its simplicity and reduced component count, which could lead to smaller, more energy-efficient, and potentially faster computing systems. This innovation may pave the way for advancements in low-power and high-density electronic devices.

Introduction

The scaling of MOSFETs (Metal-Oxide-Semiconductor Field-Effect Transistors) and advancements in the complementary metal-oxide-semiconductor (CMOS) industry play a crucial role in the progression of high-performance computing and logic gate technologies. This advancement is typically linked with higher operation frequencies, reduced power consumption, and lower cost per unit. Nevertheless, it also brings about certain challenges such as leakage current, short-channel effects, dopant impurities, and interconnect delays [1]-[3]. Overcoming these obstacles is essential for the continual enhancement of high-performance computing technologies. More complex structures like FinFETs and silicon-on-insulator (SOI) devices have been introduced as substitutes for traditional CMOS devices [4]-[7]. Nevertheless, the challenge persists in terms of the requirement for expensive processing technology and the continuous pursuit of enhancing transistor density on a chip.

Reconfigurable field effect transistors (RFET) have been proposed as an innovative alternative to traditional MOSFETs, offering the ability to switch between n-mode and p-mode operations by adjusting the bias of their

electrodes without the need for additional doping [8]-[12]. This enhanced flexibility enables a wide range of applications compared to the fixed operation mode of MOSFETs. Unlike the impurity doping used in classical CMOS to define polarities in the channel, source and drain regions, reconfigurable transistors utilize electrostatic doping to regulate carrier injection through Schottky barriers at the source and drain. In n-mode, the transistor acts as an electron-rich semiconductor, facilitating current flow between the source and drain. Conversely, in p-mode, the transistor functions as a hole-rich channel, enabling current flow in a different manner. The operational flexibility of reconfigurable transistors provides numerous advantages over MOSFET-based logic gates, including functional flexibility, optimized power consumption, adaptability to various applications, and simplified design processes.

The RFET operates based on two gates, the control gate and the program gate. The control gate controls the device on and off state, while the program gate modulates the band bending to determine which type of carrier (electron or hole) can flow from the source to the drain along the channel. It is crucial for both n-mode and p-mode operations to have identical transfer characteristics. This requires a nearly equal Schottky barrier height for both electrons and holes. In the case of a Silicon channel material, Nickel Silicide ($NiSi_2$) is used as the source/drain Schottky contact, providing a Schottky barrier height of 0.59 eV for electrons. Considering the silicon band gap value of 1.12 eV, a Schottky barrier height of 0.53 eV can be achieved for holes [13]. Experimental investigations on the electrical characteristics of a nanowire RFET have shown near-symmetric transfer characteristics. The utilization of multi-gate RFETs in circuit design has been explored to enable the development of multi-stage digital circuits with two threshold voltages [14]. Additionally, various RFET structures have been implemented using SiGe [10] and two-dimensional materials [15]-[16]. Reconfigurable transistor based on extended source configuration is presented in [17], where the $NiSi_2$ Schottky contact of the source is extended across the channel region to enhance the tunneling area. As a consequence, this modification leads to an increase in the off-state current. The impact of the position of the control gate and program gate on the efficiency of the Schottky barrier RFET is examined. The findings indicate that when tunneling happens at the junction of the source and channel areas, decreasing the spacer length in the source region leads to an increase in the on-state current [18]. In [19], a logic gate utilizing Schottky source-drain RFET technology is described. The gate architecture involves the integration of multiple transistors to facilitate the construction of both NAND and NOR gates. In the context of neural network applications, the utilization of a XNOR logic gate that relies on a metal source-drain structure is examined [20].

One of the main challenges of RFET is adjusting the Schottky barrier height to ensure identical n-mode and p-mode operation across different materials. Fermi level pinning poses a significant challenge by keeping the Schottky barrier height independent of the metal workfunction. In [21], a novel reconfigurable field effect transistor is introduced, which is based on a dual doped source/drain region. This device incorporates three gates, including two program gates located at the interface of the source and drain regions, and a control gate positioned in the middle of the channel. Furthermore, the device features a floating gate that serves as a charge storage element, enabling seamless switching between the n-mode and p-mode. However, limitation of this study is the necessity of high voltage for programming the device and storing charges. Additionally, the device requires the accumulation of opposite charges to transition from the n-mode to the p-mode, thereby limiting its dynamic switching capabilities. The feasibility of 3D dual doped RFET is considered in [22]. In [23], the proposed reconfigurable transistor has asymmetric structure. The source region is dual doped while the drain region is Schottky contact. However, the use of Schottky contacts in RFET devices is limited by the requirement for nearly equal Schottky barrier heights to achieve identical behavior in both n-mode and p-mode operation. This restriction hinders the exploration of alternative materials for this device structure, as the Schottky barrier height is typically dependent on the material used. This limitation restricts the use of different materials for the device, which can limit its potential performance and scalability. In contrast to the conventional RFET as described in [24], the program gate is situated at the source side while the control gate is positioned at the drain side. This particular arrangement serves to decrease the subthreshold swing of the device; however, the positioning of the control gate at the drain side renders it vulnerable to fluctuations in drain bias. The analog performance of the device with dual doped source and drain regions is considered in [25]. The device shows identical characteristic with respect to the Schottky barrier RFET. Different techniques for inducing negative differential resistance (NDR) region in the transfer characteristic of the RFET is investigated [26]. In [27], an RFET device containing three gates within the device structure is investigated. The two program gates are situated at the boundary between the source and channel regions, while the control gate is positioned centrally within the channel. Through the manipulation of the bias voltages applied to these gates, modulation of the charge density in the channel is achieved, leading to band-to-band tunneling occurring at the center of the channel.

The device is capable of functioning in both the TFET and conventional MOSFET modes. The drain current of the device is particularly affected by variations in the gate bias values, with higher gate bias utilized for charge accumulation.

In this study, a double gate dual-doped RFET (DDRFET) is proposed, where Silicon material is used in the source and drain regions instead of Schottky contacts. To achieve reconfigurable operation, half of the source region is doped n^+ and the other half is doped p^+. This novel structure demonstrates consistent behavior for both n-mode and p-mode operations. The device is equipped with two control gates, namely the program gate and the control gate, which play a crucial role in regulating the flow of carriers from the source to the channel. The impact of geometrical design parameters on the device performance is thoroughly evaluated. Basically, the proposed double gate architecture is more scalable than 3D FinFETs due to its ability to maintain the double gate structure even at smaller dimensions. This allows for more aggressive scaling while maintaining performance and reducing power consumption. The employment of a dual doped configuration in the source and drain region enhances the feasibility of utilizing this device with various materials.

The main focus of this research is the absence of Schottky contacts in the source and drain regions, eliminating the need for a spacer region to isolate the gate and source electrodes. This configuration allows for optimal gate controllability at the interface between the source and channel region. Additionally, the proposed DDRFET structure incorporates the XNOR logic gate utilizing a single transistor. The control gate and program gate values are designated as inputs, while the drain current level serves as the output. An essential characteristic of the XNOR gate designed with a dual doped device in comparison with Schottky barrier RFET is the drain current insensitivity to the Schottky barrier height for both electrons and holes. This paper opts for low gate bias settings for the gates, resulting in the operation of the device solely in the MOSFET mode and distinct current levels being observed. The XNOR gate design presented in this research, which relies on the level of drain current, enables the efficient implementation of logic operations within the device. The manuscript is structured in the subsequent manner: the configuration of the device and simulation models are presented in the succeeding section. Subsequently, findings concerning the influence of design parameters on the performance of the suggested DDRFET are discussed, along with the truth table of the designed logic gates. To conclude, a summary of the paper is provided in the concluding section.

Device Structure and Simulation Models

Fig. 1 illustrates the 2D diagram of the proposed reconfigurable Field-Effect Transistor (FET) that utilizes dual doped source/drain regions. The source and drain regions consist of a bilayer heavily doped p^+-n^+ junction. This device incorporates two distinct gate electrodes, which are separated by an insulating region. The control gate (V_{CG}), positioned in close proximity to the source region, is responsible for both the operational on and off states of the device, as well as the modulation of the potential barrier at the interface of the source and channel region. On the other hand, the program gate (V_{PG}) controls the band bending and determines the type of carrier that flows, whether it operates in the p-FET or n-FET mode. The p^+ and n^+ bilayers in the source and drain regions possess equal potential (zero volt in the source region and V_{DS}=1V in the drain region), effectively preventing band to band tunneling in these regions. When the control gate and program gate have a positive value, the device functions in the n-mode. In this mode, the control gate adjusts the potential barrier at the interface between the n^+ source and the channel. As a result of the accumulation of electrons in the channel, a high potential barrier forms at the interface of the p^+ source. This barrier prevents the flow of holes in the channel. Conversely, when the control gate and program gate have a negative value, the device operates in the p-mode. In this mode, the control gate modulates the potential barrier at the p^+ source, subsequently blocking the flow of the opposite charge from the n^+ source.

The initial design parameters are summarized in Table 1.

Table 1: Initial design parameters of the proposed DDRFET

Parameter	Value
Workfunction of the control gate	4.7 eV
Workfunction of the control gate	4.7 eV
Channel thickness	10 nm
Channel length	150 nm
Length of control gate	70 nm
Length of program gate	70 nm
Spacer length	10 nm
Channel doping density	Intrinsic
n^+ source/drain doping density	10^{19} cm^{-3}
p^+ source/drain doping density	10^{19} cm^{-3}
Gate oxide thickness (hfo_2)	1 nm

Silvaco device simulator [28] is utilized to conduct numerical simulations, and the simulator incorporates the following models: (a) Drift and diffusion model; The device operates on the principle of modulating the potential barrier at the interface between the source and channel region through the control gate bias. The drift current is generated as a result of carriers moving in

response to an electric field, while the diffusion current is caused by the transportation of charges due to the non-uniform concentration of charged carriers in a semiconductor material. (b) Auger and Shockley-Read-Hall (SRH) recombination models have been developed to provide a comprehensive understanding of the impact of recombination and the generation of carriers on carrier transport in the presence of defects and traps. (c) Mobility Models; The carrier drift velocity, which plays a crucial role, is greatly affected by carrier mobility. The simulation considers the impact of both horizontal and vertical electric fields on carrier mobility, utilizing relevant models. Moreover, the presence of dopants as impurities negatively affects carrier mobility, resulting in a decrease in velocity due to scattering [29]-[30]. Mobility models that incorporate the influence of dopants are also integrated. (d) Quantum Confinement Model; The Quantum Confinement effect is primarily observed in thin channel thicknesses, resulting in energy sub-bands characterized by elevated energy levels confined within the channel [31]-[33]. (e) The band-to-band tunneling model is utilized to describe the phenomenon of carriers tunneling at the interface of the n^+-p^+ region. It is important to highlight that the n^+-p^+ in the source and regions possess identical potential. Consequently, there is no carrier transport across the dual doped regions.

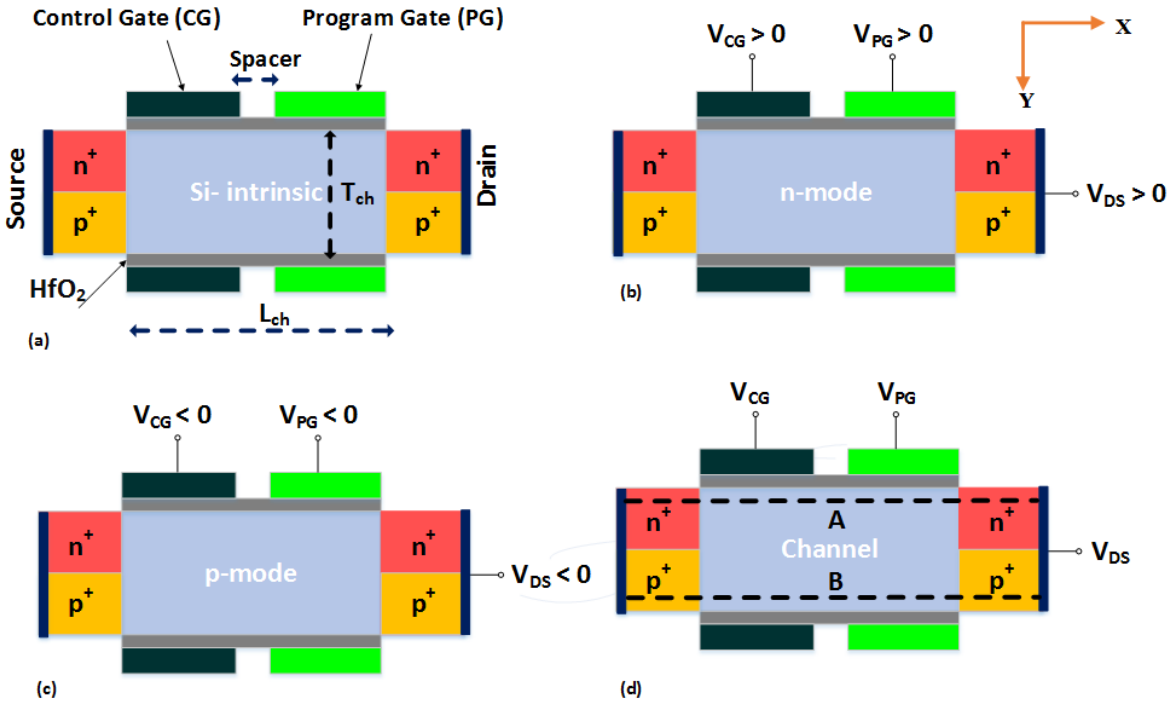


Fig. 1: (a) 2D schematics of the proposed DDRFET, (b) n-mode operation of the DDRFET, (c) p-mode operation and (d) cutline along which the energy bands are illustrated.

Results and Discussion

The distinguishing characteristic of the DDRFET is its ability to operate in both the n-mode and p-mode simply by adjusting the bias values of the control gate, program gate, and drain electrodes. This is made possible by the presence of dual doped n^+ and p^+ source and drain regions, which allow the necessary carriers to move through the device. In the n-mode operation, where all the electrodes have positive values, electrons accumulate in the channel, enabling the flow of n^+ source carriers from source to drain. Conversely, the accumulation of electrons in the channel creates a high potential barrier at the interface of the p^+ source and channel, preventing the flow of holes in the channel. Similarly, when the electrodes are negatively biased, holes accumulate in the channel, resulting in a low potential barrier at the interface of the p^+ source and channel. This facilitates the flow of holes through the channel. At the same time, a high potential barrier is formed at the interface of the n^+ source and channel, hindering the movement of carriers with opposite charge.

Fig. 2 displays the energy band diagram of the DDRFET in both the off-state and on-state for n-mode operation from the n^+ source and p^+ source. In the off-state for n-mode operation, shown in Fig. 2 (a), the electrodes are biased as follows: V_{CG}=0V, V_{PG}=1V, and V_{DS}=1V. The energy band diagram from the n+ source shows a high potential barrier at the interface of the n^+ source and channel region due to the absence of the control gate. As

illustrated in Fig. 2 (b), the energy band in the off-state from the p+ source indicates a high potential barrier for holes to flow. In this case, only minority carriers from the p^+ source region contributes to the current. As the control gate bias increases towards positive values, depicted in Fig. 2 (c), electrons accumulate in the channel, gradually reducing the potential barrier at the interface of the n+ source and the channel region. It is important to note that the program gate alters the energy band bending at the drain side, facilitating the electron flow from source to drain. Conversely, as, depicted in Fig. 2 (d) the accumulation of electrons in the channel increases the potential barrier at the p+ source side, hindering the flow of holes in the channel.

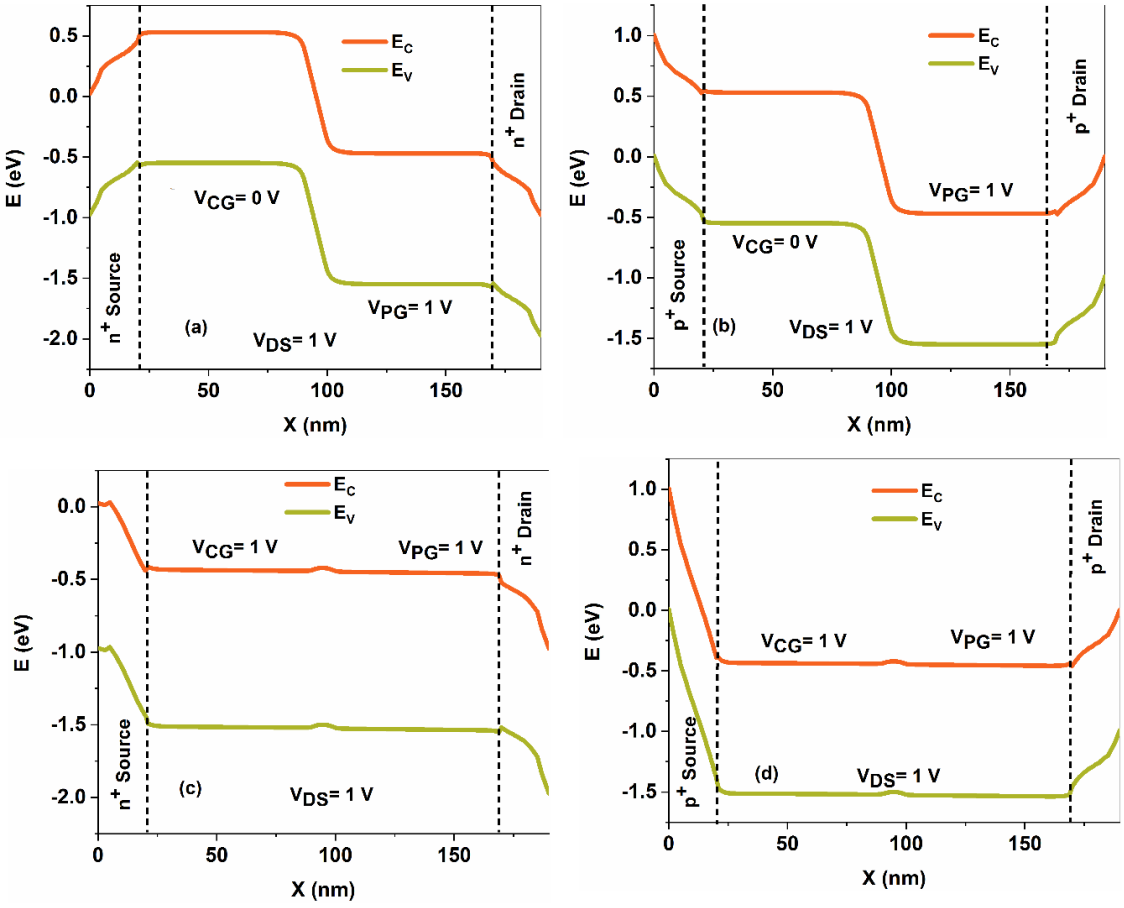


Fig. 2: Energy band diagram of DDRFET in the n-mode operation along the device from source to drain in (a) off-state condition from the n^+ source to n+ drain (cutline along the A direction), (b) off-state condition from the p^+ source to p^+ drain (cutline along the B direction), (c) on-state condition from the n^+ source to n^+ drain (cutline along the A direction), and (d) on-state condition from the p^+ source to p^+ drain (cutline along the B direction).

Similarly, the band diagram of the proposed DDRFET for p-mode operation from the n^+ and p^+ source is presented in Fig. 3. In Fig. 3 (a), under off-state conditions with V_{CG}=0V, V_{PG}=-1V, and V_{DS}=-1V, the band bending near the drain region in the channel is a result of the program gate that blocks the flow of electrons in the channel. In addition, a high potential barrier is created at the n+ source-channel interface and only minority carriers from the n+ source region contributes to the off-state current. Furthermore, Fig. 3 (b) shows a high potential barrier at the channel-p^+ source interface that hinders hole flow in the channel. Transitioning to the on-state by increasing the control gate voltage to negative values leads to hole accumulation in the channel, impeding electron flow from the n^+ source, as depicted in Fig. 3 (c). Nevertheless, the accumulated holes in the channel facilitate hole transport from the p^+ source region towards the drain, as illustrated in Fig. 3 (d).

Fig. 4 depicts the transfer characteristics of the DDRFET, showcasing its performance in both n-mode and p-mode operation. Notably, the device exhibits a nearly symmetrical behavior for both modes.

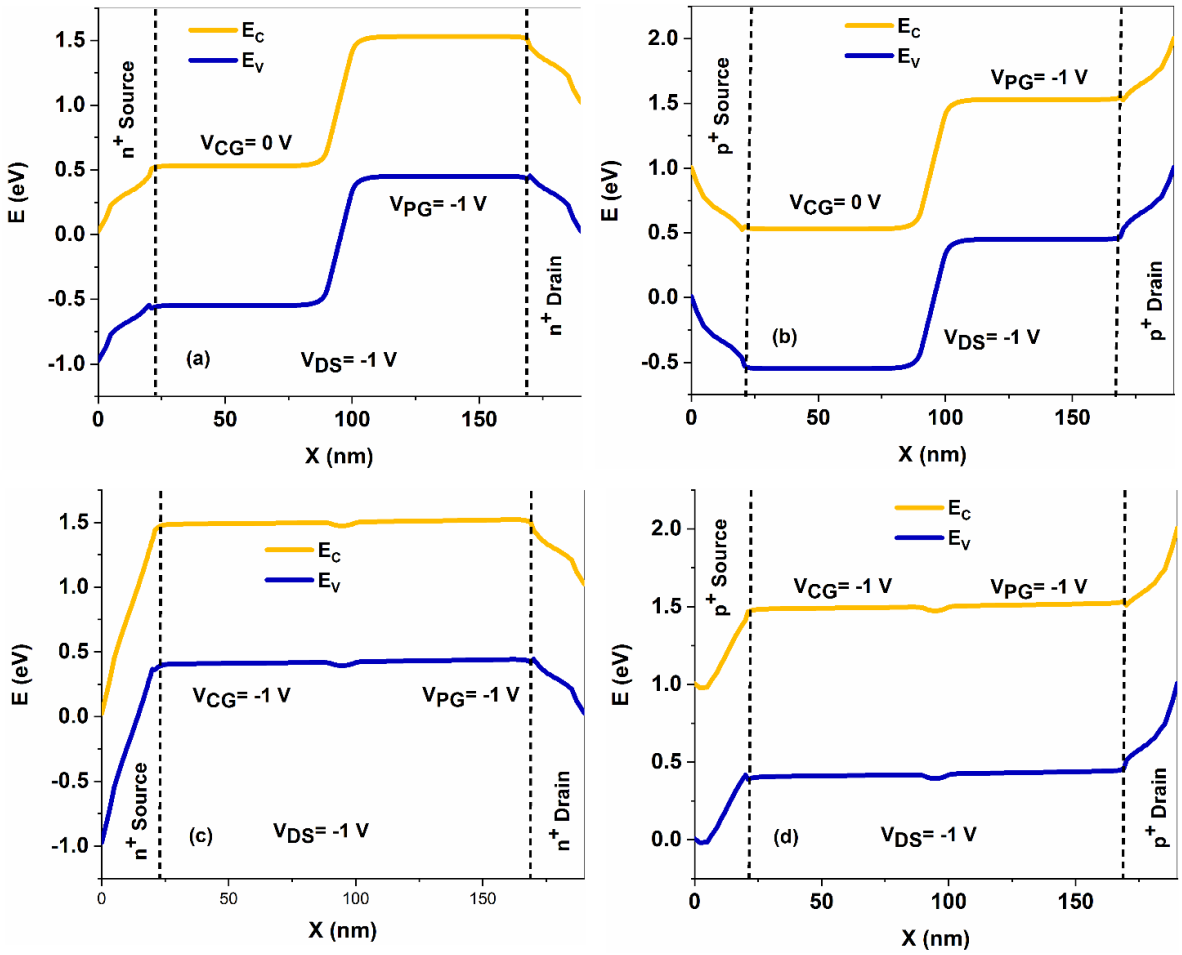


Fig. 3: Energy band diagram of DDRFET in the p-mode operation along the device from source to drain in (a) off-state condition from the n^+ source to n+ drain (cutline along the A direction), (b) off-state condition from the p^+ source to p^+ drain (cutline along the B direction), (c) on-state condition from the n^+ source to n^+ drain (cutline along the A direction), and (d) on-state condition from the p^+ source to p^+ drain (cutline along the B direction).

In the case of n-mode operation, it achieves an on-state current of 142 µA/µm, an on/off current ratio of 8.68×10^7, threshold voltage of 0.23 V and a subthreshold swing of 64.7 mV/dec. Similarly, for p-mode operation, the DDRFET attains an on-state current of 57.2 µA/µm, an on/off current ratio of 3.5×10^7, threshold voltage of -0.301 V and a subthreshold swing of 80.2 mV/dec. By definition, threshold voltage is the control gate bias level at which a drain current of 10^{-8} A/µm is attained. The gate workfunction plays a crucial role in the design of the device, significantly impacting its performance. In principle, the DDRFET functions by altering the polarity of the electrodes to achieve comparable transfer characteristics for both n-mode and p-mode operations. The difference in workfunction between the gate and the channel effectively controls the carrier density within the channel. Fig. 5 demonstrates how variations in gate workfunction affect the performance of the DDRFET in both n-mode and p-mode operations. The findings indicate that lower gate workfunction values lead to a significant increase in electron density in the channel during the off-state, resulting in higher off-state current in n-mode operation. In p-mode operation, a higher negative bias is needed initially to deplete the accumulated electrons in the channel, leading to an increase in the threshold voltage for the onset of p-mode operation. Increasing the gate workfunction reduces the electron density in the channel, thereby decreasing the off-state current. Conversely, in p-mode operation, an increase in hole density in the channel enhances the on-state current. To ensure optimal device performance, it is essential to achieve similar transfer characteristics with nearly identical threshold voltages. The study suggests that a gate workfunction value of 4.7 eV is optimal for achieving the lowest off-state current of 1.63×10^{-12} A/µm and threshold voltages of 0.23 V and -0.301 V for n-mode and p-mode operations, respectively. Beyond this optimal value, an increase in gate workfunction leads to higher hole density in the channel, resulting in increased off-state current in p-mode operation, while a reduction in electron density in n-mode operation leads to an increase in threshold voltage.

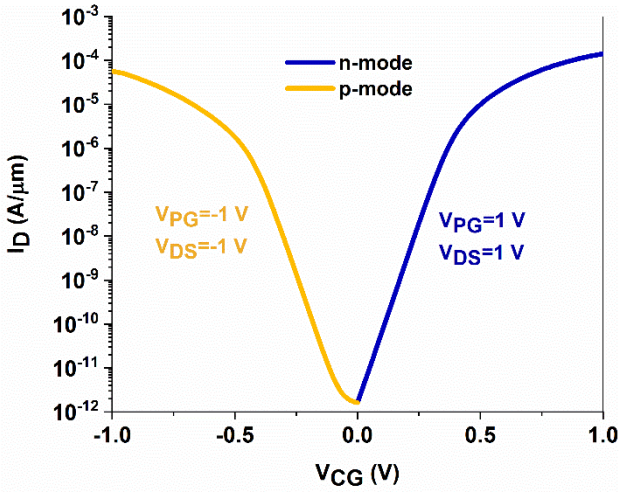


Fig. 4: Transfer characteristics of DDRFET in the n-mode and p-mode operation.

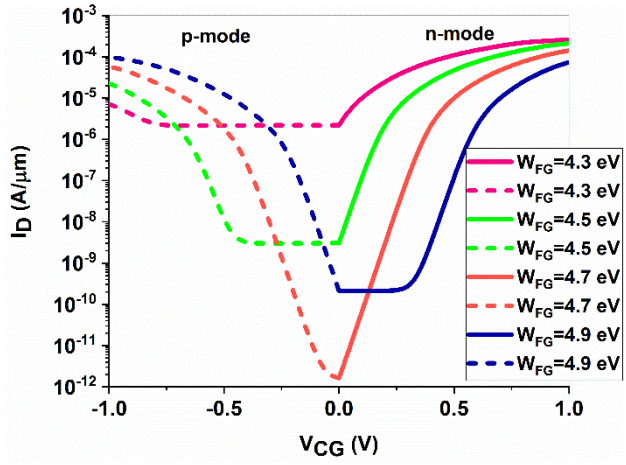


Fig. 5: Transfer characteristics of DDRFET in the n-mode and p-mode operation as the workfunction of the control gate and program gate are parameterized.

Fig. 6 demonstrates the impact of the length of the control and program gate on the on-state current of the DDRFET for n-mode and p-mode operation. When the gate length decreases, a slight decrease in the on-state current is observed due to the increase in parasitic resistance of the uncovered channel region. It is important to note that the off-state current is determined by the potential barrier height at the source-channel interface. The modulation of the potential barrier at the dual-doped source region interface by the control gate results in minimal changes in the off-state current (not depicted).

Fig. 7 depicts the impact of channel thickness on the performance of the proposed DDRFET. It is important to note that, due to the dual-doped source/drain structure, current flows from the source specific doping that matches the majority carrier accumulated in the channel. The findings indicate that as the channel thickness decreases, quantum confinement leads to an increase in energy of the states, resulting in a decrease in the density of states. Furthermore, reducing the source/drain thickness leads to a decrease in the source-channel interface area for carrier transport initiation. It is evident that as the channel thickness decreases, the density of accumulated majority carriers in the channel also decreases.

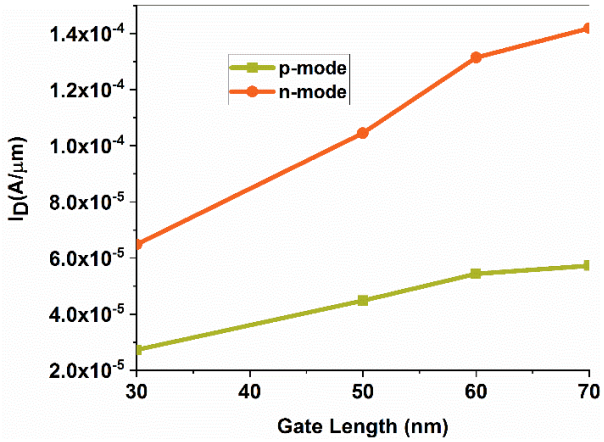


Fig. 6: On-state current of DDRFET in the n-mode and p-mode operation as the gate length is varied.

The 2D contour displayed in Fig. 8 illustrates the variation of drain current in relation to changes in the control gate and program gate. The contour reveals that the highest drain current values are observed in the top-right and bottom-left sections. In the top-right region, both the control gate and program gate exhibit positive values, leading to n-mode operation and maximum drain current output when the control and program gates reach their peak values.

Conversely, when the program gate and control gate are negative, the device operates in p-mode. It is important to note that in p-mode operation, the roles of the drain and source electrodes are reversed, with the drain electrode at 0.05 V and the source electrode at 0 V. On the contrary, when the control gate and program gate are biased oppositely, lower drain current values are anticipated, as shown in the top-left and bottom-right areas of the contour. This behavior is primarily influenced by the program gate, which creates a barrier for carrier transport, hindering the flow of accumulated carriers in the channel controlled by the control gate. A key characteristic of the proposed DDRFET is its ability to function as a logic gate within a single device. The control gate and program gate serve as inputs, while the drain current level functions as the output of the logic gate. A high gate potential of 1.0 V corresponds to logic "1", while a low gate potential of -1.0 V represents logic "0". A drain current below 10^{-12} A/μm signifies logic "0", whereas a drain current exceeding 10^{-8} A/μm indicates logic "1". The drain current remains constant at V_{DS}=0.05 V. Table 2 displays the truth table for the suggested logic gate.

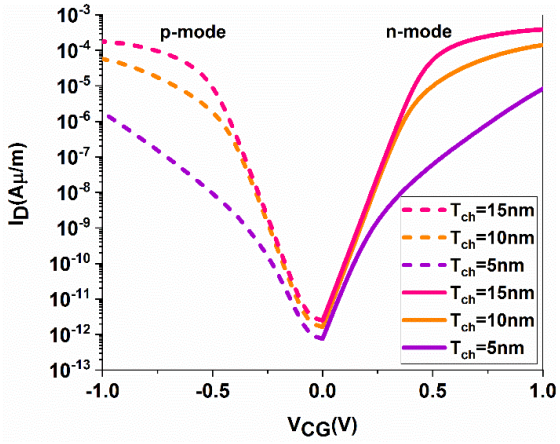


Fig. 7: I_D-V_{CG} curves of DDRFET in the n-mode and p-mode operation as the channel thickness is varied.

Table 2: Truth table of the proposed XNOR gate implemented by the proposed DDRFET

V_{CG}	V_{PG}	on-state current
0	0	**1**
0	1	**0**
1	0	**0**
1	**1**	**1**

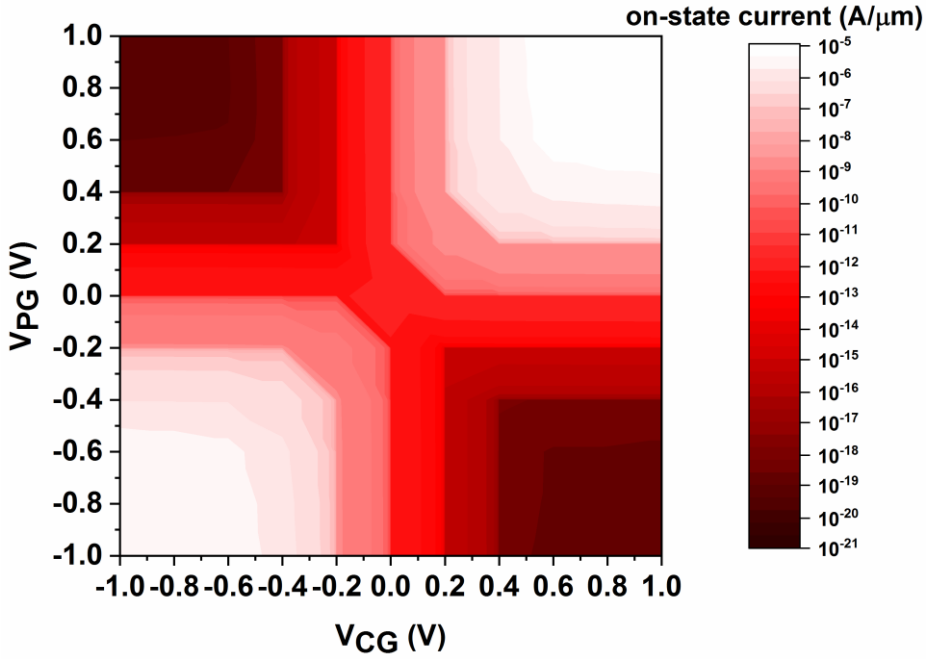


Fig. 8: 2D variation matrix of on-state current as the control gate and program gate bias is varied.

The influence of temperature on the device performance for n-mode and p-mode operation is illustrated in Fig. 9. The findings indicate that the off-state current of the device is significantly influenced by changes in temperature, attributed to the presence of increased energy carriers at elevated temperatures. Conversely, the saturation current of the device, which is controlled by the gate bias, remains unaffected by temperature variations.

The impact of drain bias on the transfer characteristics of the DDRFET being considered is illustrated in Fig. 10. The device demonstrates a minimal sensitivity of the off-state current to changes in the drain electric field, resulting in significant resistance to drain bias even with a reduced channel length. It is apparent that the rise in on-state current is attributed to the heightened carrier velocity observed at elevated drain voltages.

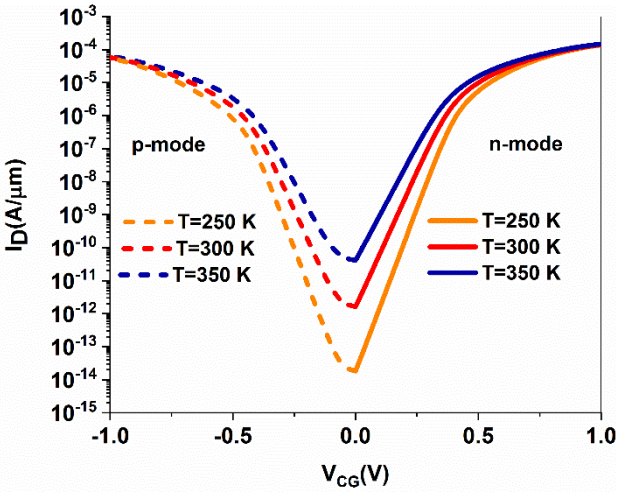


Fig. 9: I_D-V_{CG} curves of DDRFET in the n-mode and p-mode operation as the temperature is parametrized.

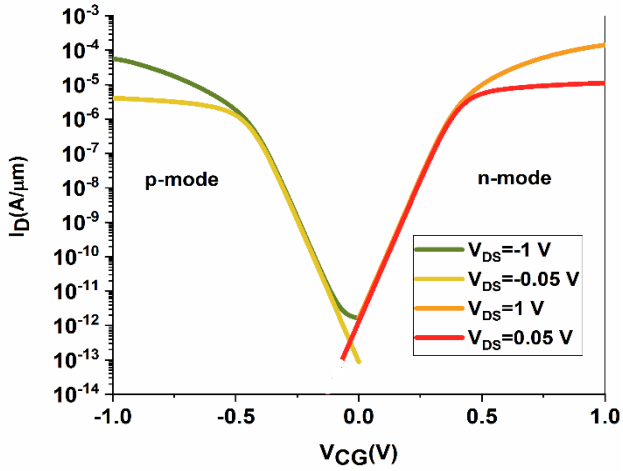


Fig. 10: Transfer characteristics of DDRFET in the n-mode and p-mode operation as the drain bias is parametrized.

Conclusion

The electrical properties of a reconfigurable field effect transistor are thoroughly examined in this study, utilizing multi doped source and drain regions. The findings indicate that the gate workfunction plays a crucial role in effectively controlling the channel charge density. To achieve similar transfer characteristics for both n-mode and p-mode operations, it is imperative to determine the optimal value for the gate workfunction. The length of the control gate and program gate has minimal impact on the device's performance, whereas extremely thin film channel thicknesses negatively affect the device on-state current. A key aspect of the proposed design is the integration of an XNOR gate within a single device, which simplifies the creation of high-speed computing circuits.

Author Contributions

Z. Ahangari discussed, simulated the results and contributed to the final manuscript.

Acknowledgment

The author would like to thank the editor and anonymous reviewers.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistors
CMOS	Complementary Metal-Oxide-Semiconductor
RFET	Reconfigurable Field Effect Transistor
DDRFET	Dual-Doped Reconfigurable Field Effect Transistor
SOI	Silicon on Insulator

References

[1] A. Chaudhry, M. J. Kumar, "Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review," IEEE Trans. Device Mater. Reliab., 4(1): 99-109, 2004.

[2] S. H. Noh et al., "Improvement in short-channel effects of the thin-film transistors using atomic-layer deposited In--Ga--Sn--O channels with various channel compositions," IEEE Trans. Electron Devices, 69(10): 5542-5548, 2022.

[3] P. Arul, K. Helen Prabha, "A comprehensive analysis of short channel effects on carbon nano tube field effect transistors," J. Nanoelectron. OptoElectron., 16(12): 1905-1912, 2021.

[4] R. K. Maurya, B. Bhowmick, "Review of FinFET devices and perspective on circuit design challenges," Silicon, 14(11): 5783-5791, 2022.

[5] A. Razavieh, P. Zeitzoff, E. J. Nowak, "Challenges and limitations of CMOS scaling for FinFET and beyond architectures," IEEE Trans. Nanotechnol., 18: 999-1004, 2019.

[6] S. Kallepelli, S. Maheshwaram, "A novel circular double-gate SOI MOSFET with raised source/drain," Semicond. Sci. Technol., 36(6): 065009, 2021.

[7] N. A. Srivastava, A. Priya, R. A. Mishra, "Analog and radio-frequency performance of nanoscale SOI MOSFET for RFIC based communication systems," Microelectron. J., 98: 104731, 2020.

[8] T. Mikolajick et al., "Reconfigurable field effect transistors: A technology enablers perspective," Solid-State Electron., 194: 108381, 2022.

[9] C. Navarro, S. Barraud, S. Martinie, J. Lacord, M.-A. Jaud, M. Vinet, "Reconfigurable field effect transistor for advanced CMOS: Advantages and limitations," Solid-State Electron., 128: 155-162, 2017.

[10] A. Fuchsberger et al., "Reconfigurable field-effect transistor technology via heterogeneous integration of SiGe with crystalline Al contacts," Adv. Electron. Mater., 9(6): 2201259, 2023.

[11] S. Larentis et al., "Reconfigurable complementary monolayer MoTe2 field-effect transistors for integrated circuits," ACS Nano, 11(5): 4832-4839, 2017.

[12] C. Roemer et al., "Physics-based dc compact modeling of schottky barrier and reconfigurable field-effect transistors," IEEE J. Electron Devices Soc., 10: 416-423, 2021.

[13] F. Fuchs, S. Gemming, J. Schuster, "Electron transport through NiSi2--Si contacts and their role in reconfigurable field-effect transistors," J. Phys.: Condens. Matter, 31(35): 355002, 2019.

[14] J. Zhang, P.-E. Gaillardon, G. De Micheli, "Dual-threshold-voltage configurable circuits with three-independent-gate silicon nanowire FETs," in Proc. 2013 IEEE International Symposium on Circuits and Systems (ISCAS): 2111-2114, 2013.

[15] W. Fei, J. Trommer, M. C. Lemme, T. Mikolajick, A. Heinzig, "Emerging reconfigurable electronic devices based on two-dimensional materials: A review," InfoMat, 4(10): e12355, 2022.

[16] R. Ranjith, R. Jayachandran, K. J. Suja, R. S. Komaragiri, "Two dimensional analytical model for a reconfigurable field effect transistor," Superlattices Microstruct., 114: 62-74, 2018.

[17] X. Li, Y. Sun, X. Li, Y. Shi, Z. Liu, "Electronic assessment of novel arch-shaped asymmetrical reconfigurable field-effect transistor," IEEE Trans. Electron Devices, 67(4): 1894-1901, 2020.

[18] Y. Yao, Y. Sun, X. Li, Y. Shi, Z. Liu, "Novel reconfigurable field-effect transistor with asymmetric spacer engineering at drain side," IEEE Trans. Electron Devices, 67(2): 751-757, 2020.

[19] G. Galderisi, T. Mikolajick, J. Trommer, "Reconfigurable field effect transistors design solutions for delay-invariant logic gates," IEEE Embedded Sys. Lett., 14(2): 107-110, 2022.

[20] J. H. Bae et al., "Reconfigurable field-effect transistor as a synaptic device for XNOR binary neural network," IEEE Electron Device Lett., 40(4): 624-627, 2019.

[21] X. Jin, S. Zhang, X. Liu, "A dual doping nonvolatile reconfigurable FET," Sci. Rep., 13(1): 5634, 2023.

[22] R. Zhang et al., "Novel 3-D fin-RFET with dual-doped source/drain to improve ON-state current," IEEE Trans. Electron Devices, 69(12): 6569-6575, 2022.

[23] X. Jin, X. Yuan, S. Zhang, M. Li, X. Liu, "Complementary doped source-based reconfigurable Schottky diode as an equivalence logic gate," ACS Omega, 8(25): 23120-23129, 2023.

[24] C. Navarro et al., "Performance of FDSOI double-gate dual-doped reconfigurable FETs," Solid-State Electron., 194: 108336, 2022.

[25] C. Navarro et al., "3D-TCAD benchmark of two-gate dual-doped Reconfigurable FETs on FDSOI28 technology," Solid-State Electron., 200: 108577, 2023.

[26] L. Sa, G. F. R. Sa, J. Sa, J. Cb, A. J. La, "Challenges and opportunities in implementing negative differential resistance mode reconfigurable field effect transistors," arXiv preprint arXiv:2312.08351, 2023.

[27] B. Lu et al., "A novel nanosheet reconfigurable field effect transistor with dual-doped source/drain," Microelectron. J., 147: 106178, 2024.

[28] ATLAS User Manual, Santa Clara, USA: Silvaco International, 2015.

[29] K. M. Liu C. P. Cheng, "Investigation on the effects of gate-source overlap/underlap and source doping gradient of n-type Si cylindrical gate-all-around tunnel field-effect transistors," IEEE Trans. Nanotechnol., 19: 382-389, 2020.

[30] Y. J. Chung, K. A. V. Rosales, K. W. Baldwin, K. W. West, M. Shayegan, L. N. Pfeiffer, "Working principles of doping-well structures for high-mobility two-dimensional electron systems," Phys. Rev. Mater., 4(4): 044003, 2020.

[31] N. Pandey, Y. S. Chauhan, "Analytical modeling of short-channel effects in MFIS negative-capacitance FET including quantum confinement effects," IEEE Trans. Electron Devices, 67(11): 4757-4764, 2020.

[32] S. Liu et al., "Performance limit of gate-all-around si nanowire field-effect transistors: An Ab initio quantum transport simulation," Phys. Rev. Appl., 18(5): 054089, 2022.

[33] R. Quhe et al., "Sub-10 nm two-dimensional transistors: Theory and experiment," Phys. Rep., 938: 1-72, 2021.

Biographies

Zahra Ahangari received her Ph.D. degree in Electrical Engineering-Electronics from Islamic Azad University, Science and Research Branch, in the year 2013. From the year 2004 onwards, Zahra served as a research assistant in the device modeling and simulation laboratory of University of Tehran under supervision of Professor Morteza Fathipour. Currently, she is the faculty member of Electrical Engineering department at Islamic Azad University, Yadegar-e-Imam Khomeini (RAH) Shahr-e-Rey Branch. Zahra's main areas of research are modeling and simulation of nanoelectronic devices, nano sensors, and solar cells.

- Email: z.ahangari@gmail.com
- ORCID: 0000-0002-1095-3150
- Web of Science Researcher ID: AAP-5468-2021
- Scopus Author ID: 26435067300
- Homepage: https://scholar.google.com/citations?user=NuvC9YMAAAAJ&hl=en

How to cite this paper:
Z. Ahangari, "Enhancing high-performance computing: A comprehensive study on dual-doped source/drain reconfigurable field effect transistor," J. Electr. Comput. Eng. Innovations, 12(2): 475-484, 2024.

DOI: 10.22061/jecei.2024.10757.732

URL: https://jecei.sru.ac.ir/article_2137.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Short-term Prediction of Bitcoin Price Based on Generative Adversarial Network

M. Moosakhani, A. Jahangard Rafsanjani*, S. Zarifzadeh

Department of Computer Engineering, Faculty of Computer Engineering, Yazd University, Yazd, Iran.

Article Info

Article History:
Received 19 March 2024
Reviewed 26 April 2024
Revised 18 June 2024
Accepted 29 June 2024

Keywords:
Generative adversarial network
Short-term
Price prediction
Cryptocurrency
Bitcoin

*Corresponding Author's Email Address: *jahangard@yazd.ac.ir*

Abstract

Background and Objectives: Investment has become a paramount concern for various individuals, particularly investors, in today's financial landscape. Cryptocurrencies, encompassing various types, hold a unique position among investors, with Bitcoin being the most prominent. Additionally, Bitcoin serves as the foundation for some other cryptocurrencies. Given the critical nature of investment decisions, diverse methods have been employed, ranging from traditional statistical approaches to machine learning and deep learning techniques. However, among these methods, the Generative Adversarial Network (GAN) model has not been utilized in the cryptocurrency market. This article aims to explore the applicability of the GAN model for predicting short-term Bitcoin prices.

Methods: In this article, we employ the GAN model to predict short-term Bitcoin prices. Moreover, Data for this study has been collected from a diverse set of sources, including technical data, fundamental data, technical indicators, as well as additional data such as the number of tweets and Google Trends. In this research, we also evaluate the model's accuracy using the RMSE, MAE and MAPE metrics.

Results: The results obtained from the experiments indicate that the GAN model can be effectively utilized in the cryptocurrency market for short-term price prediction.

Conclusion: In conclusion, the results of this study suggest that the GAN model exhibits promise in predicting short-term prices in the cryptocurrency market, affirming its potential utility within this domain. These insights can provide investors and analysts with enhanced knowledge for making more informed investment decisions, while also paving the way for comparative analyses against alternative models operating in this dynamic field.

Introduction

The unpredictable nature of Bitcoin prices presents a specific challenge for investors making decisions in the cryptocurrency market [1]. Anticipating short-term in Bitcoin's price is crucial for investors, providing a perspective on whether the value of Bitcoin is poised to rise or fall [2]. This foresight is valuable for guiding strategic investment decisions within cryptocurrency investing.

In the field of cryptocurrency prediction, utilizing deep learning methods, especially neural networks, has shown promise in improving forecasting accuracy. Deep learning models utilize multiple layers of artificial neural networks to independently extract intricate patterns and dependencies within extensive datasets. This capability makes them proficient at capturing the inherent

complexities of Bitcoin price dynamics. The use of deep learning in predicting Bitcoin not only enhances predictive capabilities but also allows adaptation to changing market conditions. This establishes Deep learning as a valuable tool in the quest for more accurate and agile forecasting models within the dynamic landscape of cryptocurrency markets [3].

Exploring GANs for short-term prediction of Bitcoin prices is an uncharted domain, despite the diverse methods employed in predicting financial market changes [4]. This research is driven by recognizing the absence of exploration in tackling this specific issue and acknowledging the potential of GANs. GANs have great potential to predict Bitcoin prices because they have proven to be effective in forecasting stock prices in traditional financial markets [5]-[7].

In the field of Bitcoin prediction, GANs have emerged as an innovative and cutting-edge approach. Proposed by Ian Goodfellow and colleagues, GANs consist of two neural networks—the generator and the discriminator—engaged in a competitive yet collaborative learning process [8]. The generator's objective is to create synthetic data resembling authentic Bitcoin price patterns, while the discriminator strives to differentiate between genuine and generated data [9]. Through adversarial training, this process facilitates the development of a generator capable of producing data indistinguishable from actual market behavior [10]. GANs offer a unique advantage in Bitcoin prediction by capturing complex temporal dependencies within prices and generating realistic future scenarios. GANs ability to create realistic data and understand complex patterns adds a new and dynamic aspect to forecasting in the cryptocurrency market [5].

This research aims to introduce a novel approach for short-term Bitcoin price prediction. While the use of GANs in financial prediction is not unprecedented and has been used in the stock market, this study pioneers a unique application by concentrating on the cryptocurrency domain, an area that has not been explored with such approaches. The lack of specific modifications to the GAN model underscores its adaptability to the distinct characteristics of cryptocurrency data.

Differing from conventional prediction models, our proposed approach enhances its predictive capacity by tapping into a wide range of data sources. This comprehensive approach encompasses fundamental data, technical data, technical indicators, and supplementary information, including tweet counts and Google trends. By integrating these diverse sources, our model aims for a comprehensive understanding of the factors shaping cryptocurrency market dynamics, striving for a detailed prediction strategy.

The main goal of this paper is to advance the knowledge in cryptocurrency prediction by introducing a method that can be compared to existing approaches. Moving forward, the proposed model creates opportunities for benchmarking against traditional methods and enables potential refinements to improve its predictive accuracy. Through this research, we seek to develop a deeper comprehension of the dynamics of Bitcoin prices, offering a valuable tool for investors navigating the intricate landscape of cryptocurrency markets.

In this study, we applied the Generative Adversarial Network (GAN) model to forecast short-term Bitcoin prices in the cryptocurrency market. The GAN architecture comprised a Generator using Gated Recurrent Units (GRU) for synthetic data generation and a Discriminator with a one-dimensional Convolutional Neural Network (CNN-1D) for distinguishing real from generated data. We curated a dataset with 738 Bitcoin-related features, applying the Gray Wolf algorithm for feature selection. The GAN model, trained on 70% of the dataset, demonstrated strong predictive capabilities, validated by low Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) values. This affirms the GAN model's effectiveness in short-term Bitcoin price prediction within the cryptocurrency market.

The structure of the continuation of this paper is as follows: In the second section, we will provide a review of relevant studies. The third section will elaborate on the methodology employed in this research. In the fourth section, we will describe the data used in this study and the operations conducted on them. In the subsequent section, we present the obtained results and the accuracy of the model. Finally, we conclude and outline the sources utilized.

Literature Review

So far, numerous works have been conducted in the field of cryptocurrency price prediction using various methods and models [11]-[14]. Some researchers have employed machine learning or deep learning approaches for this prediction, while others have utilized sentiment analysis methods in this domain. Additionally, a group of researchers has undertaken such predictions using traditional statistical methods. In the following, we will examine some examples of the works carried out in this area.

Alonso-Monsalve et al. investigated the applicability of Convolutional Neural Network (CNN) as an alternative to the traditional Multilayer Perceptron (MLP) in the realm of classification, specifically focusing on the cryptocurrency market trends through high-frequency technical analysis [15]. They gathered and processed data for six widely used cryptocurrencies (Bitcoin, Dash,

Ethereum, Litecoin, Monero, and Ripple). The analysis, based on one-minute exchange rates, spanned a three-month period (July 1, 2018, to September 30, 2018). Utilizing 18 technical indicators, they evaluated the model's performance using the Sensitivity metric. Their results indicated that, overall, CNN exhibited superior performance compared to MLP, particularly for Bitcoin and Ethereum, showing promising outcomes. Litecoin, Dash, and Ripple also demonstrated incremental improvements, while Monero did not achieve satisfactory results.

Oyedele et al. explored the feasibility of predicting closing prices for various cryptocurrencies (Bitcoin, Ethereum, Litecoin, Binance Coin, Dogecoin, and Stellar) using three deep learning techniques (CNN, GRU, and DFNN), along with three tree-based techniques (Adaptive Boosting, GBM, and XGB) [16]. The study employed VIF for feature selection. Prediction accuracy was assessed using datasets from multiple sources, including Yahoo Finance (January 1, 2018, to December 31, 2021), UK Investing (July 1, 2021, to March 2, 2022), and Bitfinex (January 1, 2021, to July 6, 2021). The datasets encompassed five closing price features, such as closing price, highest price, lowest price, opening price, and daily volume for each cryptocurrency, along with weighted average and two technical indicators, SMA, and EMA. Evaluation metrics included NSE, EVS, t-test, and MAPE. The researchers concluded that, overall, predictions made by deep learning models, especially Convolutional Neural Network models, outperformed tree-based models.

Nakano et al. investigated Bitcoin investment using an Artificial Neural Network (ANN) that extracted trading signals from past time series data every 15 minutes, enabling the prediction of Bitcoin prices in the next 15 minutes [17]. In this study, they explored an artificial neural network model with varying numbers of layers, different activation functions, multiple classification types (3 classes, 4 classes, and 5 classes), and different inputs (to assess the impact of technical indicators). The authors collected Bitcoin price data, including closing price, highest price, lowest price, and volume, from the cryptocurrency exchange Poloniex at 15-minute intervals spanning from July 31, 2016, at 15:00 (GMT) to January 21, 2018, at 7:30 (GMT). They utilized cumulative returns for model performance evaluation. Ultimately, they concluded that predictive performance improves with an increase in the number of layers, the use of the Leaky ReLU activation function, a reduction in the number of classes in classification, and the incorporation of technical indicators as inputs.

Hansun et al. employed three methods—LSTM, Bi-LSTM, and GRU—to predict the next day's prices for several cryptocurrencies, including Bitcoin, Ethereum, Binance Coin, Cardano, and Tether [18]. The data used in this study comprised the maximum daily data available for each cryptocurrency collected from Yahoo Finance. Each dataset included historical features such as Open, Close, High, Low, and Volume for the respective cryptocurrency. They innovatively utilized a multi-variable approach and compared three models: LSTM, Bi-LSTM, and GRU. Performance evaluation metrics included MAE, MAPE, and RMSE. They iteratively developed and evaluated each deep learning network (LSTM, Bi-LSTM, and GRU) ten times and concluded that GRU emerged as the preferred deep learning method among the three considered.

Serafini et al. investigated the predictive power of sentiment analysis on the network, exploring statistical and deep learning methods for forecasting the future price of Bitcoin [19]. They analyzed financial features and sentiments extracted from economic data and crowd-sourced data, demonstrating how sentiments play a crucial role in predicting market movements for Bitcoin. The data used in this study were collected from April 2017 to October 2019, spanning a total of 944 days. Each sample in the dataset comprised daily Bitcoin volume features, weighted Bitcoin price, Bitcoin tweet sentiments, and Bitcoin tweet volume. The authors compared two models, ARIMAX and LSTM (a type of RNN), for predicting Bitcoin prices. They used the mean squared error as their evaluation metric and found that both models achieved optimal results in new predictions due to the inclusion of sentiment features (with a mean squared error of less than 0.14%). However, they noted that the ARIMAX model outperformed the LSTM, with a mean squared error of 0.00030187 in new predictions.

Zou et al. proposed a multi-faceted model for predicting extreme prices in Bitcoin [20]. This model takes input from related assets (Ethereum and gold), technical indicators, and Twitter content. The study aimed to investigate whether social media discussions among the public about Bitcoin have predictive power for significant price changes. For this purpose, a dataset of 5,000 daily tweets (containing the keyword "Bitcoin") from 2015 to 2021 was collected, amounting to a total of 9,435,437 tweets in the study. A new dataset included tweets, candlestick data, prices of related assets (Ethereum and gold), and a set of technical indicators collected from January 1, 2015, to May 31, 2021. In the presented model, FinBERT embeddings were used to represent the complete content of tweets at the sentence level. These embeddings were then combined with a convolutional neural network to create a predictive model for significant market movements. The goal of the study was to optimally utilize the content of social media beyond sentiment scores. To evaluate the performance of their model and demonstrate its practical application, the

authors proposed a simple but long trading strategy and reported Backtesting results. They stated that the superior performance of the combined model presented in the paper is confirmed with a threshold of 0.95 in risk-adjusted measures such as the Sortino ratio and maximum drawdown.

Methodology

Certainly, the domain of cryptocurrency prediction has seen the rise of deep learning approaches, particularly neural networks, as a promising pathway for improving forecasting accuracy. Deep learning models, utilizing multiple layers of artificial neural networks, can independently identify complex patterns and relationships within large datasets. This feature makes them effective in grasping the inherent complexities of Bitcoin price dynamics [21].

Indeed, the use of deep learning in Bitcoin prediction not only enhances predictive capabilities but also enables adaptation to evolving market conditions. This positions deep learning as a valuable tool in the pursuit of more accurate and responsive forecasting models within the dynamic landscape of cryptocurrency markets [22].

In this paper, we have delved into short-term price prediction of Bitcoin using the GAN model, a methodology not yet explored in the cryptocurrency domain.

The aim of this article is to investigate the usability of the GAN model in the cryptocurrency market.

The GAN (Generative Adversarial Network) is a type of artificial neural network architecture consisting of two main components: the Generator and the Discriminator. These components engage in a competitive yet collaborative learning process. The Generator is responsible for predicting future Bitcoin prices using historical Bitcoin data it receives as input. These predictions should resemble real price patterns of Bitcoin. Subsequently, this generated data is presented as input to the Discriminator. The Discriminator is responsible distinguishing between real data and data generated by the Generator, and endeavors to differentiate between actual and generated data. Throughout the training process, the Generator and Discriminator networks iteratively improve in a reciprocal manner. In essence, the Generator attempts to produce data that the Discriminator cannot distinguish from real data, while the Discriminator strives to always provide the best possible discrimination. This adversarial training process enhances the quality of the data generated by the Generator [23], [24].

Each of the two components, Generator and Discriminator, is itself a neural network. In this study, we utilized GRU as the Generator and CNN-1D as the Discriminator. The architecture of the GAN model used in this research is show in Fig. 1.

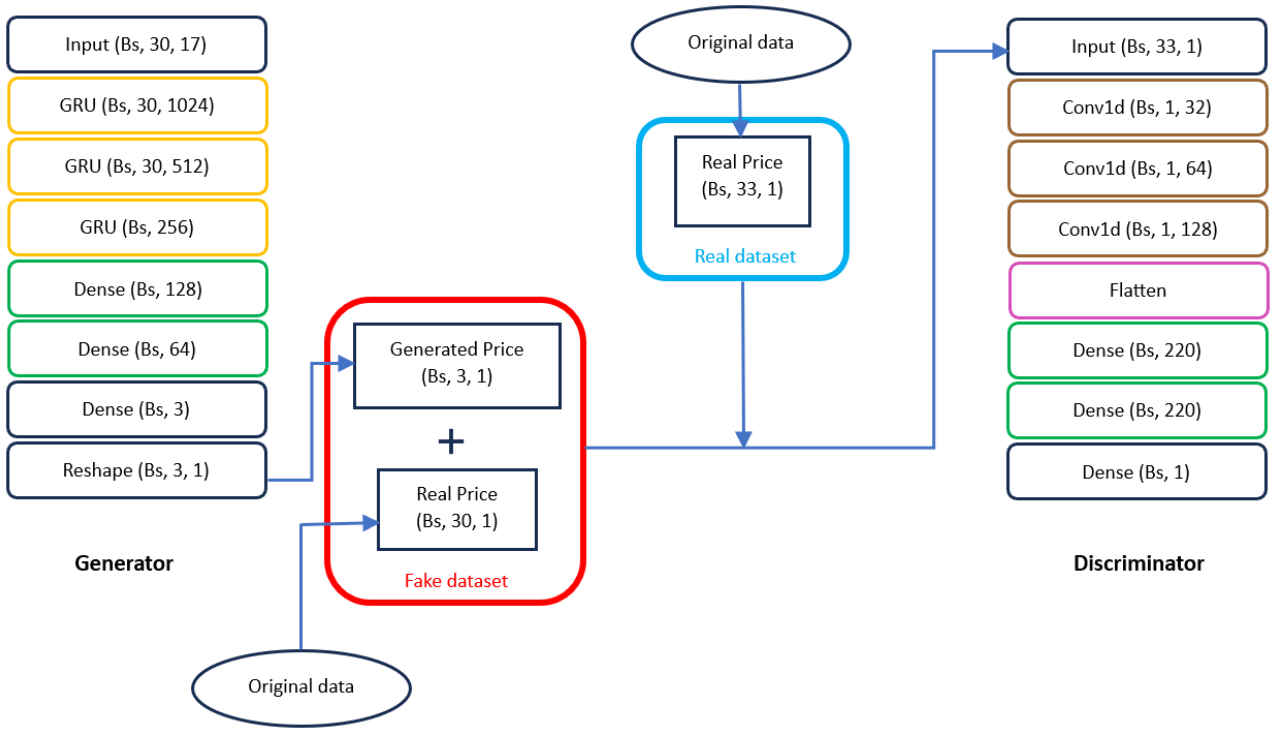


Fig. 1: GAN Architecture.

As illustrated in Fig. 1, historical Bitcoin data, which represents real values, is given to the Generator as input. In this study, GRU is utilized as the Generator. The details related to the data are thoroughly explained in the "Data Collection and Preprocessing" section. Then, the GRU, considering historical Bitcoin data and learning patterns related to them, predicts future values of Bitcoin prices. We combined values generated by the Generator with the real Bitcoin prices, and the result was considered as input for the discriminator. This combination process leads to an increase in the length of the data and, consequently, enhances the discriminator's accuracy in learning classification.

In this study, we train the Generator to minimize its objective function, defined as (1) [25].

$$Minimize \log(1 - D(G(z))) \quad (1)$$

Additionally, the objective of the Discriminator is to maximize the probability of assigning correct labels to samples, and its objective function is defined according to (2) [25].

$$Maximize \log(D(x)) + \log(1 - D(G(z))) \quad (2)$$

where z is the input data for generator, x is the target from the real dataset, G(z) is the generated data by the generator.

This study employs cross-entropy for calculating the loss for both Generator and Discriminator in the presented GAN model structure. The loss function for the Generator is defined according to Equation 3, and for the Discriminator, it is defined according to (3) [25].

$$-\log(1 - D(G(z))) \quad (3)$$

$$-\log(D(x)) - \log(1 - D(G(z))) \quad (4)$$

whatever loss function value minimizes during the training process, the better results obtained.

Generally, in the context of (1) to (4), where z represents the input data for the generator, x denotes the target from the real dataset, and G(z) indicates the data generated by the generator. Subsequently, this generated data serves as input for the discriminator, denoted by D(G(z)). In fact, G and D represent the generator and the discriminator respectively.

Every model of machine learning algorithms has a specific set of parameters, such as the number of layers, learning rate, the number of neurons, and some other parameters. These parameters need to be defined through the training process for achieving the highest performance score of the model [26].

In this study, Bayesian Optimization has been employed in the training process that utilize Bayes' Theorem to finding parameters that can create minimize or maximize the given objective function score.

Hyperparameters Tuning

This section outlines the specific configuration of key hyperparameters in the experimental design of the study, shedding light on the choices made to fine-tune the Generative Adversarial Network (GAN) model for predicting short-term Bitcoin prices.

1. Learning Rate Range (Hyperparameter Tuning): The learning rate, a critical hyperparameter in training neural networks, was meticulously chosen within the range of 0.00003 to 0.00016. This deliberate range selection reflects a systematic exploration of different learning rates. The learning rate is pivotal in determining the step size during optimization, influencing how quickly or slowly the model adapts to the dataset. The specified range indicates a nuanced approach to finding an optimal balance between convergence speed and model stability.

2. Number of Epochs (Training Iterations): To capture the dynamics of the learning process, the number of epochs, or complete passes through the dataset, was varied between 150 and 300. This parameter is crucial in gauging the model's exposure to the dataset and its learning capacity over time. The range chosen suggests a comprehensive investigation into the model's performance across different durations of training, offering insights into convergence patterns and potential trade-offs between underfitting and overfitting.

3. Batch Size (Data Processing Units): The batch size, set at 128, determines the number of data samples processed in each iteration during training. This parameter balances computational efficiency and model generalization. A batch size of 128 indicates a pragmatic choice, as it is commonly used in practice. It strikes a balance between leveraging computational efficiency, especially in parallel processing, and facilitating the model's ability to generalize patterns from the data.

By explicitly detailing these hyperparameters, the section not only provides transparency regarding the experimental setup but also offers valuable information for reproducibility and comparison with future studies. The meticulous selection of these parameters reflects a thoughtful approach to optimizing the GAN model for accurate short-term Bitcoin price prediction [27].

Data Collection and Preprocessing

The required data for this research was collected from the website https://bitinfocharts.com [28] using a web scraper written in Python 3.11.4. The collected data consists of 738 features related to the Bitcoin cryptocurrency that overall encompass technical data, fundamental data, technical indicators, and additional data such as the number of tweets and Google trends. The data has been collected daily in the time range from

October 1, 2014, to October 14, 2023.

The study employs a set of technical indicators to enhance the analysis of Bitcoin (BTC) price time series features. These indicators encompass diverse metrics such as Simple Moving Average (SMA), Exponential Moving Average (EMA), Relative Strength Index (RSI), Weighted Moving Average (WMA), Standard Deviation (STD), Variance (VAR), Triple Moving Exponential (TRIX), and Rate of Change (ROC). Computed across various periods, including end-of-day, 7, 30, and 90 days, these technical indicators provide nuanced insights into BTC price dynamics.

The raw features, constituting the foundation for these indicators and presented in Table 1, are derived from end-of-day closing prices, serving as fundamental reference values. These indicators transcend the inherent characteristics of raw features, unveiling intricate properties such as variances and standard deviations over time.

Specifically, they illuminate the relationship between BTC price and critical factors like the standard deviation of transactions or hashrate over 30-day periods. This nuanced approach moves beyond raw transaction and hashrate data, offering a refined understanding of BTC price dynamics through the perspective of calculated technical indicators.

Table 1: Raw features from which the technical indicators are created

Features	Discription
Transactions	The number of sent and received Bitcoin payments
Block size	Transactional information is cryptographically linked within the blockchain, with the maximum block size presently established at 1 megabyte
Sent from addresses	These are distinct Bitcoin addresses from which payments are made everyday
Difficulty	The daily average mining difficulty. The difficulty is computed by the network after a specified number of blocks have been created so that the time it requires to mine a block remains around 10 min
Average transaction value	The average value of the transactions in Bitcoin
Mining profitability	The profitability in USD/day for 1 terahash per second (THash/s)
Sent BTC	The total Bitcoins sent daily
Fee-to-reward ratio	The ratio of the fee sent in a transaction to the reward for verifying that transaction by the other users
Median transaction fee	The median of transaction fees in Bitcoin
Average transaction fee	In each transaction, the sender can include a transaction fee, and this fee is received by the miners who verify the transaction. Transactions offering higher fees serve as incentives for Bitcoin miners to prioritize and process them more promptly compared to transactions with lower fees
Block time	The time required to mine one block. Usually, it is around 10 min but can fluctuate depending on the hashrate of the network
Hashrate	The total daily computational capacity of the Bitcoin network, known as hashrate, signifies the speed at which a computer can perform operations
Median transaction value	The median value of the transactions in Bitcoin
Active addresses	The number of unique addresses participating in a transaction by either sending or receiving Bitcoins
Top 100 to total	The ratio of Bitcoins stored in the top 100 accounts to all the other accounts of Bitcoin

Once the data was collected, and obtaining clean and processable data, we added a label column to the data as a classification format (0, 1). In this format, 0 indicates a decrease in price compared to the previous day, and 1 indicates an increase in price compared to the previous day.

Next, we utilized the Gray Wolf optimization algorithm as a feature selection algorithm [29]. This led to identifying 17 influential features out of the 738 available features for predicting Bitcoin prices.

Table 2 displays 17 features obtained from running the Gray Wolf algorithm, along with some associated values.

Table 2: Selected features of Gray Wolf algorithm

Selected features by Gray Wolf algorithm
bitcoin-transactions
mediantransactionvalue-btc
transactions-btc-ema-7
transactions-btc-wma-7
size-btc-rsi-7
size-btc-std-14
transactionvalue-btc-trx-14
fee-to-reward-btc-roc-3
confirmationtime-btc-var-30
hashrate-btc-ema-14
hashrate-btc-trx-3
hashrate-btc-mom-7
mediantransactionvalue-btc-var-90
mediantransactionvalue-btc-mom-7
top100cap-btc-rsi-14
price-btc-ema-14
price-btc-rsi-90

As evident from Table 2, the features obtained from running the Gray Wolf algorithm, selected as the most influential features among the collected 738 features, are listed in Table 2.

After running the Gray Wolf algorithm and obtaining the influential features, we divided the data into a 70-30 ratio for training and testing, respectively. In fact, we have utilized 70% of the data for model training and the remaining 30% for testing the model.

This model employs a method for preparing the dataset for supervised learning by dividing it with a rolling window set to 1. Fig. 2 provides an illustration of this process. The original dataset, which is two-dimensional, undergoes a reshaping operation to transform it into three dimensions based on the specified timesteps.

Fig. 3 presents the output of the dataset generated by the generator, where the number of output units is set to 1. In our model, adjustments to the time step can be made in Fig. 2, while modifications to the output step can be implemented in Fig. 3. This paper has constructed a many-to-many model with a timestep of 30 and an output step of 3.

The time window used in this study is structured in a way that it leverages the past 30 days of Bitcoin data to predict its price for the next 3 days.

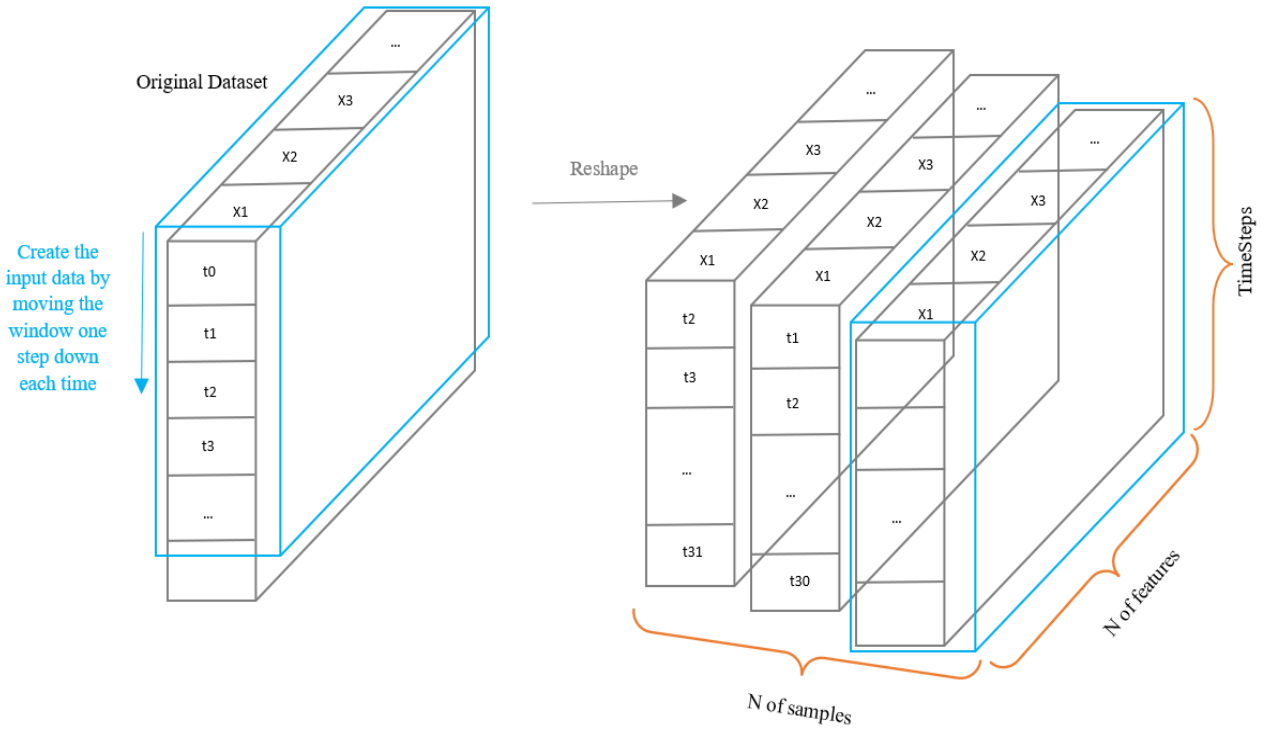


Fig. 2: Input data.

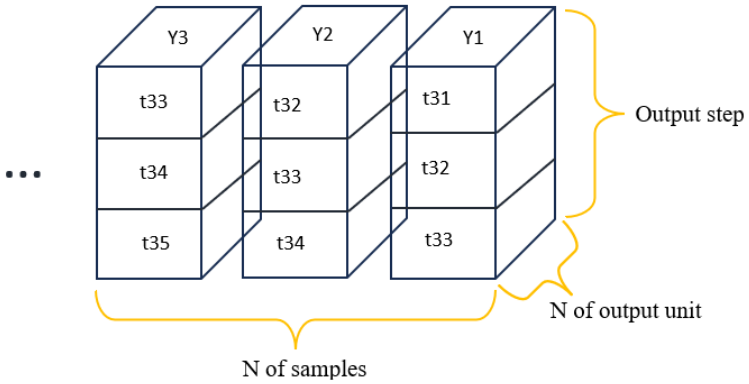


Fig. 3: Output data.

Evaluation and Result

In this paper, for evaluating the performance of the presented GAN model, we have employed the RMSE, MAE and MAPE metrics, defined by (5)-(7) respectively [30].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{N}} \quad (5)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{n}| x_i - \hat{x}_i | \quad (6)$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{n}| \frac{x_i - \hat{x}_i}{x_i} | \quad (7)$$

In (5)-(7), N represents the number of data points, x_i denotes the actual Bitcoin price, and $\hat{x}_i$ represents the predicted Bitcoin price.

Fig. 4 and Fig. 5 illustrate the model's performance accuracy during the training and testing processes, respectively.

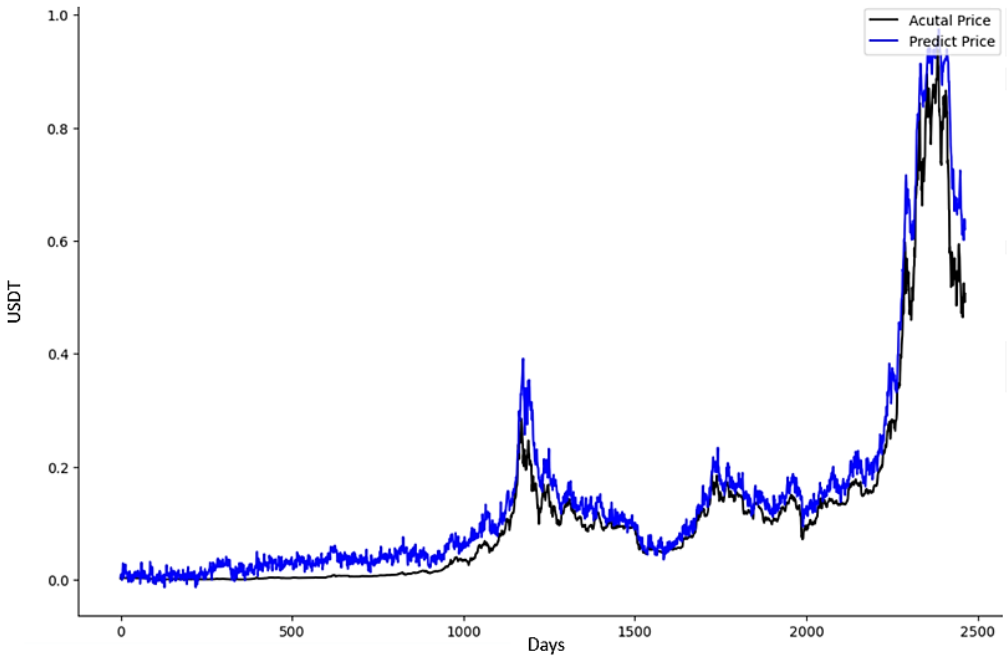


Fig. 4: The performance of the GAN model during the training phase.

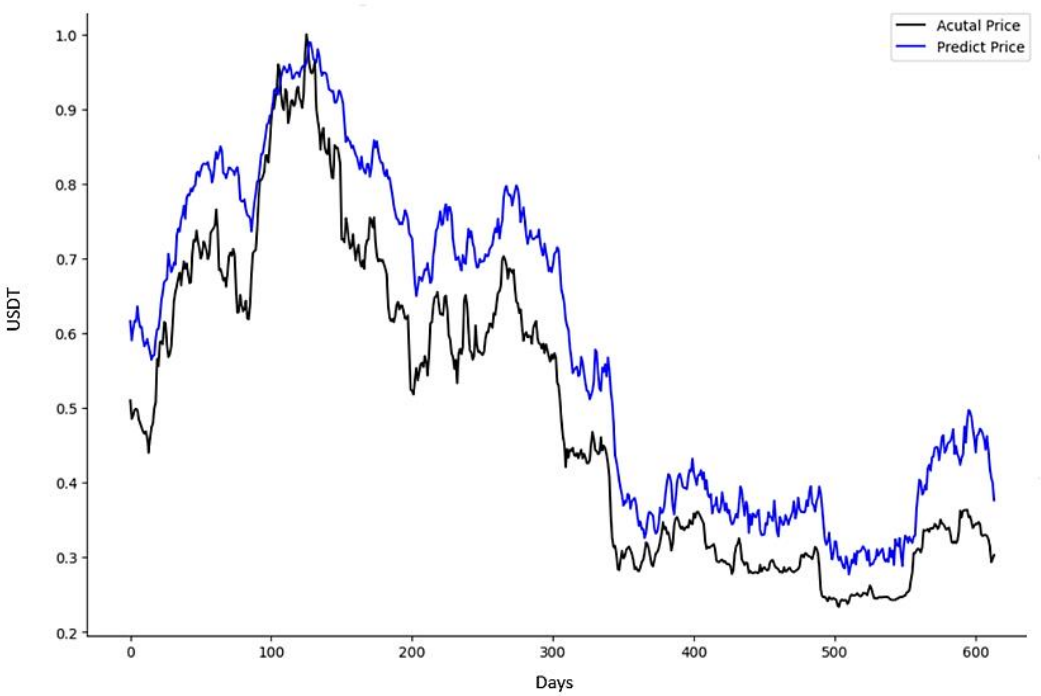


Fig. 5: The performance of the GAN model during the testing phase.

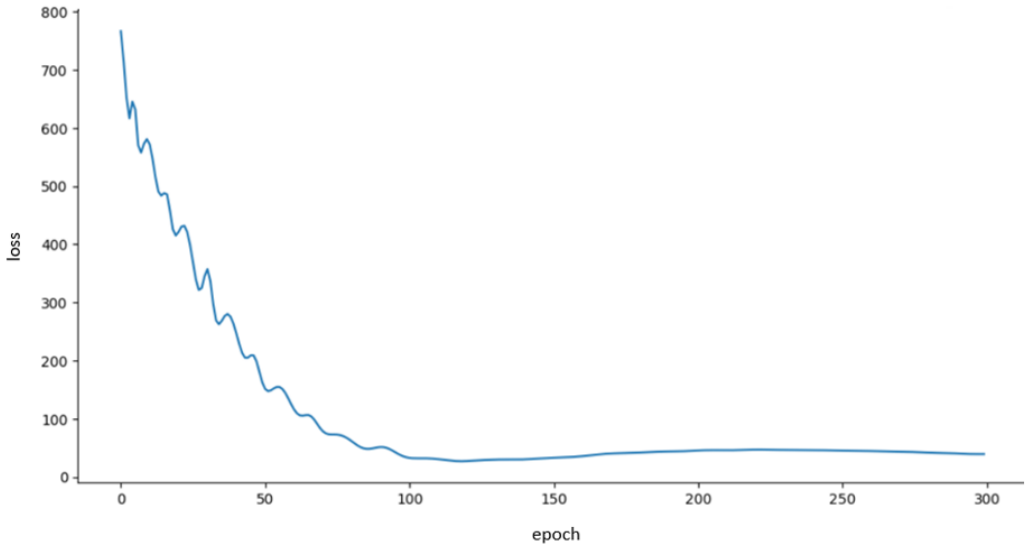


Fig. 6: GAN Loss Plot.

Fig. 4 illustrates the performance of the GAN model during the training process, and Fig. 5 displays its performance in the testing phase. The horizontal lines in Fig. 4 and Fig. 5 represent days, while the vertical lines denote the Bitcoin prices on each respective day. The black line in Fig. 4 and Fig. 5 represents the actual Bitcoin prices for each day, while the blue line depicts the predicted prices for Bitcoin on each day.

The RMSE, MAE and MAPE values obtained are also presented in Table 3, Table 4, and Table 5 respectively.

As indicated in Tables 3-5, these values indicate that overall, the GAN model has demonstrated good performance in predicting short-term Bitcoin prices, suggesting its potential utility in the cryptocurrency market. Fig. 6 shows the trend of loss reduction in this study.

As Fig. 6 illustrates, the loss value has significantly decreased from the beginning of the training process to around epoch 100, after which it has stabilized with minor fluctuations.

Table 3: RMSE values in GAN model

Stage	RMSE
Training	0.046
Testing	0.099

Table 4: MAE values in GAN model

Stage	MAE
Training	0.033
Testing	0.077

Table 5: MAPE values in GAN model

Stage	MAPE
Training	33.52%
Testing	75.24%

Conclusion and Future Suggestions

In this research, our primary focus was on examining the viability of implementing the Generative Adversarial Network (GAN) model within the cryptocurrency market. Specifically, we employed the GAN model to forecast short-term Bitcoin prices, utilizing a neural network architecture comprising two essential components: Generator and Discriminator, both functioning as distinct artificial neural networks. The Generator, responsible for producing synthetic data, was constructed using Gated Recurrent Units (GRU), while the Discriminator, designed to distinguish between real and generated data, employed a one-dimensional Convolutional Neural Network (CNN-1D). To conduct a thorough investigation, we amassed a dataset containing 738 features related to Bitcoin, spanning technical, fundamental, and additional data such as tweet counts and Google trends, collected daily from 01/10/2014 to 14/10/2023.

To pinpoint the most influential features within this extensive dataset, we applied the Gray Wolf algorithm. Subsequently, we partitioned the dataset, allocating 70% for training the GAN model and reserving the remaining 30% for assessing its predictive capabilities. The model's performance was then quantified using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics. The experimental results, underscored by low RMSE values (0.046 during training and 0.099 in testing), validate the GAN model's efficacy in short-term Bitcoin price prediction. This substantiates the model's potential utility within the cryptocurrency market, opening avenues for comparative analyses against alternative models operating in this dynamic domain.

Through comparisons, architectural explorations, hyperparameter optimization, and training process refinements, researchers can discover strategies to boost the GAN model's effectiveness in the dynamic and complex cryptocurrency market.

1. Model Comparison: The first recommendation is to compare the GAN model with a variety of other models such as GRU, CNN, LSTM, Bi-LSTM, ARIMAX, etc. This comparative analysis aims to provide a deeper understanding of the relative strengths and weaknesses of each model. By evaluating how well the GAN model performs in comparison to these alternatives, researchers can gain insights into which model may be more suitable for specific scenarios or market conditions.

2. Exploring Different Architectures: The paragraph suggests exploring different artificial neural networks as both Generator and Discriminator. This means considering alternative configurations or types of neural networks for these crucial components of the GAN model. By experimenting with different architectures, researchers can identify which combinations yield the most accurate and reliable predictions. This exploration could involve testing variations in network depth, width, or incorporating novel neural network structures.

3. Optimizing Hyperparameters: The success of machine learning models heavily relies on choosing appropriate hyperparameters. Therefore, the paragraph highlights the importance of fine-tuning hyperparameters for the GAN model. This involves systematically adjusting parameters like learning rates, batch sizes, and layer configurations to optimize the model's performance. Bayesian Optimization, as mentioned earlier, is one approach that can be employed for this purpose.

4. Refining Training Processes: Continuous refinement of the training process is crucial for improving model performance. Researchers may consider experimenting with different optimization algorithms, regularization techniques, or introducing advanced training strategies. Bayesian Optimization, mentioned earlier in the context of hyperparameter tuning, can also be applied to refine the training process.

Author Contributions

Everyone contributed equally

Conflict of Interest

The authors declare that the authors have no competing interests as defined by JECEI, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

References

[1] L. Kristoufek, "Will Bitcoin ever become less volatile?," Finance Res. Lett., 51: 103353, 2023.

[2] M. Shu, W. Zhu, "Real-time prediction of Bitcoin bubble crashes," Physica A, 548: 124477, 2020.

[3] S. E. Charandabi, K. Kamyar, "Prediction of cryptocurrency price index using artificial neural networks: a survey of the literature," Eur. J. Bus. Manage. Res., 6: 17-20, 2021.

[4] F. Eckerli, J. Osterrieder, "Generative adversarial networks in finance: An overview. arXiv 2021," arXiv preprint arXiv:2106.06364.

[5] K. Zhang, G. Zhong, J. Dong, S. Wang, Y. Wang, "Stock market prediction based on generative adversarial network," Procedia Comput. Sci., 147: 400-406, 2019.

[6] X. Zhou, Z. Pan, G. Hu, S. Tang, C. Zhao, "Stock market prediction on high-frequency data using generative adversarial nets," Math. Probl. Eng., 2018.

[7] H. Lin, C. Chen, G. Huang, A. Jafari, "Stock price prediction using generative adversarial networks," J. Comp. Sci, 17: 188-196, 2021.

[8] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, "Generative adversarial networks: An overview," IEEE Signal Process. Mag., 35: 53-65, 2018.

[9] A. Aggarwal, M. Mittal, G. Battineni, "Generative adversarial network: An overview of theory and applications," Int. J. Inf. Manag. Data Insights, 1: 100004, 2021.

[10] A. Staffini, "Stock price forecasting by a deep convolutional generative adversarial network," Front. Artif. Intell, 5, 2022.

[11] M. M. Patel, S. Tanwar, R. Gupta, N. Kumar, "A deep learning-based cryptocurrency price prediction scheme for financial institutions," J. Inf. Secur. Appl., 55: 102583, 2020.

[12] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, M. Alazab, "Stochastic neural networks for cryptocurrency price prediction," IEEE Access, 8: 82804-82818, 2020.

[13] G. Kim, D. H. Shin, J. G. Choi, S. Lim, "A deep learning-based cryptocurrency price prediction model that uses on-chain data," IEEE Access, 10: 56232-56248, 2022.

[14] S. Biswas, M. Pawar, S. Badole, N. Galande, S. Rathod, "Cryptocurrency price prediction using neural networks and deep learning," in Proc. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS): 408-413, 2021.

[15] S. Alonso-Monsalve, A. L. Suárez-Cetrulo, A. Cervantes, D. Quintana, "Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators," Expert Syst. Appl., 149: 113250, 2020.

[16] A. A. Oyedele, A. O. Ajayi, L. O. Oyedele, S. A. Bello, K. O. Jimoh, "Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction," Expert Syst. Appl., 213: 119233, 2023.

[17] M. Nakano, A. Takahashi, S. Takahashi, "Bitcoin technical trading with artificial neural network," Physica A, 510: 587-609, 2018.

[18] S. Hansun, A. Wicaksana, A. Q. Khaliq, "Multivariate cryptocurrency prediction: Comparative analysis of three recurrent neural networks approaches," J. Big Data, 9: 1-15, 2022.

[19] G. Serafini, P. Yi, Q. Zhang, M. Brambilla, J. Wang, Y. Hu et al., "Sentiment-driven price prediction of the bitcoin based on statistical and deep learning approaches," in Proc. International Joint Conference on Neural Networks (IJCNN): 1-8, 2020.

[20] Y. Zou, D. Herremans, "A multimodal model with Twitter FinBERT embeddings for extreme price movement prediction of Bitcoin," arXiv preprint arXiv:2206.00648, 2022.

[21] H. Mujlid, "A survey on machine learning approaches in Cryptocurrency: Challenges and Opportunities," in Proc. 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET): 1-6, 2023.

[22] G. Pv, B. Jackson, "Machine learning-based timeseries analysis for cryptocurrency price prediction: a systematic review and research," in Proc. 2023 International Conference on Networking and Communications (ICNWC): 1-5, 2023.

[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair et al., "Generative adversarial networks," Commun. ACM, 63: 139-144, 2020.

[24] L. Gonog, Y. Zhou, "A review: Generative adversarial networks," in Proc. 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA): 505-510, 2019.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair et al., "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014.

[26] S. Singh, D. K. Verma, K. Malik, "A review on machine learning approach for predicting cryptocurrency prices," Automation and Computation: 295-299, 2023.

[27] H. Berard, G. Gidel, A. Almahairi, P. Vincent, S. Lacoste-Julien, "A closer look at the optimization landscapes of generative adversarial networks," arXiv preprint arXiv:1906.04848, 2019.

[28] M. Mudassir, S. Bennbaia, D. Unal, M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: A machine learning approach," Neural Comput. Appl.: 1-15, 2020.

[29] H. Pan, S. Chen, H. Xiong, "A high-dimensional feature selection method based on modified Gray Wolf Optimization," Appl. Soft Comput., 135: 110031, 2023.

[30] A. De Myttenaere, B. Golden, B. Le Grand, F. Rossi, "Mean absolute percentage error for regression models," Neurocomput., 192: 38-48, 2016.

Biographies

Mahnaz Moosakhani received the B.Sc. degree in Software Engineering from Malayer University, Hamedan, Iran, in 2016, and the M.Sc. degree, also in Software Engineering, from Shiraz University, Shiraz, Iran, in 2019. She is currently a student in the Ph.D. program in the Department of Computer Engineering at Yazd University, Yazd, Iran.

- Email: m.moosakhani@stu.yazd.ac.ir
- ORCID: 0009-0009-7520-1724
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Amir Jahangard-Rafsanjani received the B.Sc. degree in Software Engineering from Shahid Beheshti University, Tehran, Iran in 2003, and the M.Sc. degree also in Software Engineering from Sharif University of Technology, Tehran, Iran in 2005. He also received the Ph.D. degree in Software Engineering from Sharif University of Technology in 2014. He is currently an Assistant Professor in the Department of Computer Engineering. His research interests include Database, Data and Text Mining and Software Testing.

- Email: jahangard@yazd.ac.ir
- ORCID: 0000-0003-2638-5722
- Web of Science Researcher ID: NA
- Scopus Author ID: 35173443500
- Homepage: https://yazd.ac.ir/people/jahangard

Sajjad Zarifzadeh is currently an Associate Professor in Yazd University, Iran. He got his Ph.D. from University of Tehran in 2012. His main research interests are on big data, data mining and applications of machine learning.

- Email: szarifzadeh@yazd.ac.ir
- ORCID: 0000-0001-5627-0542
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: https://yazd.ac.ir/people/zarifzadeh

How to cite this paper:
M. Moosakhani, A. Jahangard Rafsanjani, S. Zarifzadeh, "Short-term prediction of bitcoin price based on generative adversarial network," J. Electr. Comput. Eng. Innovations, 12(2): 485-496, 2024.

DOI: 10.22061/jecei.2024.10760.734

URL: https://jecei.sru.ac.ir/article_2138.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Hybrid Convolutional Neural Network with Domain adaptation for Sketch based Image Retrieval

A. Gheitasi, H. Farsi, S. Mohamadzadeh*

Department of Electrical Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

Article Info

Article History:
Received 22 March 2024
Reviewed 29 April 2024
Revised 17 June 2024
Accepted 29 June 2024

Keywords:
Sketch Based Image Retrieval (SBIR)
Hybrid CNN
domain adaptation
deep learning

*Corresponding Author's Email Address:
s.mohamadzadeh@birjand.ac.ir

Abstract

Background and Objectives: Freehand sketching is an easy-to-use but effective instrument for computer-human connection. Sketches are highly abstract to the domain gap, that exists between the intended sketch and real image. In addition to appearance information, it is believed that shape information is also very efficient in sketch recognition and retrieval.

Methods: In the realm of machine vision, comprehending Freehand Sketches has grown more crucial due to the widespread use of touchscreen devices. In addition to appearance information, it is believed that shape information is also very efficient in sketch recognition and retrieval. The majority of sketch recognition and retrieval methods utilize appearance information-based tactics. A hybrid network architecture comprising two networks—S-Net (Sketch Network) and A-Net (Appearance Network)—is shown in this article under the heading of hybrid convolution. These subnetworks, in turn, describe appearance and shape information. Conversely, a module known as the Conventional Correlation Analysis (CCA) technique module is utilized to match the range and enhance the sketch retrieval performance to decrease the range gap distance. Finally, sketch retrieval using the hybrid Convolutional Neural Network (CNN) and CCA domain adaptation module is tested using many datasets, including Sketchy, Tu-Berlin, and Flickr-15k. The final experimental results demonstrated that compared to more sophisticated methods, the hybrid CNN and CCA module produced high accuracy and results.

Results: The proposed method has been evaluated in the two fields of image classification and Sketch Based Image Retrieval (SBIR). The proposed hybrid convolution works better than other basic networks. It achieves a classification score of 84.44% for the TU-Berlin dataset and 82.76% for the sketchy dataset. Additionally, in SBIR, the proposed method stands out among methods based on deep learning, outperforming non-deep methods by a significant margin.

Conclusion: This research presented the hybrid convolutional framework, which is based on deep learning for pattern recognition. Compared to the best available methods, hybrid network convolution has increased recognition and retrieval accuracy by around 5%. It is an efficient and thorough method which demonstrated valid results in Sketch-based image classification and retrieval on TU-Berlin, Flickr 15k, and sketchy datasets.

Introduction

Simple Sketches are prevalent in daily life because, Freehand sketching is an easy-to-use but effective instrument for communication and computer-human

connection. Since, touch screens are so commonly used in portable electronics, finding simple Sketches has received more attention. Sketches typically depict location and shape, but actual photographs also include additional useful details like color and texture [1]. Compared to actual images, Sketches are considered as highly sparse signals, and since they are abstract and lack input data [2], their analysis might be difficult. As a result, comparing low-detail images with real photos with a high pixel density is challenging [3]. The following are some of the difficulties in this field:

1. While Sketches are highly abstract and have very little shape information, natural images provide a great deal of color and texture information.

2. On the other hand, as every individual displays an object's sketch uniquely, there will be a wide range of Sketch styles for an object.

3. The training data, may not include every query that might be asked.

4. Another significant obstacle in this sector is the domain gap, that exists between the intended sketch and real image.

5. In each class of the accessible dataset, a small number of Sketches are compared to real images [4]. Strategies for sketch-based image recognition and retrieval are generally similar. Handcrafted features including the Fisher vector, Scale Invariant Feature Transform (SIFT) Strategies for sketch-based image recognition and retrieval are generally similar. Handcrafted features including the Fisher vector, SIFT [5], GF-HOG [6], Histograms of Oriented Gradients (HOG) and Structural Similarity Index Measure (SSIM) were employed in earlier methods. To acquire broad features, these descriptors are frequently paired with a Bag Of Words (BOW) [7]. The main factor in differentiating learning Sketches is their unique and potent features [8]. Until now, the majority of models have solely considered account visual details like color and texture. Yet, shape information has not been the subject of many investigations. Simultaneous consideration of appearance and shape information leads to considerable improvements in sketch-based image recognition and retrieval.

In image processing researches, the use of various features has always been confirmed, taking into account low-level and high-level features at the same time, features such as depth, shape, appearance [9], semantic features, map features [10], visual and time features [11], texture, hierarchy [12], etc. Based on the problem nature, the appearance and shape features are considered at the same time, so that both the general appearance information of the Sketch and the features related to its shape are considered since, the Sketch consists of a series of simple lines. Therefore, the most comprehensive features that can be effective are these two features. Although there have been other studies in the past, that have used these two features in the retrieval problem like [9], the idea of the upcoming research is to use the architecture of Inception V1 in the branch of extracting the appearance feature with the aim of using filters with different dimensions at the same time, which improves the performance. In addition, the use of the domain adapting module, along with this hybrid network, had a positive effect on the retrieval accuracy, which has not been used in similar research.

Therefore, the current study aims to improve the performance of the SBIR system, by only using and analyzing the Sketches, not real images. Accordingly, the best model has been proposed by simultaneously considering the appearance features and shape features through a two-branch network and testing the architectural combination of different networks. In addition, with the aim of reducing the domain distance between Sketches and images, and achieving a more accurate recovery system, the CCA technique module, has been used [13].

Related Works

Since the early 1990s, SBIR was explored, and between 1990 to 1994, Content-Based Image Retrieval (CBIR) was a popular topic, which is still an interesting topic for the researches. Research on SBIR started in 1994 [7]. SBIR is a cross-domain retrieval research where the dataset is made up of real images and the query is run from an abstract sketch [14]. As a result, the majority of studies focus on identifying traits that Sketches and images have in common. For SBIR, training methods extract shallow features [4]. The previously stated SIFT [5], GF-HOG [6], HOG [7], SSIM [15], Soft Histogram of Edge Logal Orientations (SHELO) [16], Fisher vector [17], etc. are the major emphasis of low-level features. Furthermore, several studies also aim to get features at the medium level.

Pixel-based methods often have limited flexibility and high processing costs. Subsequently, a feature extraction module was added to extract several features, with significant edge changes. Sabeti [18], first decomposed the image into blocks of equal size and then extracted a feature vector from each block. Subsequently, a tree-structured hierarchical clustering technique, was developed to divide the blocks into multiple classes based on the extracted features and train a classifier for each class to determine whether a block is from the same overlapping image. In some studies like [19], spectral and spatial features have been used simultaneously for image processing.

In addition, a binary mask was used for objects that matched the original image spatially. The gallery display module also combines the best-bin-first and K-means tree. The amalgamation of these two methods resulted in

a multiplication of the retrieval speed. Moreover, the multiscale representation of edge maps was used to illustrate changes in the degree of detail contained in Sketches created by humans. They found that the intricacies of the sketch were preserved by the mixture of scales [20]. In another research, the K-means fuzzy clustering is used in combination with the convolution neural network [21]. Oriented Chamfer Matching (OCM) is another pixel-based method that examines the sketch's closest edge pixel that matches the image. Sain et al. [22] introduced a novel network that can cultivate certain hierarchies and make use of them to match the Sketch with the image in the appropriate hierarchical levels.

Research in the field of SBIR has recently taken on a new shape with the advent of deep learning and the application of deep neural networks [23]. Due to Deep Convolutional Neural Network's (DCNN's) effectiveness, researchers have also created potent deep learning-based methods that can model intricate sketch aspects. Convolutional networks decrease ambiguities and defects in data and are thorough and effective in image processing [24]. In general, neural network-based methods outperform conventional statistical models and offer stronger data pattern identification, faster processing, and greater resistance to environmental changes [25]. Custom architectures for feature ranking and prediction have been employed recently, including multi-objective ranking networks [26], hybrid CNN models [9], Alex-Net, and Google-Net [27]. In [9], although the combined network of S_net and A_net has used, it has not done anything to reduce the distance of sketch and image nor studied about different network architectures. [28] has extracted the deep features of the sketch using a hybrid CNN model based on ImageNet, but it can be criticized because it did not use a proper pre-processing and only used canny edge extraction. In [29], a deep convolutional three-branch neural network named Sketch Net with the Soft_max function as the cost function is used in a weakly supervised method. Bhattacharjee et al. [30] claims that by proposing an adaptive search method, the query is able to locate small objects in complex background, but only used appearance information to disambiguate between object proposals and refine search results. [31] solved the SBIR problem as a three-stage solution consisting of a self-awareness module for feature learning, a mutual-awareness module to generate probabilistic matches, and a kernel-based communication network to aggregate local matches. Prolonging the process of solving the problem and not considering the pre-processing stage of the data are among the points that have not been analyzed in this research. Chauhan [32] developed a novel method for Zero Shot Sketch Based Image Retrieval (ZS-SBIR) called aligned adaptability and generalizability Adaptability Balance Domain and Generalizability (ABDG). Generally speaking, this method aligns discriminability with generalizability learning for each instance by applying a two-stage advanced knowledge sketch. An end-to-end deep network termed DVSF is presented in [33] for sketch recognition. It feeds deep data to visual and sequential networks at the same time to acquire temporal and spatial features. Some researchers have proposed and solved the problem in a fuzzy way [34], [35]. A deep structure known as Network In Network (NIN)—a micro-neural network with a complicated structure—was shown in research [36]. A robust performance estimator, the multilayer perceptron, is used in the creation of this network. Wang et al.'s network with contrast cost function [37] was one of the first studies on multi-branch networks for sketch retrieval (3D objects). Each branch's weights were individually trained to correspond with the sketch and image domains. Some other examples of research that have studied multi-branch and hybrid networks in the subject of image retrieval are introduced below. Visual Geometry Group (VGG) Net is a CNN-based network that uses data augmentation techniques in conjunction with dimensionality reduction of the output layer [38]. Qi et al. [39] recently suggested a bifurcated Siamese network with contrast cost function. A triple CNN—an anchor branch—is used by Tu et al. [40] as a reference object. The use of triple CNNs for face recognition [41], tracking, visual search for photography, and queries intended to refine search inside a class [3] have recently been investigated. Similarly, Sangkloy et al. [42] used the Sketchy system to implement a Fine-Grained SBIR method.

Additionally, a multi-stage, triple network based on CNN is suggested in [43]. A straight forward and effective method that learns the semantic embedding space from a vision model and does not require a lot of computer training resources is shown in [44]. Indeed, a pre-trained Image Net CNN is used here along with three cost functions: semantic knowledge retention, domain adapter, and semantic classifier. According to the experts, utilizing a completely shared network is preferable to utilizing a bifurcated network without weight sharing. On the other hand, the authors of [42] argued that considering cross-category retrieval, it is better to avoid layer sharing. Additionally, Bui et al. [45] analyzed a hybrid sketch that shared certain layers but had identical architecture in both branches.

Experiments conducted on actual datasets demonstrate that deep features outperform handmade descriptors in SBIR. A sketch is first transformed into a collection of points, and then CNNs are used to describe the shape information based on the set of points to acquire strong shape features. Finally, CNN's two-stream Sketch combines appearance and shape information to provide unique features. To lessen the domain gap, the domain adaptation module is also utilized under the title of the conventional correlation analysis approach module CCA [13].

Proposed Method

This section presents the bifurcated hybrid convolutional network's architecture and discusses the specific details of each component.

A. Hybrid Convolutional Network Architecture

Generally, shape and appearance data may be used to characterize a sketch. To integrate appearance and shape information, a bifurcated hybrid convolutional network is therefore suggested. Fig. 1 depicts this network's overall structure. The definition of the A-Net network, the first branch, describes appearance data, whereas the S-Net network, the second branch, describes and extracts shape features.

The two networks have distinct inputs. The A-Net network's input directly determines the main sketch image. The probabilities of each object class constitute the branch's output, and following feature extraction, the feature vectors are sent to all linked layers.

On the other hand, many points are initially sampled from the sketch image to acquire the shape's features. The feature vectors related to the shape are then extracted using layers like sketch layers, alignment layers, perceptron, and pooling. S-Net specifications are covered in more depth later.

We have both appearance features and shape features when A-Net and S-Net are trained separately. Subsequently, the two feature vectors undergo independent concatenation, resulting in a composite feature representation. In other words, appearance and shape feature vectors are obtained from A-Net and S-Net branches, respectively. These vectors are normalized independently and separately, and then we put them together and use them as a combined feature vector for sketch-based image retrieval and classification. Finally, pattern classification and pattern-based image retrieval are achieved through the utilization of the hybrid features. The hybrid features are fed into a Support Vector Machine (SVM) classifier for sketch classification.

The studied sketch features and the resulting edge maps are taken from the image dataset for vector SBIR. The distance is then compared to determine the ranking results. The tests' findings demonstrate how image classification and retrieval are enhanced when hybrid appearance and shape data are used.

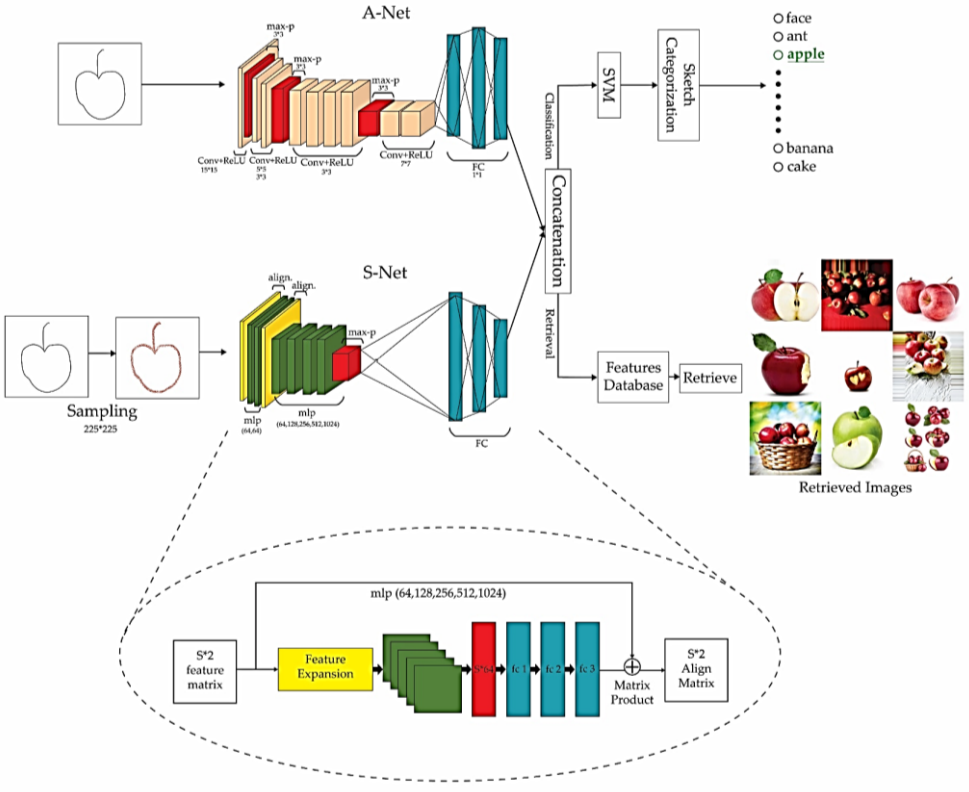


Fig. 1: The general framework of the proposed bifurcated hybrid convolutional network.

To show the shape information for the S-Net network, the sketch is first transformed into a set of points, from which features are then retrieved. Compared to other descriptors, the shape features represented by the point set are distinct. There are some problems in this regard: The finished drawings have distinct styles when they are done by various artists. For various wheels and Sketch styles, the learned representation of sample points should be uniform. As mentioned above, a Sketch is described by extracting a large number of points. Various

starting points for sampling might result in different perturbations for the same Sketch.

S-Net architecture is suggested as a solution for these problems. The class label is the output of the S-Net network, which receives the collection of points as input. A pre-processing step is necessary to create the set of points that make up the S-Net input from the sketch. Sampling with 512 points, has been chosen based on testing this sampling method with a variety of points and evaluating the network's performance, to minimize the computation's time complexity.

The set of collected points is aligned with a conventional space after acquiring the sampled point image. Subsequently, an MLP network extracts the features. Alternatively, the set of points is aligned in the feature space using a transformation matrix. The sketch's high-level shape features are extracted using a Multi-Layer Perceptron (MLP) network that has fully linked layers and a pooling layer. In the below the general Algorithm of the proposed method has been shown.

Algorithm 1: Algorithm of the proposed method

INPUT:

Train Img Set: training set that is images are randomly selected in 9 set 8, 16, 24, 32, 40, 48, 56 and 72 of Sketches
Test Img Set: randomly selected of 10% of the remaining data
Pre-Train: $\mathbf{I_A}$: All input Sketches are resized to 225*225
$\mathbf{I_S}$: 512- point Farthest Point Sampling (FPS) & Resized to 225*225

1. Normalization of the training set for classification and retrieval
2. For a query sketch $\mathbf{I_A}$
   Concatenate the deep visual and extract appearance features
   Return the probability of each of each object class
3. For a query sketch $\mathbf{I_S}$
   Concatenate the alignment to predict a M = affine transformation matrix and extract shape features
   Return the label of class
4. Training step:
   For e = 1:E do E = number of epochs
   a) Feed the sketch in to A-Net & S-Net
      $\mathcal{F}_{fc_2}^{app}(\boldsymbol{I_A}) \leftarrow \boldsymbol{I_A}$ /*sketch appearance feature mapping*/
      $\mathcal{F}_{fc_2}^{shape}(\boldsymbol{I_s}) \leftarrow \boldsymbol{I_s}$ /*sketch shape feature mapping*/
   b) Calculate the ranking loss according to (1) and update the model parameters
   c) $\mathcal{F}^{+} \longleftarrow \left[\mathcal{F}_{fc_2}^{app}, \mathcal{F}_{fc_2}^{shape}\right]$ combining feature vectors

   End
5. Ranking:
   a) Classification: using SVM classification
   b) SBIR: - using CCA convolutional correlation analysis technique module
      - using K-Nearest Neighbors (KNN) to find the closest answer to the query sketch between the sketch query and top-k samples in testing set

OUTPUT:

Some of the closest real images to the query sketch

1. Preprocessing

The images are initially resized to 225 × 225 during the pre-processing phase. The background is represented by one and lines by zero. Then, as X $= \{x_1, x_2, \ldots, x_n\}$, a subset of points taken from the sketch are chosen. Although the set of points might be selected at random, the findings indicate that it is preferable to have a minimum amount of space between the points. As a result, the Farthest Point Sampling (FPS) sub-set of points' furthest repetitive point is chosen using sampling.

Compared to random sampling, using this sampler offers greater coverage for a sketch and as close to uniform sampling as possible. As previously stated, 512 points are sampled for each sketch, and the coordinates (x,y) of each point are used to signify it.

2. Alignment Stage

Rotation and transformation/conversion are two examples of geometric transformations in which the semantic information of the sampled points of the Sketches should be invariant. To align the input space, an alignment network is built to predict a two-by-two affine transformation matrix. The set of input points is then multiplied directly in the anticipated transformation matrix. The whole input alignment procedure is depicted in Fig. 1.

The alignment phase is carried out twice in the suggested network architecture: once in the input space and once in the feature space. Although the transformation matrix in the feature space is 64x64, the alignment in these two spaces is similar. Three completely

linked layers with output sizes of 1024, 512, and 256 are inserted in the continuation of Max Pooling. The output sizes of each layer are 64, 128, 256, 512, and 1024, respectively.

A regularization component is introduced to the cross entropy as a function of training cost because optimization becomes challenging when the feature space's transformation matrix has much larger dimensions than the original input space.

$$\ell = \mathcal{T}(\theta) + \Im \times \mathcal{L} \tag{1}$$

$$\mathcal{T}(\theta) = \frac{-1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\left(y^{(i)} = j\right) log \frac{exp(a_k)}{\sum_{k^*=1}^{c} exp(a_{k^*})}\right] \tag{2}$$

$$\mathcal{L} = \|I - MM^T\|_2^2 \tag{3}$$

$$a_k = \sum_j W_{kj} Z_j \tag{4}$$

where $\Im$ is the $\mathcal{L}$ function's alignment weight. Also, it was found that $\Im = 0.01$ provides the best accuracy based on the testing and errors. The output of hidden unit j is Z_j. The feature alignment matrix is M, and the input is I [9]. By including regularization into the cost function, the orthogonal transformation preserves the original data, resulting in stable optimization and improved model performance. Three input layers, one hidden layer, and one output layer exist in every MLP. In the suggested MLP, five hidden layers—64, 128, 256, 512, and 1024 neurons—are added to the MLP after the second alignment, and two hidden layers—64 and 64 neurons—are added to the S-Net network following the first alignment. The points' irregularity is crucial for the shape descriptor.

Max Pooling is used to collect data from every point to ensure the consistency of the collection of sampled points. In this case, n vectors that were computed by the preceding MLP module are the input that Max Pooling uses. As a result, a fresh vector that supplies the completely linked layers is produced.

Fully linked layers apply an offset bias and multiply their input by a weight matrix. The S-Net network consists of three completely linked layers, with the number of Sketch categories being the output of the third layer.

B. Hybrid Convolutional Network Training

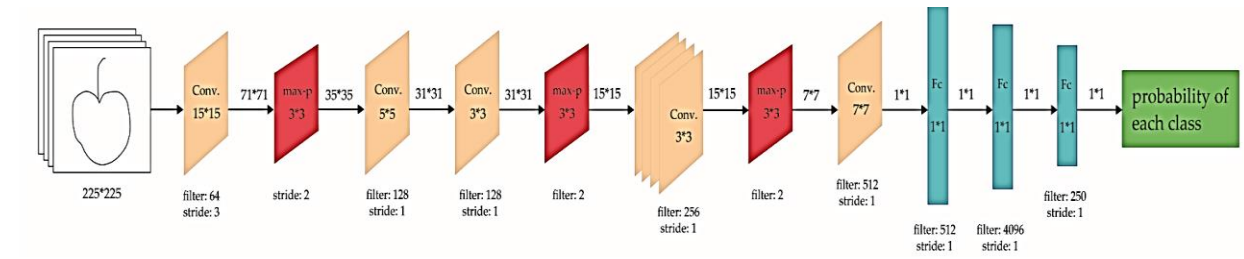
The appearance features and shape features are combined using a hybrid procedure that joins $\mathcal{F}_{fc_2}^{app}$ and $\mathcal{F}_{fc_2}^{shape}$ shape with ℓ_1 normalization specified by $\mathcal{F}^+$ after the A-Net and S-Net have been trained separately [9].

$$\mathcal{F}^+ \leftarrow \left[\mathcal{F}_{fc_2}^{app}, \mathcal{F}_{fc_2}^{shape}\right] \tag{5}$$

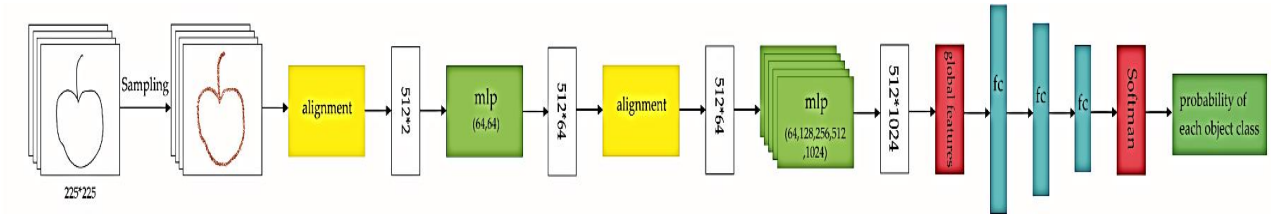
A binary SVM is separately trained for every category to classify the sketch. Prediction scores are produced for each category at the time of evaluation, which makes direct use of the integrated features of each design. The next section will provide the features of the sketch classification.

The edge maps of the natural image data set are initially extracted using gpb to reconstruct the image-based plan.

Subsequently, features are extracted on the resultant edge maps and the sketch in question separately using the hybrid CNN. Finally, K top candidate images for the given sketch are retrieved using K-Nearest Neighbors (KNN) classification with Euclidean distance. The hybrid CNN network's retrieval performance is covered in the evaluation section.



(a)



(b)

Fig. 2: Architectural details of (a) A-net and (b) S-net.

Fig. 2 shows the architecture details of two branches of the proposed network. As shown, a CNN network consists of convolution and max pooling layers, which end up with fully connected layers. Testing various examples of CNN, including VGG, Google Net, Alex-Net, Sketch A Net and Inception V1 for A-Net branch in combination with alignment network (S-Net) demonstrated that the CNN network with the Inception V1 module has the best performance.

Results and Discussion

This section will analyze the suggested hybrid CNN network. First, a description of the evaluation's data set is given. Then, the network's performance and its results are discussed.

A. Evaluation Criteria

1. Accuracy precision

One of the most frequently used evaluation criteria in classification problems is accuracy. It is determined by dividing the total number of identified samples (both properly and wrongly categorized) by the ratio of correctly classified samples [16]. The following formula is used to calculate accuracy.

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (6)$$

where FP and TP represent samples that were incorrectly identified and those that were correctly identified.

2. Recall

The following formula may be used to represent recall, which is a metric derived from the ratio of properly classified samples to the total of correctly identified and incorrectly rejected samples [46].

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (7)$$

3. MAP (Mean Average Accuracy)

The weighted mean of the accuracy at each threshold is used to compute the mean accuracy. Recall gains weight over the preceding level. The mean AP for every class determines the mean accuracy [47].

$$AP = \frac{1}{N} \sum_{r} P_{interp}(r) \quad (8)$$

where MAP is the mean AP in each data class and $P_{interp}(r)$ is the precision in calling point(r).

B. Datasets

The two datasets used for classification evaluation are Sketchy and TU-Berlin.

1. TU-Berlin

This collection of 20,000 hand-drawn images, created by regular people, is divided into 250 groups, the majority of which include common everyday items like eyes, aircraft, apples, and bananas. Each design measures 1111×1111. There are eighty sketch images collected for every class.

2. Sketchy

There are 125 classifications in the collection, and there are 75471 human Sketches for items. In this study, the remaining Sketches were utilized for training, and 50 Sketches from each category were used for testing.

3. Flickr 15k

15024 images make up this data collection, from which samples of Sketches have been taken. Ultimately, 33 categories have been created from the labeled sketch images, with 10 Sketches in each category. Using data augmentation methods, this data set has been augmented during the review process. Among the methods employed for data augmentation were: 1. Vertical and horizontal flip 2. Turning angles of ±10 and ±15 degrees 3. Magnification ±1.2 and ±1.5 times the design height; 4. Taking into account the planar movement of Sketches, translation in both horizontal and vertical directions with ±10 and ±15 pixels is taken into account.

C. Training and Testing

TensorFlow's Stochastic Gradient Descent (SGD) is used to train the model. The batch size is 64, the momentum is 0.9, and the learning rate is 0.01 for the A-Net network. Every period of the optimization procedure involves a 0.3 reduction, and it is terminated after 15 optimization periods. The initiating weights of the bias terms have a Gaussian distribution with δ=0.01 and μ=0, and they are initialized from 0.1. The final three completely linked layers of the S-Net network employ elimination at a rate of 0.7. For batch normalization, the decay rate initiates at 0.5 and rises steadily to 0.99. Additionally, an Adam optimizer with a batch size of 64, a momentum of 0.9, and a learning rate of 0.001 was utilized. The learning rate in each of the five courses is split in two.

Nine educational division models are put to the test. For every class, a random selection of 8, 16, 24, 32, 40, 48, 56, and 72 Sketches is used as the training data set. To estimate robust model parameters, however, 10% of the remaining data are chosen at random to serve as the validation data set.

D. Evaluation

Table 1 and Table 2 show that the following results can be attained:

- Compared to other basic networks, the proposed hybrid convolution performs better. As shown, it achieves a classification score of 84.44% for the TU-Berlin dataset and 82.76% for Sketchy.
- Tables 1 and 2's results indicate that NIN performs better than VGG-Net when there are less than 50

training plans, while VGG-Net performs better than NIN when there are more than 50 training plans. However, as compared to base networks, the hybrid CNN performs better across all training set sizes.

- Deep features outperform hand-made features in terms of performance; nonetheless, the more images in the training plan, the better the performance to the extent that the classification accuracy is higher than human accuracy, even with a significant amount of training images. With 45 training Sketches, the system achieves superhuman accuracy on average for each category. However, a hybrid convolutional network needs just 35 training samples to attain this limit.

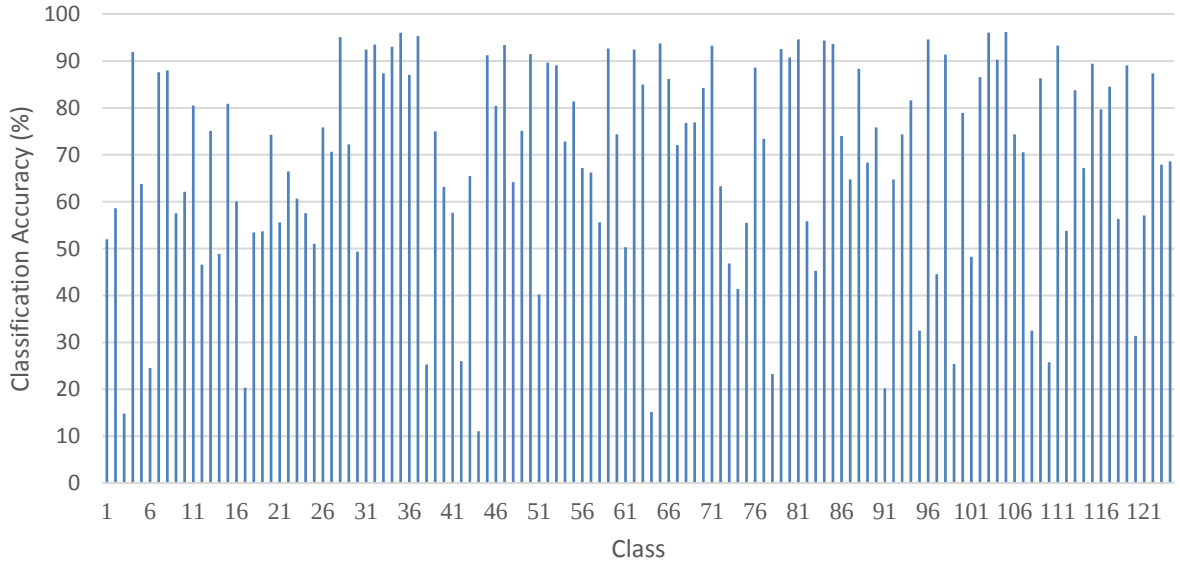
Table 1: Classification accuracy of different features on the TU-Berlin dataset

Method	Size of training set						
	10	**20**	**30**	**40**	**50**	**60**	**70**
DVSF [33]	60.52	67.43	71.85	73.96	75.87	77.03	78.94
Sketch Net [29]	59.91	65.82	70.06	73.02	75.41	76.25	78.00
VGG Net [38]	51.22	61.33	66.43	66.96	70.09	74.26	86.02
NIN [36]	54.13	64.42	67.34	70.74	71.87	73.00	74.86
Fisher Vector Size 24 [17]	44.96	54.21	58.32	62.18	65.14	66.58	67.86
Fisher Vector Size 16 [17]	42.56	50.37	54.17	58.40	60.85	63.64	65.03
Eitz (SVM soft) [48]	35.00	42.20	45.63	50.12	51.11	54.00	55.20
Eitz (SVM hard) [48]	33.10	39.42	44.16	48.04	49.14	52.13	53.00
SSIM [15]	28.97	38.00	43.52	46.11	48.35	50.91	52.01
Proposed Method	**61.81**	**67.91**	**73.00**	**75.21**	**76.14**	**80.20**	**84.44**

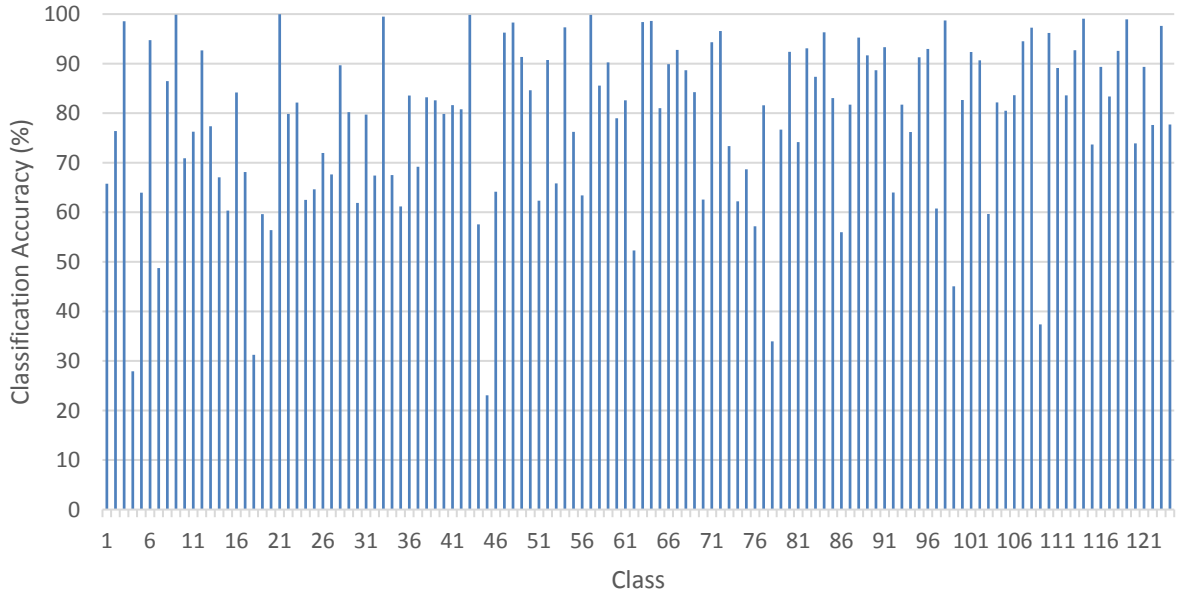
Table 2: Classification accuracy of different features on Sketchy dataset

Method	Size of training set						
	10	**20**	**30**	**40**	**50**	**60**	**70**
DVSF [33]	58.03	64.05	69.21	72.76	74.00	75.18	77.58
Sketch Net [29]	57.53	63.41	68.92	71.94	73.26	74.88	76.84
VGG Net [38]	51.28	59.98	62.27	66.54	70.11	72.85	74.12
NIN [36]	52.02	61.51	65.44	68.07	69.16	70.19	71.96
Fisher Vector Size 24 [17]	43.48	51.27	55.05	59.09	62.20	64.38	66.45
Fisher Vector Size 16 [17]	42.89	50.23	54.08	58.98	61.03	63.67	65.13
Eitz (SVM soft) [48]	32.24	38.76	43.37	45.88	49.14	51.96	53.12
Eitz (SVM hard) [48]	30.24	37.07	42.16	46.14	47.10	48.29	51.94
SSIM [15]	26.92	35.61	40.66	43.07	45.18	46.71	48.96
Proposed Method	**59.98**	**65.78**	**71.16**	**73.26**	**74.97**	**77.53**	**82.76**

Using the Sketchy dataset, the classification performance is evaluated with and without taking the shape feature into account. Fig. 3 illustrates how, when shape features are included, the classification accuracy of seven categories is less than 45% and the accuracy of more than 50% of categories is above 70%.



(a)



(b)

Fig. 3: Classification performance of the hybrid CNN on Sketchy dataset a) shape feature-free b) including shape feature.

As shown in Fig. 3, the performance of the S-Net branch is lower than the A-Net branch. In addition, considering shape information, the proposed combined network has performed better than the A-Net branch alone. These results indicated that the extraction and use of shape features can enhance the classification performance of the scheme.

Another experiment was conducted to investigate the impact of different network architectures and choosing the best architecture for the proposed network. The results of this experiment are shown in Fig. 4.

The results of Fig. 4, indicated that the convolution with Inception V1 module has performed better than other architectures in the field of classification. After Inception V1, Alex-net and Google-net are the architectures that can be used in the field of image retrieval with a slight difference.

Table 3 compares the results of several methods using the Sketchy dataset and the MAP criterion.

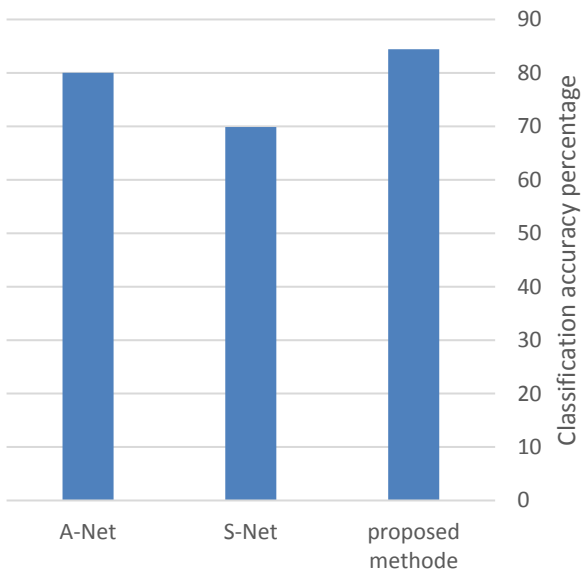


Fig. 4: The effect of each branch of the proposed network alone and combined.

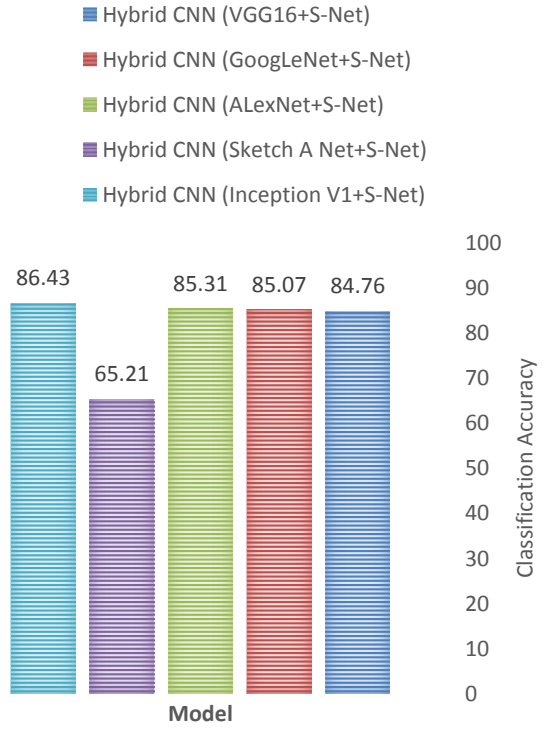


Fig. 5: Testing different architectures in combination with S-Net network.

The SBIR is tested using the expanded 15k Flickr dataset; for this evaluation, 2/3 of the dataset is used for training and 1/3 for testing. The evaluation's findings on the expanded Flickr 15k dataset are displayed in Table 4.

It is evident from the table data that all deep learning-based methods outperform hand-made feature extraction methods.

Table 3: Results of SBIR based on the MAP criterion on the Sketchy dataset

Method	MAP
HOG [49]	0.115
GF-HOG [6]	0.157
SHELO [50]	0.161
Siamese With Contrastive Loss [39]	0.195
RST-SP-SHELO [16]	0.200
Sketch A Net [26]	0.208
Triplet Sketch-Edgemap [45]	0.244
Query adaptive Reranking CNN [30]	0.323
Sketchy Triplet [42]	0.359
Sket RET [32]	0.437
Siamese CNN [39]	0.481
Cross Modal [22]	0.523
GN Triplet [42]	0.529
SBT-Net [44]	0.553
CNN with Multi Stage Regression [43]	0.565
Hybrid CNN [9]	0.574
Proposed Method	**0.585**

Table 4: The MAP criterion-based SBIR results on the extended Flickr 15k dataset

Method	MAP
Sketch A Net [26]	0.084
SIFT [51]	0.101
HOG [49]	0.119
Query adaptive Reranking CNN [30]	0.130
GF-HOG [6]	0.132
Sketchy Triplet [42]	0.145
Sket RET [32]	0.177
Siamese CNN [39]	0.195
Cross Modal [22]	0.212
GN Triplet [42]	0.214
SBT-Net [44]	0.224
CNN with Multi Stage Regression [43]	0.229
Triplet CNN [40]	0.245
Hybrid CNN [9]	0.287
Proposed Method	**0.293**

Conclusion

This research presented the hybrid convolutional framework, which is based on deep learning, for sketch retrieval. In addition to appearance information, it is thought that shape information is also very efficient in sketch recognition and retrieval. A hybrid network architecture comprising two networks—S-Net and A-Net—is shown in this article under the heading of hybrid convolution.

These subnetworks, in turn, describe appearance and shape information.

The classification accuracy of the proposed method reached 84.44% on the comprehensive

TU-Berlin dataset and 82.76% accuracy on the challenging Sketchy dataset, which is a significant value. It was also shown how the operation of two branches of the network simultaneously can be effective in the image retrieval process than when each branch is used alone, and this highlights the importance of using appearance and shape features at the same time. On the other hand, using the CCA module to reduce the domain distance of the sketch and image, as well as choosing the appropriate network architecture, improved the efficiency of the proposed retrieval system. Compared to the best available methods, hybrid network convolution has increased recognition and retrieval accuracy by around 5%. It is an efficient and thorough method which demonstrated valid results in sketch-based image classification and retrieval on TU-Berlin, Flickr 15k wide, and sketchy datasets.

Author Contributions

A. Gheitasi, H. Farsi, and S. mohammadzadeh designed the experiments. A. Ghatasi collected the data. A. Ghatasi carried out the data analysis. A. Gheiaasi, H. Farsi, and S. Mohamadzadeh interpreted the results and A. Mahboobi wrote the manuscript. H. Farsi and S. Mohamadzadeh Corrected editing errors.

Acknowledgment

I am sincerely grateful for the guidance and help of honorable professors Mr. Hassan Farsi and Sajjad Mohammadzadeh. This work is completely self-supporting, thereby no any financial agency's role is available.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

SBIR	Sketch Base Image Retrieval
S_Net	Shape Network
A_Net	Appearance Network
CCA	Conventional Correlation Analysis
CNN	Conventional Neural Network
TU_Berlin	Technische Universität Berlin
SIFT	Scale Invariant Feature Transform
HOG	Histograms of Oriented Gradients
SSIM	Structural Similarity Index Measure
BOW	Bag Of Words
OCM	Oriented Chamfer Matching
DCNN	Deep Conventional Neural Network
ZS_SBIR	Zero Shot Sketch-Based Image Retrieval
ABDG	Adaptability Balance Domain and Generalizability
DVSF	Deep Visual-Sequential Fusion
NIN	Network In Network
VGG	Visual Geometry Group
SVM	Support Vector Machine
MLP	Multi-Layer Perceptron
Img	Image
FPS	Farthest Point Sampling
KNN	K-Nearest Neighbors

MAP	Mean average Accuracy Percent
SHELO	Soft Histogram of Edge Logal Orientations
SGD	Stochastic Gradient Descent
GN	Google Net

References

[1] D. Birari, D. Hiran, V. Narawade, "Survey on sketch based image and data retrieval," in Proc. 2nd International Conference on Communications and Cyber Physical Engineering (ICCCE 2019): 285-290, 2020.

[2] A. Chaudhuri, A. K. Bhunia, Y. Z. Song, A. Dutta, "Data-free sketch-based image retrieval," arXiv preprint arXiv:2303.07775, 2023.

[3] A. Sain, A. K. Bhunia, Y. Yang, T. Xiang, Y. Z. Song, "Stylemeup: Towards style-agnostic sketch-based image retrieval," in Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[4] P. Xu, T. M. Hospedales, Q. Yin, Y. Zh. Song, T. Xiang, L. Wang, "Deep learning for free-hand sketch: A survey," IEEE Trans. Pattern Anal. Mach. Intell., 45(1): 285-312, 2023.

[5] Z. Hossein-Nejad, H. Agahi, A. Mahmoodzadeh, "Remote sensing image registration based on a geometrical model matching," J. Inf. Syst. Telecommun. (JIST), 5(36): 41, 2021.

[6] R. Hu, S. James, T. Wang, J. Collomosse, "Markov random fields for sketch based video retrieval," in Proc. the 3rd ACM conference on International Conference on Multimedia Retrieval (ICMR): 279-286, 2013.

[7] Y. Li, W. Li, "A survey of sketch-based image retrieval," Mach. Vision Appl., 29(7): 1083-1100, 2018.

[8] S. Mohamadzadeh, S. Pasban, J. Zeraatkar-Moghadam, A. K. Shafiei, "Parkinson's disease detection by using feature selection and sparse representation," J. Med. Biol. Eng., 41(4): 412-421, 2021.

[9] X. Zhang, Y. Huang, Q. Zou, Y. Pei, R. Zhang, S. Wang, "A hybrid convolutional neural network for sketch recognition," Pattern Recognit. Lett., 130: 73-82, 2020.

[10] P. Xu, T. M. Hospedales, Q. Yin, Y. Zh. Song, T. Xiang, L. Wang, "Deep learning for free-hand sketch: A survey," IEEE Trans. Pattern Anal. Mach. Intell., 45(1): 285-312, 2020.

[11] P. Xu, Y. Huang, T. Yuan, K. Pang, Y. Zh. Song, T. Xiang, T. M. Hospedales, Zh. Ma, J. Guo, "Sketchmate: Deep hashing for million-scale human sketch retrieval," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018): 8090-8098, 2018.

[12] A. Dutta, Z. Akata, "Semantically tied paired cycle consistency for zero-shot sketch-based image retrieval," in Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019): 5089-5098, 2019.

[13] H. Yu, M. Huang, J. J. Zhang, "Domain adaptation problem in sketch based image retrieval," ACM Trans. Multimidia Comput. Commun. Appl., 19(3): 1-17, 2022.

[14] A. K. Bhunia, A. Sain, P. H. Shah, A. Gupta, P. N. Chowdhury, T. Xiang, Y. Zh. Song, "Adaptive fine-grained sketch-based image retrieval," in Proc. 17th European Conference Computer Vision (ECCV 2022), Part XXXVII: 163-181, 2022.

[15] E. Shechtman, M. Irani, "Matching local self-similarities across images and videos," in Proc. 2007 IEEE Conference on Computer Vision and Pattern Recognition: 1-8, 2007.

[16] J. M. Saavedra, "Rst-shelo: Sketch-based image retrieval using sketch tokens and square root normalization," Multimedia Tools Appl., 76(1): 931-951, 2017.

[17] R. G. Schneider, T. Tuytelaars, "Sketch classification and classification-driven analysis using fisher vectors," ACM Trans. Graphics (TOG), 33(6): 1-9, 2014.

[18] V. Sabeti, "An improved approach to blind image steganalysis using an overlapping blocks idea," J. Electr. Comput. Eng. Innovations, 11(2): 263-276, 2023.

[19] M. Imani, "Target detection using multispectral images, A case study: Wheat detection in Chenaran County in Iran," J. Electr. Comput. Eng. Innovations, 9(1): 11-24, 2020.

[20] M. Rezaei, M. Rezaei, "Foreground-back ground segmentation using k-means clustering algorithm and support vector machine," J. Inf. Syst. Telecommun. (JIST), 1(41): 65, 2023.

[21] S. Fooladi, H. Farsi, S. Mohamadzadeh, "Segmenting the lesion area of brain tumor using convolutional neural networks and fuzzy K-means clustering," Int. J. Eng., 36(8): 1556-1568, 2023.

[22] A. Sain, A. K. Bhunia, Y. Yang, T. Xiang, Y. Zh. Song, "Cross-modal hierarchical modelling for fine-grained sketch based image retrieval," arXiv preprint arXiv:2007.15103, 2020.

[23] A. K. Bhunia, P. N. Chowdhury, A. Sain, Y. Yang, T. Xiang, Y. Zh. Song, "More photos are all you need: Semi-supervised learning for fine-grained sketch based image retrieval," in Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 4247-4256, 2021.

[24] A. K. Bhunia, S. Koley, A. F. U. R. Khilji, A. Sain, P. N. Chowdhury, T. Xiang, Y. Zh. Song, "Sketching without worrying: Noise-tolerant sketch-based image retrieval," in Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019): 999-1008, 2022.

[25] A. Gheitasi, H. Farsi, S. Mohamadzadeh, "Estimation of hand skeletal postures by using deep convolutional neural networks," Int. J. Eng., 33(4): 552-559, 2020.

[26] Q. Yu, F. Liu, Y. Zh. Song, T. Xiang, T. M. Hospedales, Ch. Ch. Loy, "Sketch me that shoe," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016): 799-807, 2016.

[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 1-9, 2015.

[28] N. Kumar, R. Ahmed, V. B. Honnakasturi, S. Sowmya Kamath, V. Mayya, "Sketch-based image retrieval using convolutional neural networks based on feature adaptation and relevance feedback," in Proc. International Conference on Emerging Applications of Information Technology: 103-113, 2022.

[29] H. Zhang, S. Liu, Ch. Zhang, W. Ren, R. Wang, X. Cao, "Sketchnet: Sketch classification with web images," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 1105-1113, 2016.

[30] S. D. Bhattacharjee, J. Yuan, W. Hong, X, Ruan, "Query adaptive instance search using object sketches," in Proc. the 24th ACM International Conference on Multimedia: 1306-1315, 2016.

[31] F. Lin, M. Li, D. Li, T. Hospedales, Y. Zh. Song, Y. Qi, "Zero-shot everything sketch-based image retrieval," in Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023): 23349-23358, 2023.

[32] R. Chavhan, Zero-Shot Sketch Based Image Retrieval, Indian Institute of Technology Bombay, 2021.

[33] J. Y. He, X. Wu, Y. G. Jiang, B. Zhao, Q. Peng, "Sketch recognition with deep visual-sequential fusion model," in Proc. the 25th ACM International Conference on Multimedia: 448-456, 2017.

[34] H. Zhao, M. Liu, M. Li, "Feature fusion and metric learning network for zero-shot sketch-based image retrieval," Entropy, 25(3): 502, 2023.

[35] E. Askari, S. Motamed, "Computational model for image processing in the minds of people with visual agnosia using fuzzy cognitive map," J. Inf. Syst. Telecommun. (JIST), 2(42): 102, 2023.

[36] M. Lin, Q. Chen, S. Yan, "Network in network," arXiv preprint arXiv:1312.4400, 2013.

[37] F. Wang, L. Kang, Y. Li, "Sketch-based 3d shape retrieval using convolutional neural networks," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition: 1875-1883, 2015.

[38] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," arXiv preprint arXiv:1405.3531, 2014.

[39] Y. Qi, Y. Zh. Song, H. Zhang, J. Liu, "Sketch-based image retrieval via siamese convolutional neural network," in Proc. 2016 IEEE International Conference on Image Processing (ICIP): 2460-2464, 2016.

[40] T. Bui, L. Ribeiro, M. Ponti, J. Collomosse, "Generalisation and sharing in triplet convnets for sketch based visual search," arXiv preprint arXiv:1611.05301, 2016.

[41] M. Rohani, H. Farsi, S. Mohamadzadeh, "Deep multi-task convolutional neural networks for efficient classification of face attributes," Int. J. Eng., 36(11): 2102-2111, 2023.

[42] P. Sangkloy, N. Burnell, C. Ham, J. Hays, "The sketchy database: learning to retrieve badly drawn bunnies," ACM Trans. Graphics (TOG), 35(4): 1-12, 2016.

[43] T. Bui, L. Ribiro, M. Ponti, J. Collomosse, "Sketching out the details: Sketch-based image retrieval using convolutional neural networks with multi-stage regression," Computers & Graphics, 71: 77-87, 2018.

[44] O. Tursun, S. Denman, S. Sridharan, E. Goan, C. Fookes, "An efficient framework for zero-shot sketch-based image retrieval," Pattern Recognit., 126: 108528, 2022.

[45] T. Bui, L. Ribeiro, M. Ponti, J. Collomosse, "Compact descriptors for sketch-based image retrieval using a triplet loss convolutional neural network," Comput. Vision Image Understanding, 164: 27-37, 2017.

[46] C. Bai, J. Chen, Q. Ma, P. Hao, Sh. Chen, "Cross-domain representation learning by domain-migration generative adversarial network for sketch based image retrieval," J. Visual Commun. Image Represent., 71: 102835, 2020.

[47] A. P. R. G. G. Rajput, "Sketch based image retrieval in large databases using edge features," Int. J. Recent Technol. Eng. (IJRTE), 08: 2277-3878, 2020.

[48] M. Eitz, J. Hays, M. Alexa, "How do humans sketch objects?," ACM Trans. Graphics (TOG), 31(4): 1-10, 2012.

[49] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," in Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1: 886-893, 2005.

[50] J. M. Saavedra, "Sketch based image retrieval using a soft computation of the histogram of edge local orientations (shelo)," in Proc. 2014 IEEE International Conference on Image Processing (ICIP): 2998-3002, 2014.

[51] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision, 60: 91-110, 2004.

Biographies

Azita Gheitasi received a Bachelor's degree in Electrical Engineering from Payam Noor University in Birjand center in 2014 and a Master's degree in Electrical Engineering from Shaukat Abad University in Birjand in 2017. Currently, she is a doctoral student in the field of Electrical Engineering and telecommunication Systems. Her research interests include image processing, deep learning, convolutional networks, cloud computing and machine learning.

- Email: A.gheitasi@birjand.ac.ir
- ORCID: 0009-0005-3953-9630
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Hassan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D. in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D. degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as Associate Professor in Communication Engineering in department of Electrical and Computer Eng., University of Birjand, Birjand, IRAN.

- Email: hfarsi@birjand.ac.ir
- ORCID: 0000-0001-6038-9757
- Web of Science Researcher ID: NA
- Scopus Author ID: 16202385600
- Homepage: https://cv.birjand.ac.ir/hasanfarsi/en

Sajad Mohamadzadeh received the B.Sc. degree in Communication Engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. and Ph.D. degree in Communication Engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012 and 2016, respectively. Now, he works as Associte Professor at department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Deep Neural Network, Pattern recognition, Digital Signal Processing, Sparse Representation, and Deep Learning.

- Email: s.mohamadzadeh@birjand.ac.ir
- ORCID: 0000-0002-9096-8626
- Web of Science Researcher ID: NA
- Scopus Author ID: 57056477500
- Homepage: https://cv.birjand.ac.ir/mohamadzadeh/en

How to cite this paper:
A. Gheitasi, H. Farsi, S. Mohamadzadeh, "Hybrid convolutional neural network with domain adaptation for sketch based image retrieval," J. Electr. Comput. Eng. Innovations, 12(2): 497-510, 2024.

DOI: 10.22061/jecei.2024.10778.735

URL: https://jecei.sru.ac.ir/article_2139.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

A Robust Control Centrality Applicable to Genetic Regulatory Networks with Structured Uncertainties

Z. Ghassemi Zahan [1], ***S. Ozgoli*** [1,*], ***S. Bolouki*** [2]

[1] *Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran.*

[2] *Department of Mechanical Engineering, Polytechnique Montréal, Quebec, Canada.*

Article Info

Article History:
Received 03 March 2024
Reviewed 17 April 2024
Revised 26 June 2024
Accepted 30 July 2024

Keywords:
Network analysis and control
Network centrality
Network controllability
Uncertain systems

*Corresponding Author's Email Address: *ozgoli@modares.ac.ir*

Abstract

Background and Objectives: In genetic network control, RC-Centrality is introduced as a new control centrality measure to address the control of linear time-invariant networks. The objective of this study is to propose an optimal control centrality metric that quantifies the centrality of individual nodes or groups of nodes within a network. Specifically, RC-Centrality identifies key nodes or node groups that can act as controllers, such as genes regulating the gene expression process. To assess the effectiveness of this method, RC-Centrality is compared with standard centralities in a real genetic network. Additionally, the research delves into the role of uncertainty structure in altering the priority order of RC-Centrality.

Methods: The RC-Centrality measure is introduced based on an optimal control problem to address weighted, directed, and signed networks. Robust controllers are designed to ensure Lyapunov stability under uncertainty. A cost function is introduced to measure the performance metric represented by input energy in the presence of uncertainty.

Results: The study presents RC-Centrality as an effective measure for identifying key nodes in genetic networks suitable for control. In-silico simulations are conducted to evaluate its performance in comparison to standard centralities. The research highlights the impact of uncertainty structure on the priority of RC-Centrality.

Conclusion: RC-Centrality offers a promising approach to identify essential nodes in genetic networks for control purposes. Its performance is demonstrated through simulations, and the study emphasizes the influence of uncertainty structure on the centrality measure's prioritization. This research has implications for understanding and controlling genetic networks, particularly in the presence of uncertainty.

Introduction

Graph theoretical representations are widely utilized to model biological networks and the relationships between various cellular components such as genes, proteins, mRNAs, and metabolites. The nodes of the graph, also referred to as agents, represent biological units, while links represent connections between them [1]. Identifying influential biological units plays an important role in designing and analyzing growth, death, division, and survival processes. Centrality analysis is proposed as a valuable guide to predict and identify key units in biological networks, such as genetic networks [2]-[6].

By combining centrality analysis with optimal control strategies, an effective approach has been developed to steer biological networks in the presence of uncertainty toward their desired states. This paper aims to contribute

to the subject by investigating novel control centrality that exploits the network's structure and performance metrics for optimal control in uncertain genetic networks.

A. Related Work

Centrality measures are employed to quantify the importance of components in biological, brain, social and urban traffic networks [7]-[9]. Here, several individual centrality measures for each agent have been listed for unweighted, undirected, and unsigned networks, used in our examples later on. Degree centrality $C_D(i)$ is defined as the number of agent i's neighbors. Closeness centrality $C_{cl}(i)$ is formulated as an inverse sum of the length of shortest paths between agent i and all other agents. The basic definition of Betweenness centrality measure $C_B(i)$ measures the extent to which an agent i lies on the shortest paths between other agents as an intermediate agent. The eigenvector centrality of agent i, $C_E(i)$, captures the agent's influence in the network based on the importance of its neighbors [7]. Although there are many methods for generalizing individual centralities, several papers are made on reformulations of individual centrality measures to extend them to group centrality [10], [11].

Energy optimization approach, which is also the approach taken in this work, is a branch of recent research used in Gramian-based centrality measures to control the given network in a desired direction [12], [13]. The control centralities $C_p(i)$ and $C_q(i)$ are formulated as the trace of the controllability and observability Gramian for node i, respectively [13]. The extension of this method to the discrete case introduces the proposed control centrality measures $C_W(i)$ and $C_M(i)$, formulated as the trace of the controllability and observability Gramian for node i, respectively [26]. Many recent papers on the subject of control energy requirements depend on the properties of controllability Gramian [14]-[16]. For uncertain networks, the influence of uncertainty on the energy is formulated in terms of the Lyapunov equation to minimize the required energy that guarantees stability [17]-[19]. The common characteristic of the efforts mentioned above in centrality measures' definition is that they are all proposed to quantify the importance of the focal node in the network. However, a great deal of research focuses on trying to quantify the power of a specific agent in controlling a network named control centrality [20], [21]. Several methods are used to quantify the influence of each agent or link in uncertain networks as their centrality measure [22], [23].

Despite research in the field of control centrality, some limitations are significant. First, the links' direction and weight in genetic regulatory networks (GRNs) model provide a more detailed and accurate description of gene expression process in biological systems' behavior. Second, interactions between genes are generally described by a signed graph, where the positive and negative weights represent the transcriptional activators and repressors attached to them, respectively. However, it should be noted that the specific studies discussed in [2], [6] do not address these key aspects. Furthermore, the dynamics of genetic networks are paramount for comprehending gene regulation processes and capturing the time-dependent behaviors of the genes and their interactions. However, references [3]-[5] do not specifically delve into the dynamics of their networks.

While in a genetic regulatory network, energy consumed by each gene from the external environment utilizes in gene expression process to control the network in a desired behavior to reach a given target. In practice, injecting the input energy into the cell for expressing genes in controlling biological processes is a costly and time-consuming process as they need to be done on living organisms [24]. On the other hand, the intrinsic nature of biological networks encompasses inherent uncertainty [25]. Thus, control centrality of genetic networks in presence of uncertainty according to optimal energy consumption is an important quantifier to identify the central genes to prevent off-target effects.

B. Contributions

While research on control centrality for weighted, directed, and signed networks with control inputs has been conducted (see [13] and [26]), these measures face limitations in dealing with network uncertainty, a common feature in biological networks. Furthermore, existing measures are solely based on network dynamics and do not adequately adapt to different control scenarios. Conversely, measures proposed in [22] and [23] consider network dynamics and performance metrics but are not suitable for directed, signed networks, or those with control inputs. To address these gaps, this paper proposes a comprehensive control centrality measure that incorporates dynamics, performance metrics, and control inputs, making it applicable across various weighted, directed, and signed network types and scenarios. In this paper, the RC-Centrality is proposed to characterize the centrality of feedback controllers in weighted, directed, and signed (WDS) networks without and in the presence of uncertainty. The proposed robust control centrality quantifies both individual and group centralities as a measure of control that each input set has on the rest of the WDS network in the presence of uncertainty (WDS+U). By introducing a Gramian-based control centrality, not only is the Lyapunov stability of closed-loop system achieved, but also the optimal value of required energy for each input set to place a target set states in their steady state from initial time to final time is minimized. An input set is a set of independent control agents (e.g., controlling genes) which receive input signals to steer target set, a set of agents that cannot be

controlled directly due to their biological characteristics, to desired states (e.g., gene concentrations).

Table 1 shows the comparison between our control centrality measure with established ones.

Table 1: Comparison of the proposed control centrality with existing centrality measures

Centrality measures	Considers dynamics of network	Applicable to weighted graph	Applicable to directed graph	Applicable to signed graph	Considers uncertainty in model	Considers effect of input signal(s)	Extendable to group centrality	Network & performance-driven measure
[13]	✓	✓	✓	✓	✖	✓	✓	✖
[26]	✓	✓	✓	✓	✖	✓	✖	✖
[20]	✓	✖	✓	✖	✖	✓	✖	✖
[22], [23]	✓	✓	✖	✖	✓	✖	✖	✓
[21]	✓	✖	✓	✖	✓	✓	✖	✖
[27], [28]	✖	✓	✓	✖	✖	✖	✖	✖
[29]	✖	✖	✓	✖	✖	✖	✓	✖
[30], [11], [31], [6]	✖	✓	✖	✖	✖	✖	✖	✖
[3]	✖	✖	✓	✖	✖	✖	✖	✖
[32], [4]	✖	✖	✖	✖	✖	✖	✓	✖
[33]	✓	✖	✖	✖	✖	✖	✓	✖
[2], [34], [35], [5]	✖	✖	✖	✖	✖	✖	✖	✖
Our measure	✓	✓	✓	✓	✓	✓	✓	✓

C. Paper Organization

The rest of this paper is organized as follows: In Section 2, the preliminaries of graph theory, network analysis, and a linear time-invariant (LTI) model of WDS genetic networks are represented. In Section 3, the RC-Centrality bounded by the eigenvalues of a proposed Gramian, is derived. Section 4 is about a control centrality algorithm based on the mathematical relation to stability conditions. Several numerical examples in matched and unmatched structures of uncertainties are simulated in Section 5 to illustrate the effectiveness of the proposed approach. The paper is summarized in Section 6, where future research directions are stated.

Preliminaries

The set of $n \times m$ matrices with real entries is denoted by $R^{n\times m}$. Given a matrix A, the symbol A_{ij} denotes the (i,j)-th entry of A, while A^T and A^{-T} mean the transpose and the inverse of the transpose of A.

A generic WDS network is denoted by a graph $G = (V,E,A)$ where $V = \{v_1, \dots, v_n\}$ and $E \subset V \times V$ are the set of agents and the set of links, respectively, where a link is drawn from v_i to v_j, as $e_{ij} \in E$, if the agent v_j interacts with the agent v_i. In other words, if v_j influences v_i through its dynamics. In addition, $a(e_{ij}), : E \to R$ also denoted by a_{ij}, is the strength of the link e_{ij}, which captures the significance of the agent v_j to v_i.

Moreover, for simplicity, it is defined that $a_{ij} = 0$ if $e_{ij} \notin E$ and $a_{ii} = 0$, $\forall i = 1, \dots, n$. Then, the adjacency matrix $A \in R^{n\times n}$ comprised of coefficients a_{ij} is used to describe the network topology. For each agent v_i, a_i is defined as the total significance of other agents to v_i as [8]:

$$a_i = \sum_{j=1}^{n} a_{ij}\,.$$

The $n \times n$ Laplacian matrix of G, denoted by L, is defined as

$$L = \mathcal{D} - A$$

in which $\mathcal{D}$ is the degree matrix of graph G as [36]:

$$\mathcal{D}_{ij} = \begin{cases} a_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The mathematical state of each agent $i \in V$ is described by a scalar state variable $x_i(t)$ at time $t \geq 0$. It is assumed that only an arbitrary, but fixed, subset $S \subset V$, which is the input set, receives control inputs. The control inputs can then be represented as:

$$u(t) = [u_1(t) \dots u_n(t)]^T$$

in which $u(t) \in R^n$ represents all possible control inputs, while $B \in R^{n\times n}$ restrict them to the subset S, where only a selected subset of control inputs is active. The input matrix B defines the agents in which the activated control inputs are injected. The structure of matrix B may vary depending on the location of the activated control inputs in the network.

RC-Centrality

Consider the following class of uncertain LTI systems with the state vector $x(t) \in R^n$ and the adjacency matrix $A \in R^{n\times n}$.

$$\dot{x}(t) = Ax(t) + Bu(t) + \Delta(p)v(t) \tag{1}$$

The control input is partitioned into two parts: 1) $u(t) \in R^n$ is the "certain" part, which affects the system via matrix $B \in R^{n\times n}$ that has no uncertainty, and 2) $v(t) \in R^n$ is the second part that enters the system via the bounded uncertainty $\Delta(p) \in R^{n\times n}$, where $p \in P$ is an uncertain parameter vector, and no uncertainty is imposed upon A and B matrices.

It is also assumed that the pair (A,B) is stabilizable and controllable.

A. Robust Control Energy level for a Network

In the biological context, control input could be a regulatory signal coming from outside of the considered process or an externally supplied agent applied to a cell. The process of applying control input into a cell by an experimenter can be very expensive and time consuming [24]. Thus, in this subsection, the cost function is proposed due to the importance of the control inputs in GRNs. An energy-based quantity is defined to consider the effect of uncertainty on the consumed energy of the WDS+U network $\mathcal{G}$.

Definition 1. The robust control energy level is defined as follows that tries to be a reflection of the consumed energy of uncertain LTI system (1) from the initial time $t_0 = 0$ to the final time $t_f \to \infty$:

$$J_R = \frac{1}{2}\int_0^\infty (u^T(t)Ru(t) + v^T(t)Qv(t))dt \tag{2}$$

where $R,Q \in R^{n\times n}$ are symmetric and positive definite weighting matrices.

RC-Centrality for an Input Set

Uncertainty is a significant factor influencing the consumed energy to steer a network from its initial states to desired final states. To capture this influence, the following control centrality measure is introduced for the activated input set S in the WDS+U network $\mathcal{G}$.

Definition 2. For a WDS+U network $\mathcal{G} = (\mathcal{V},\mathcal{E},A)$ described by (1), the *RC-Centrality* of input set S from the initial state $x(t_0)$ at $t_0 = 0$ to the given final state $x(t_f)$ at $t_f \to \infty$ is defined as

$$C_{RC}(S) = \frac{1}{J_R^*(B)} \tag{3}$$

where $J_R^*(B)$ is the optimal value of robust control energy level (2). The diagonal matrix B is designed in such a way that S is the activated input set. Only nonzero elements in matrix B are selected according to the set S to receive control inputs.

Energy-based Characterization of C_{RC}

In this subsection, we relate the optimal value of the robust control energy level (2) to the RC-Centrality by an auxiliary system. Consider the following LTI system:

$$\dot{x}(t) = Ax(t) + Bu(t) + \gamma(I - BB^\dagger)v(t) \tag{4}$$

where $B^\dagger \in R^{n\times n}$ is pseudo-inverse matrix of B defined as $B^\dagger = (B^T B)^{-1} B^T$ such that $B^T B$ is non-singular and γ is a design parameter which satisfies an inequality to guarantee the stability of the uncertain LTI system in (1) characterized later on in Section 4.2.

In the next theorem, it is proved that characterizing $C_{RC}(S)$ for the uncertain LTI system in (1) is equivalent to finding the inverse of the optimal value of input energy in (2), J_R^*, for the auxiliary system described as (4).

Theorem 1. Consider the uncertain LTI system described as (1). For any input set S, the robust control centrality $C_{RC}(S)$ can be expressed as:

$$C_{RC}(S) = \left(x(t_f) - e^{At_0}x(t_0)\right)^{-1}\left(\int_0^\infty e^{A\tau}\left[BR^{-1} + \gamma^2(I - BB^\dagger)Q^{-1}(I - BB^\dagger)^T\right]e^{A^T\tau}d\tau\right)\left(x(t_f) - e^{At_0}x(t_0)\right)^{-T} \tag{5}$$

Proof. According to the Hamilton-Jacobi-Bellman (HJB) equation for optimality [19], the Hamiltonian form for the auxiliary system (4) based on the cost function (2) is represented as

$$H(t) = \frac{1}{2}(u^T Ru + v^T Qv) + \lambda^T(Ax + Bu + \gamma(I - BB^\dagger)v)$$

where $\lambda(t) \in R^n$ is a Lagrange multiplier. Using the stationary conditions, the optimal control laws are obtained as:

$$\begin{aligned} u^*(t) &= R^{-1}B^T e^{A^T t}\left(\int_0^\infty e^{A\tau}\left[BR^{-1}B^T + \gamma^2(I - BB^\dagger)Q^{-1}(I - BB^\dagger)^T\right]e^{A^T\tau}d\tau\right)^{-1}\left[x(t_f) - e^{At_0}x(t_0)\right], \\ v^*(t) &= \gamma Q^{-1}(I - BB^\dagger)^T e^{A^T t}\left(\int_0^\infty e^{A\tau}\left[BR^{-1}B^T + \gamma^2(I - BB^\dagger)Q^{-1}(I - BB^\dagger)^T\right]e^{A^T\tau}d\tau\right)^{-1}\left[x(t_f) - e^{At_0}x(t_0)\right] \end{aligned} \tag{6}$$

The optimal value of robust control energy level affected by the optimal control laws in (6) is determined as

$$J_R^*(B) = \left[x(t_f) - e^{At_0}x(t_0)\right]^T \left(\int_0^\infty e^{A\tau} \left[BR^{-1}B^T + \gamma^2(I - BB^\dagger)Q^{-1}(I - BB^\dagger)^T\right] e^{A^T\tau} d\tau \right)^{-1} \left[x(t_f) - e^{At_0}x(t_0)\right] \quad (7)$$

which is the inverse of robust control centrality $C_{RC}(S)$, that completes the proof. ■

Remark 1. The proposed control centrality is consistent with physical interpretation in the sense that the higher the $C_{RC}(S)$ of input set S, the less input energy it consumes to steer the WDS+U network $\mathcal{G}$ to the given final states.

Remark 2. The proposed control centrality incorporates the dynamics of the network and the performance metric. Even with the fixed structure of the network, by changing the weighting matrix of the control inputs, the centrality of node or set of nodes will change accordingly.

Encapsulation

The robust reachability Gramian is defined as

$$G_R(B) = \int_0^\infty e^{A\tau} \left[BR^{-1}B^T + \gamma^2(I - BB^\dagger)Q^{-1}(I - BB^\dagger)^T\right] e^{A^T\tau} d\tau \quad (8)$$

This is in fact the reachability Gramian of [37] modified to uncertain system (1) from the initial time t_0 to the final time t_f. The final state difference vector $d_f = \left[d_{f_1} \dots d_{f_n}\right] \in R^n$ is defined as the difference between the final states and the zero-input solution as

$$d_f = x(t_f) - e^{At_0}x(t_0) \quad (9)$$

Based on Definition 2 and (5), the control centrality measure $C_{RC}(S)$ defined in (3) can be encapsulated as follows:

$$C_{RC}(S) = d_f^{-1}G_R(B)d_f^{-T} \quad (10)$$

Remark 3. Considering the challenges in applying control input, especially to genetic networks, such as the high cost of practical biological experiments, it is desirable to determine whether a selected input set S is central to control the WDS+U network $\mathcal{G}$. The concept of proposed control centrality concerns the optimal amount of required input energy in the controllable LTI system to steer the initial states to their final states.

Next, some lower and upper bounds for $C_{RC}(S)$ is established.

Bounds of RC-Centrality

The following theorem is provided to drive theoretical bounds of the RC-Centrality of (10). Lemma 1 is needed to later prove the theorem.

Lemma 1. [38] If $q(X) = X^TAX$ is a quadratic form in $X = [x_1, \dots, x_n]^T$ for a square matrix A, then there exists an invertible orthogonal matrix F such that

$$q(X) = \lambda_1\tilde{x}_1^2 + \cdots + \lambda_n\tilde{x}_n^2$$

where

$$\begin{bmatrix}\tilde{x}_1 \\ \vdots \\ \tilde{x}_n\end{bmatrix} = F^T \begin{bmatrix}x_1 \\ \vdots \\ x_n\end{bmatrix}$$

and $\{\lambda_1, \dots, \lambda_n\}$ is the set of A's eigenvalues. In the following theorem, the n eigenvalues of $G_R(B)$ are denoted, from the lowest to the highest, as

$$\lambda_{min}\left(G_R(B)\right) \le \cdots \le \lambda_{max}\left(G_R(B)\right).$$

Theorem 2. Consider the controllable LTI system in the presence of uncertainty in (1). The control centrality measure $C_{RC}(S)$ is bounded below and above as follows:

$$\sum_{i=1}^n \frac{\lambda_{min}\left(G_R(B)\right)}{\tilde{d}_{f_i}^2} \le C_{RC}(S) \le \sum_{i=1}^n \frac{\lambda_{max}\left(G_R(B)\right)}{\tilde{d}_{f_i}^2}$$

where

$$\begin{bmatrix}\tilde{d}_{f_1} \\ \vdots \\ \tilde{d}_{f_n}\end{bmatrix} = F^T \begin{bmatrix}d_{f_1} \\ \vdots \\ d_{f_n}\end{bmatrix}$$

for the existing orthogonal matrix F.

Proof. According to the Lemma 1 and the definition of $C_{RC}(S)$, we have:

$$\begin{aligned} C_{RC}(S) &= d_f^{-1}G_R(B)d_f^{-T} \\ &= \lambda_{min}\left(G_R(B)\right)\left(\tilde{d}_{f_1}^{-T}\right)^2 + \cdots \\ &+ \lambda_{max}\left(G_R(B)\right)\left(\tilde{d}_{f_n}^{-T}\right)^2 \end{aligned}$$

This can be expressed in matrix form as

$$C_{RC}(S) = \left[\lambda_{min}\left(G_R(B)\right) \quad \dots \quad \lambda_{max}\left(G_R(B)\right)\right] \left[\frac{1}{\tilde{d}_{f_1}^2} \quad \dots \quad \frac{1}{\tilde{d}_{f_n}^2}\right]^T$$

So, the bounds for $C_{RC}(S)$ can be found, which are functions of $\lambda_{min}(G_R(B))$ and $\lambda_{max}(G_R(B))$ as follows:

$$\lambda_{min}\left(G_R(B)\right)\sum_{i=1}^n \frac{1}{\tilde{d}_{f_i}^2} \le C_{RC}(S) \le \lambda_{max}\left(G_R(B)\right)\sum_{i=1}^n \frac{1}{\tilde{d}_{f_i}^2}$$

■

The bounds on $C_{RC}(S)$ for input set S could be helpful, especially in the case that characterizing the exact value of $C_{RC}(S)$ is computationally too complex. In the following corollary, the worst-case energy is considered that the WDS+U network $\mathcal{G}$ consumes the maximum amount of input energy to steer initial states to final states for matrix B.

Corollary 1. For any input set S, the upper bound on the control energy of $C_{RC}(S)$ is proportional to the inverse of the smallest eigenvalue of $G_R(B)$.

Proof. According to the definition of $C_{RC}(S)$, the bounds of input energy function are exactly the same as the method proposed in Theorem 2 as follows:

$$\frac{\sum_{i=1}^{n} \tilde{d}_{f_i}^2}{\lambda_{max}\big(G_R(B)\big)} \leq d_f^T G_R^{-1}(B) d_f \leq \frac{\sum_{i=1}^{n} \tilde{d}_{f_i}^2}{\lambda_{min}\big(G_R(B)\big)} < \infty$$

In the worst-case, the upper bound on the input energy would be obtained as

$$\max_{S} d_f^T G_R^{-1}(B) d_f \propto \lambda_{min}^{-1}\big(G_R(B)\big)$$ ■

The importance of the bounds is that for a controllable network in theory, the network may still not be controllable in practice. For instance, unreasonable values of the input energy are required to steer biological complex systems in some practical direction [39].

Special Case: In the Absence of Uncertainty

We compare the proposed control centrality with other known standard centrality measures (degree, closeness, betweenness, and eigenvector centrality) in Case Study 1. Therefore, the control centrality is defined based on the RC-Centrality's definition to quantify the impact of input set S in the WDS network $\mathcal{G}$ to compare other measures.

Definition 3. For a WDS network $\mathcal{G} = (\mathcal{V},\mathcal{E},A)$ described by the LTI controllable system in (11) as:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{11}$$

The, the control centrality of input set S is defined as

$$C_C(S) = C_{RC}(S) - d_f^{-1}\left(\int_0^{\infty} e^{At}\gamma^2(I - BB^{\dagger})Q^{-1}(I - BB^{\dagger})^T e^{A^T t} dt\right) d_f^{-T} \tag{12}$$

The value of $C_C(S)$ can be encapsulated in the following form:

$$C_C(S) = d_f^{-1} G(B) d_f^{-T} \tag{13}$$

where the reachability Gramian is represented as [37]

$$G(B) = \int_0^{\infty} e^{A\tau} BR^{-1}B^T e^{A^T\tau} d\tau \tag{14}$$

Remark 4. In the control centrality defined in (13), the importance of input set S is the inverse of the optimal value of input energy $J = \frac{1}{2}\int_0^{\infty} u^T(t)Ru(t)dt$ to steer the initial state $x(t_0)$ at $t_0 = 0$ to their desired state $x(t_f)$ at $t_f \to \infty$ through input set S.

Calculation of RC-Centrality

In the previous section, we introduced a novel control centrality measure designed for systems operating under uncertainty. The RC-Centrality emerged as the optimal solution for energy optimization in a scenario without uncertainty, a solution that we demonstrated to be applicable to systems in presence of uncertainty, as proven in the preceding section. In this section, we delve into the role of robust controllers in ensuring system stability. We establish that the control centrality of these robust controllers, acting as driver nodes, provides valuable insights into their placement while simultaneously guaranteeing the Lyapunov stability of the closed-loop system across individual or grouped inputs.

A. A Lyapunov-based Solution for $G_R(B)$

The first step in calculating $C_{RC}(S)$ in (10) is characterizing $G_R(B)$ formulated in the following theorem.

Theorem 3. Consider a WDS+U network $\mathcal{G} = (\mathcal{V},\mathcal{E},A)$ described by (1). Suppose that there exists a design parameter γ such that satisfies (17). Then, the robust reachability Gramian, $G_R(B)$ is the unique solution that is positive definite of the following Lyapunov equation:

$$\begin{aligned} AG_R(B) + G_R(B)A^T &+ (BR^{-1}B^T \\ &+ \gamma^2(I - BB^{\dagger})Q^{-1}(I - BB^{\dagger})^T) \\ &= 0 \end{aligned} \tag{15}$$

Proof. The solution to the Lyapunov equation in (15) is obtained using the robust reachability Gramian in (8).

$$\begin{aligned} G_R(B) + G_R(B)A^T &= e^{At}[BR^{-1}B^T \\ &+ \gamma^2(I \\ &- BB^{\dagger})Q^{-1}(I - BB^{\dagger})^T]e^{A^T t}\Big|_{t=0}^{\infty} \end{aligned}$$

It is clear that if A is Hurwitz, then $\lim_{t\to\infty} e^{At} = 0$. Therefore, $G_R(B)$ is the solution of the Lyapunov equation in (15). The proof of the uniqueness of $G_R(B)$ as the solution is straightforward and omitted. Consequently, $G_R(B)$ is the unique solution of the Lyapunov equation in (15), thereby completing the proof. ■

A necessary and sufficient condition for the analysis of Lyapunov stability of systems in the presence of uncertainty is the existence and uniqueness of solution to Lyapunov equation [19]. The result of the theorem is used in the next subsection to prove that the condition for the Lyapunov stability of the uncertain LTI system in (1) is satisfied.

The γ Condition on Robust Stability

A main step in measuring the RC-Centrality indices for input set S in (10) is finding design parameter γ. The result of Lemma 2 can be used to characterize γ, which is done next.

Lemma 2. Assume that $G_R(B)$ is the unique solution, which is positive definite, of the Lyapunov equation (15) for the LTI system with uncertainty in (1). Then, the optimal control laws are obtained as follows:

$$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} R^{-1}B^T G_R(B) \\ \gamma Q^{-1}(I - BB^{\dagger})^T G_R(B) \end{bmatrix} x(t) = \begin{bmatrix} K \\ L \end{bmatrix} x(t) \quad (16)$$

Proof. The proof is very similar to that of Theorem 1 and is thus removed. ■

The coefficients of state feedback K and L are now utilized to find a suitable value for γ.

Theorem 4. Consider the controllable LTI system in (1) with bounded uncertainty $\Delta(p)$. The robust control centrality $C_{RC}(S)$ indicates the centrality of controllers $u(t) = Kx(t)$ activated by input set S which robustly stabilizes the closed-loop system (1) for all $p \in P$. The design parameter γ and feedback gains K and L can be found in such a way to satisfy the following inequality:

$$\text{inv}(\gamma) L^T Q \Delta(p) L > -\frac{1}{2}(K^T R K - L^T Q L + 2K^T R B^{\dagger} \Delta(p) L) \quad (17)$$

Proof. Suppose that the robust control energy level defined in (2) is the Lyapunov function candidate for the uncertain system in (1) of the following form:

$$V(x) = \min_{u,v} \int_0^{\infty} (u^T(t) R u(t) + v^T(t) Q v(t)) \mathrm{d}t \quad (18)$$

the HJB equation is satisfied by the proposed $V(x)$, which reduces to

$$\min_{u,v} \left(u^T(t) R u(t) + v^T(t) Q v(t) + V_x^T (Ax(t) + Bu(t) + \gamma(I - BB^{\dagger}) v(t)) \right) \quad (19)$$

where $V_x \triangleq \frac{\partial V(x)}{\partial x}$. The optimal control laws $u(t) = Kx(t)$ and $v(t) = Lx(t)$ must make the derivative of $u^T(t)Ru(t) + v^T(t)Qv(t) + V_x^T(Ax(t) + Bu(t) + \gamma(I - BB^{\dagger})v(t))$ zero with respect to $u(t)$ and $v(t)$, respectively [18], [19]. Hence,

$$2x^T(t) K^T R + V_x^T B = 0 \quad (20)$$

$$2x^T(t) L^T Q + V_x^T \gamma (I - BB^{\dagger}) = 0 \quad (21)$$

It is clear that the integral type of Lyapunov function $V(x)$ proposed by (18) is a positive definite function for system in (1):

$$V(x) > 0 \quad x \neq 0$$
$$V(x) = 0 \quad x = 0$$

It will be shown that $\dot{V}(x) < 0$ for all $x \neq 0$ using (1). According to the HJB equation (19) and Lemma 2, we have:

$$\dot{V}(x(t)) = -x^T K^T R K x - x^T L^T Q L x + V_x^T B B^{\dagger} \Delta(p) L x + V_x^T (I - BB^{\dagger}) \Delta(p) L x - V_x^T \gamma (I - BB^{\dagger}) L x$$

which, combined with (20) and (21), implies that

$$\dot{V}(x) = -x^T K^T R K x + x^T L^T Q L x - 2x^T K^T R B^{\dagger} \Delta(p) L x - 2\gamma^{-1} x^T L^T Q \Delta(p) L x \quad (22)$$

and consequently,

$$\dot{V}(x) = x^T(-K^T R K + L^T Q L - 2K^T R B^{\dagger} \Delta(p) L - 2\gamma^{-1} L^T Q \Delta(p) L) x \quad (23)$$

Equation (23) has the form of $x^T H(p) x$, then by (17)

$$\dot{V}(x) < 0 \quad x \neq 0$$
$$\dot{V}(x) = 0 \quad x = 0$$

Then, it is clear that $V(x)$ is a Lyapunov function for system (1). Therefore, according to the Lyapunov stability theorem, the system in (1) using the optimal control laws $u(t) = Kx(t)$ and $v(t) = Lx(t)$ for all $p \in P$ would be asymptotically stable. ■

Corollary 2. For any input set $S \subseteq V$, the Lyapunov candidate function is the optimal value of robust control energy level to achieve the following property:

$$V(x) = J_R^*(B) = \frac{1}{C_{RC}(S)} \quad (24)$$

This shows the impact of robust controller $u(t)$ activated via matrix B, which holds the above property. In the next subsection, an algorithm is proposed for calculating the control centrality $C_{RC}(S)$ of input set S.

Algorithm to Calculate C_{RC}

The proposed control centrality measure is summarized as Algorithm 1. It should be noted that the first iteration starts with the nominal value of uncertainty. Moreover, during the calculation of RC-Centrality Algorithm, the optimal controllers $u(t)$ and $v(t)$ in (16) stabilize the LTI system (1) in the presence of uncertainty.

Algorithm 1: Robust Control Centrality

Input: Input set S, A, $B_{\Delta}(p)$, t_0, t_f, $x(t_0)$, $x(t_f)$
Output: $C_{RC}(S)$ or $C_C(S)$

1. Pick $s \in S$.
2. Compute d_f using (9).
3. If uncertainty exists % RC-Centrality
   3.1. Decompose $B_{\Delta}(p)$ into B and $\Delta(p)$ using s and (1).
   3.2. Find $G_R(B)$ and γ using (16) subject to (15) and (17).
   3.3. Calculate $C_{RC}(s)$ using (10).
   3.4. Repeat steps 3.1 to 3.3 for each individual or group set of S.
4. else % Control Centrality
   4.1. Find $G(B)$ using (14).
   4.2. Calculate $C_C(s)$ using (13).
   4.3. Repeat steps 4.1 to 4.2 for each individual or group set of S.
5. End

It is noted that step 3. 2 in Algorithm 1 could be done by convex optimization method in MATLAB. Also, in this step, metaheuristic methods such as particle swarms and genetic algorithms can be used. However, these approaches were not used in the case studies presented in this paper.

In the next section, the controllable and asymptotically stable system (1) in the presence of bounded uncertainty $\Delta(p)$ is considered. The RC-Centrality is used to identify input set's importance in the closed-loop system with robust control law $u(t)$. In this WDS+U genetic network $\mathcal{G}$, maximization of $C_{RC}(S)$ corresponds to minimizing input energy of input set S to reach the desired final state $x(t_f)$.

Numerical Examples

In this section, we present an example of a real genetic regulatory network to demonstrate the effectiveness of the proposed control centrality strategy using Algorithm 1. The GRN, consisting of five nodes, is relatively simple but serves to illustrate the method's effectiveness in various scenarios, including the absence and presence of uncertainty with different structures. Additionally, we evaluate the feasibility of numerical computation in MATLAB for the parameters γ and $G_R(B)$.

Let $\mathcal{A} = C = -\text{diag}(1)_{5\times5}$, $D = \text{diag}(0.8)_{5\times5}$, and the Hill function is $f(x) = \frac{x^2}{(1+x^2)}$. (Robust) control centrality is calculated to transfer the states of the linear GRN system in (11) from the initial concentrations of protein and mRNA, $M(0) = P(0) = [5]_{5\times1}$, to the final states at the origin. In this part, it is assumed that control input $u_i(t)$ is a repressor transcription factor of gene i to prevent the target genes from being upregulated [40].

A. Case Study 1: A Non-WDS Network

A case study is here given to show the differences and similarities between the control centrality and standard centrality measures such as degree, closeness, betweenness and eigenvector centrality for unweighted and undirected networks. Consider a GRN without loops, shown in Fig. 1.

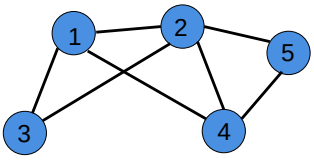


Fig. 1: The GRN model of Case Study 1.

The coupling matrix is $G_1 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$.

The control centrality in (13) is compared with the well-known aforementioned centrality measures in Tables 2 and 3.

According to Tables 2 and 3, the proposed control centrality agrees with the previous individual centrality measures for simple (non WDS) networks. Node 2 is more central, and nodes 1 and 4 have the same centrality based on degree, closeness, betweenness, eigenvector, and the proposed control centrality measures. Table 2 presents the order of individual and group centralities to illustrate the similarity and difference between the control centrality and other established measures.

Table 3: Priorities based on various centrality measures

Order of Centralities
$C_D(\{1,2,4\}) = C_D(\{2,4\}) = C_D(\{1,2\}) = C_D(\{1,4\}) = C_D(\{2\}) > C_D(\{1\}) = C_D(\{4\})$
$C_{cl}(\{1,2,4\}) > C_{cl}(\{2,4\}) = C_{cl}(\{1,2\}) = C_{cl}(\{1,4\}) > C_{cl}(\{2\}) > C_{cl}(\{1\}) = C_{cl}(\{4\})$
$C_B(\{1,2,4\}) > C_B(\{2,4\}) = C_B(\{1,2\}) > C_B(\{2\}) > C_B(\{1,4\}) > C_B(\{1\}) = C_B(\{4\})$
$C_E(\{2\}) > C_E(\{1\}) = C_E(\{4\})$
$C_C(\{1,2,4\}) > C_C(\{2,4\}) = C_C(\{1,2\}) > C_C(\{1,4\}) > C_C(\{2\}) > C_C(\{1\}) = C_C(\{4\})$

Table 2: Scores based on various individual and group centrality methods

	Input set S	$C_D(S)$	$C_{cl}(S)$	$C_B(S)$	$C_E(S)$	$C_C(S)$
Individual Centrality	{1}	0.75	0.2	0.14	3.1	4.5
	{2}	1	0.25	0.37	3.7	4.7
	{4}	0.75	0.2	0.14	3.1	4.5
Group Centrality	{1,2}	1	0.33	0.66	N.A.	5.5
	{1,4}	1	0.33	0.25	N.A.	5.3
	{2,4}	1	0.33	0.66	N.A.	5.5
	{1,2,4}	1	0.5	1	N.A.	6

N.A.: Not applicable

It is clear from Table 3 that based on control centrality, $C_C(\{1,2,4\})$ also gets the highest score as other methods to prevent upregulation of target set $\{3,5\}$. The degree centrality scores of input sets $\{1,2\}$,$\{1,4\}$,$\{2,4\}$,$\{1,2,4\}$ are all equal to 1 and the scores of the input sets $\{1,2\}$,$\{1,4\}$,$\{2,4\}$ are equal per the group closeness centrality. However, the groups which contain individual with high centrality score, $\{1,2\}$,$\{2,4\}$, inherit some of these scores in group betweenness and input centrality than the other group centrality $\{1,4\}$.

Case Study 2: A Real Genetic Network

In this subsection, we consider the application of control centrality to a real genetic regulatory network (GRN), which has been both theoretically predicted and experimentally validated in *Escherichia coli* [41]. We specifically focus on the dynamics of the repressilator, a well-studied system. The repressilator is composed of three repressor genes ($lacl$, $tetR$, and cl) along with their corresponding promoters.

Consider the WDS genetic network with five genes drawn in Fig. 2 [41], in which each agent describes a gene. The $\downarrow$ sign denotes an activation link, and the $\perp$ sign denotes a repression link. The coupling matrix is

$$G_2 = 0.5 \times \begin{bmatrix} 0 & -0.5 & 0.75 & 0 & 0 \\ -0.5 & 0 & 0 & 0.5 & 1 \\ 0 & 0.75 & 0 & 0 & 0 \\ 1 & -0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

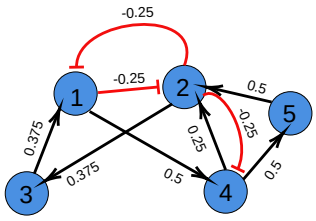


Fig. 2: Genetic regulatory network model [41].

This network has the same nodes and connections, as the one in previous case study, but weights and directions and signs are also added.

The individual and group control centralities for this network are calculated using (13) and are depicted in descending order in Fig. 3.

The target set $\{3,5\}$ in this genetic network will be paid attention to again. Based on control centrality investigation in Case Study 1, input nodes $\{1\}$ and $\{4\}$ gain the same centrality measures, and it can't be determined which one is more important. However, in this example, input node $\{1\}$ has a higher rank than $\{4\}$ since, according to the coupling matrix G_2, tf_1 has a more repression influence over all other genes in this network topology rather than tf_4, which explains why input node $\{1\}$ requires less input energy to steer the system towards the desired final state.

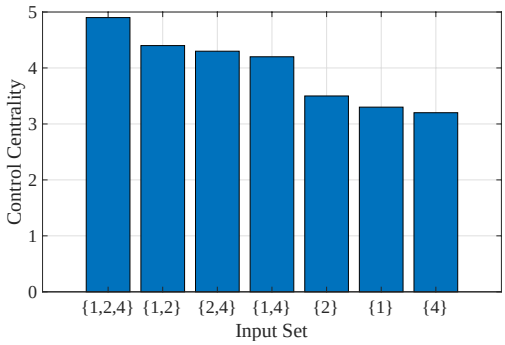


Fig. 3: Control Centrality Without Uncertainty.

In comparison between the control centrality measures for weighted, directed, and signed networks considering the effect of control inputs proposed in [13] and [26] and our measure, we use Table 4 to demonstrate the similarities between them.

In the following subsection, the control centrality for GRN would be calculated in the presence of uncertainty.

Table 4: Priorities of control centrality measures for the WDS network

Order of Centralities
$C_C(\{1,2,4\}) > C_C(\{12\}) > C_C(\{24\}) > C_C(\{1,4\}) > C_C(\{2\}) > C_C(\{1\}) > C_C(\{4\})$
$C_p(\{1,2,4\}) > C_p(\{12\}) > C_p(\{24\}) > C_p(\{1,4\}) > C_p(\{2\}) > C_p(\{1\}) > C_p(\{4\})$
$C_q(\{1,2,4\}) > C_q(\{12\}) > C_q(\{24\}) > C_q(\{1,4\}) > C_q(\{2\}) > C_q(\{1\}) > C_q(\{4\})$
$C_W(\{2\}) > C_W(\{1\}) > C_W(\{4\})$
$C_M(\{2\}) > C_M(\{1\}) > C_M(\{4\})$

RC-Centrality in the Genetic Network

Genetic regulatory networks often contain a significant amount of uncertainty. This uncertainty is categorized into two broad classes. Matched uncertainty stems from the lack of knowledge or inherent biological phenomena in the intrinsic dynamics of the system under examination without input. In GRNs, for example, noise in gene expression can induce uncertainty in the products. The second type of uncertainty results from the applied input to the system. The unmatched uncertainty is profoundly present due to practical limitation and measurement errors or noises of the inputs in various experimental experiences, such as a moderate change of the input in a transition region [25].

In this part, the uncertainty should be decomposed into matched and unmatched parts defined mathematically next. It is proven that the unmatched uncertainty needs more input energy to satisfy the condition (17) in order to guarantee robust stability. This condition depends on the design parameter γ that can be found by computer programming, e.g., the MATLAB toolbox and Yalmip.

Case Study 3: Matched Uncertainty

If the uncertainty $\Delta(p)$ can be written in the form of $B\varphi(p)$ for some $\varphi(p)$, it can be considered as the matched uncertainty. In this subsection, matched uncertainty is applied to the model using degree structure for the uncertainty $\Delta(p)$.

The degree uncertainty is a matched uncertainty form that can be considered in RC-Centrality. In this case, the dynamics of the state space model are given by

$$\dot{x}(t) = Ax(t) + Bu(t) + \mathcal{D}(p)v(t), p \in [-1,1] \qquad (25)$$

where p is an uncertain parameter vector, and $D(p) = p.\text{diag}([d_1 \dots d_5])$ is the degree matrix of the underlying coupling graph of the GRN. In this case study, the matrix $D(p)$ is $D(p) = p.\text{diag}([0.125,0.5,0.375,0.25,0.5])$ and a design parameter γ for the input set S, that satisfies condition (17), can be obtained using Yalmip.

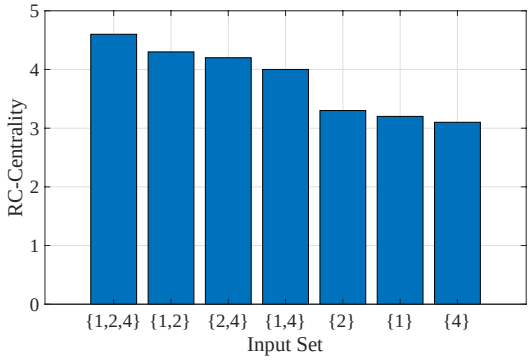


Fig. 4: RC-Centrality for the Degree Uncertainty.

According to the RC-Centrality depicted in Fig. 4, parameter γ is not sufficiently large to change the priority of $C_{RC}(S)$ compared to $C_C(S)$ in Fig 3. These plotted figures reflect that the value of $C_C(S)$ of each input set is larger than $C_{RC}(S)$ obtained for the uncertain model, that is since a cost applies to deal with uncertainty.

As a consequence of the above case study, the matched uncertainty cannot change the order of priority for the robust control centrality. The case of unmatched uncertainty will be examined next.

Unmatched Uncertainty

In this section, two types of unmatched uncertainty $\Delta(p)$ are considered, which cannot be written in the form of $B\varphi(p)$, and the RC-Centrality would be calculated for each case.

Case Study 4: Laplacian Uncertainty

The second type is Laplacian uncertainty. The updated equation of the state space model can now be described as follows:

$$\dot{x}(t) = Ax(t) + Bu(t) + L(p)v(t), p \in [-1,1] \qquad (26)$$

where $L(p) = p.L$ is the Laplacian uncertainty matrix. The value of the RC-Centrality can be computed and drawn in Fig. 5.

According to Fig. 5, while agents 2 and 1 are the most individual central agents responding to the matched uncertainty, respectively, they are the least central agents in the presence of Laplacian uncertainty. This observation can be explained as the following: since the non-zero entries of unmatched uncertainty of each input set are larger, the input set needs more energy to deal with the uncertainty by activated input set, which is compatible with our simulation results to characterize design parameter γ and feedback gains K and L in Theorem 4.

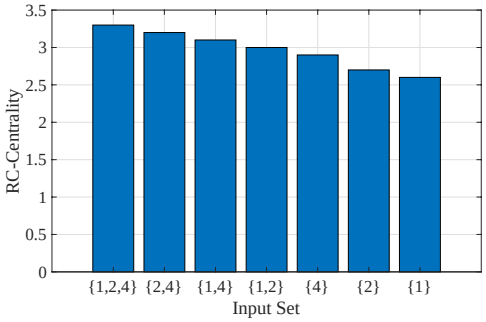


Fig. 5: RC-Centrality for the Laplacian Uncertainty.

Case Study 5: Adjacency Uncertainty

The third type of uncertainty can be considered due to the communication with its regulatory neighbors of the input set. The following relation is taken into account:

$$G(p) = G_{rep}(p) + G_{act}(p)$$

where, $G_{act}(p)$ and $G_{rep}(p)$ are the repression and activation uncertainties, respectively, defined as:

$$G_{act}(p) = \begin{cases} p\alpha_{ij} & \text{if } e_{ij} > 0 \\ 0 & \text{otherwise} \end{cases},$$

$$G_{rep}(p) = \begin{cases} -p\alpha_{ij} & \text{if } e_{ij} < 0 \\ 0 & \text{otherwise} \end{cases}$$

This uncertainty model applies to the adjacency matrix of the underlying coupling of the GRN.

$$\dot{x}(t) = Ax(t) + Bu(t) + G(p)v(t), p \in [-1,1] \qquad (27)$$

with regard to the adjacency uncertainty $G(p)$, the RC-Centrality for all input sets is depicted in Fig. 6.

It is worth noting that while Fig. 3-5 imply that input set $\{1,2,4\}$ is the most central set, it can be seen from Fig. 6 that based on the RC-Centrality, input node $\{2\}$ is the most central set with respect to the proposed type of unmatched uncertainty. That is due to its position with other genes in the network, i.e., input node $\{2\}$ has many repression links with both the target set and other agents to steer states to the origin and consumes less energy based on Theorem 4 to deal with the adjacency unmatched uncertainty matrix entries. However, it is observable that the input set $\{1,2,4\}$ needs much more input energy to reach the desired states, which can be deduced from the adjacency unmatched uncertainty matrix.

According to Fig. 6, it is also concluded that the group RC-Centrality may improve or diminish compared to the individual indices depending on the structure of unmatched uncertainty, which confirms the utility of the proposed centrality measure in the presence of unmatched uncertainty.

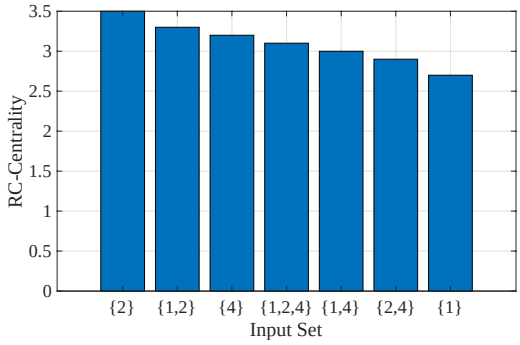


Fig. 6: RC-Centrality for the Adjacency Uncertainty.

We note that the term expressed in the last part of (12) is a function of the parameter γ and weighting matrix Q, which are designed to properly with uncertainty. In the presence of unmatched uncertainty, γ requires a larger amount of input energy to satisfy the condition (17). So, it is expected to see more energy for the unmatched uncertainty highly related to the non-zero entries of $\Delta(p)$.

Conclusion

A Robust Control Centrality measure is introduced in this paper based on an optimal value of input energy level needed for steering the states of an LTI network from the initial values at t_0 to their desired states at the final time t_f in presence of uncertainty.

In biological context, the proposed measure can be used to characterize the amount of control inputs used in controlling biological process. The proposed control centrality measure is related to the reachability Gramian of the weighted, directed, signed genetic networks that can be associated with the individual agents and agent groups. In genetic networks, the presence of inherent uncertainty poses a significant challenge in characterizing a robust control centrality. However, this challenge is addressed by adopting an optimal control approach in terms of the centrality of the robust controllers to stabilize the uncertain controllable GRN. Finally, the effectiveness of the proposed centrality measure is verified via simulation of a real GRN network under different scenarios of matched and unmatched uncertainties and comparison with established measures. The input resource allocation framework discussed in this paper is general, therefore, studying strategies for fixed budget constraints or variation in the cost of different control inputs in the network forms an interesting future research direction.

Results and Discussion

In future work, we shall introduce centrality measures as energy calculation for unstable systems with regard to the optimal energy to reach a desired steady state. Moreover, questions surrounding the terminal time and final steady states as two essential elements, which may affect the structure of the optimal energy, will be investigated.

Appendix A. An LTI model for GRNs

GRNs describe relationships between genes in a cell. They can be modeled as a WDS network [42]. The network's agents are genes, and the agent's state is the concentration of mRNA and protein of the gene expression. Each weighted and directed link in the WDS network represents power and direction of regulatory relationships, respectively. The activator or repressor transcription factor associates, respectively, with the positive or negative sign of the link in GRN's model.

A modified state space model for GRNs, based on the models provided in [41], [43] is proposed in this paper as follows:

$$\begin{aligned} \dot{m}_i(t) &= -\sigma_i m_i(t) + \sum_{j=1}^{n} b_{ij}(p_j(t)) + \mathcal{B}_i u_i(t) \\ \dot{p}_i(t) &= -c_i p_i(t) + d_i m_i(t) \qquad i = 1,2,\dots,n \end{aligned} \qquad (28)$$

where the concentration of mRNA and protein of gene i are denoted by $m_i(t)$, $p_i(t) \in R$. σ_i and c_i's are the degradation rates of mRNA and protein, respectively, and d_i is a bounded constant. Control input u_i is applied to gene i if it is activated via B_i; otherwise, it is zero. $b_{ij}(.)$ is expressed by the Hill form as:

$$b_{ij}\left(p_j(t)\right)=\begin{cases}\alpha_{ij}\frac{(p_j(t)/\beta)^H}{1+(p_j(t)/\beta)^H} & \text{if } a_{ij}>0\\ \alpha_{ij}\frac{1}{1+(p_j(t)/\beta)^H} & \text{if } a_{ij}<0\\ 0 & \text{otherwise}\end{cases} \quad (29)$$

where β is a positive constant, H is the Hill coefficient, and transcription rate α_{ij} is a bounded constant that determines the power of gene j to gene i in the regulation mechanism. The coupling matrix between genes is defined as

$$G_{ij}=\begin{cases}\alpha_{ij} & \text{if } a_{ij}>0\\ -\alpha_{ij} & \text{if } a_{ij}<0\\ 0 & \text{otherwise}\end{cases}$$

We know that:

$$\alpha_{ij}\frac{\left(p_j(t)/\beta\right)^H}{1+\left(p_j(t)/\beta\right)^H}=\alpha_{ij}-\alpha_{ij}\frac{1}{1+\left(p_j(t)/\beta\right)^H}$$

Then, (28) can be rewritten as

$$\begin{aligned}\dot{m}_i(t)&=-\sigma_i m_i(t)+\sum_{j=1}^{n}G_{ij}g\left(p_j(t)\right)+\mathcal{B}_i u_i(t)\\&\quad+l_i\\ \dot{p}_i(t)&=-c_i p_i(t)+d_i m_i(t),\qquad i=1,2,\dots,n\end{aligned} \quad (30)$$

where $l_i=\sum_{j\in V_{i1}}\alpha_{ij}$ is the basal rate, V_{i1} is the set of repressors of gene i, i.e., $a_{ij}<0$, and $g(p)=\frac{(p/\beta)^H}{1+(p/\beta)^H}$. In compact matrix form, (30) can be rewritten as

$$\begin{aligned}\dot{m}(t)&=\mathcal{A}m(t)+Gg\big(p(t)\big)+\mathcal{B}u(t)+l\\ \dot{p}(t)&=Cp(t)+Dm(t)\end{aligned} \quad (31)$$

where $G=G_{ij}\in R^{n\times n}$, $m(t)=[m_1(t),\dots,m_n(t)]^T$, $p(t)=[p_1(t),\dots,p_n(t)]^T$, $\mathcal{A}=\mathrm{diag}(-\sigma_1,\dots,-o_n)$, $\mathcal{B}=\mathrm{diag}(\mathcal{B}_1,\dots,\mathcal{B}_n)$, $C=\mathrm{diag}(-c_1,\dots,-c_n)$, $D=\mathrm{diag}(d_1,\dots,d_n)$, $l=[l_1,\dots,l_n]^T$, $g\big(p(t)\big)=\left[g\big(p_1(t)\big),\dots,g\big(p_n(t)\big)\right]^T$.

Consider an equilibrium point (m^*,p^*,u^*) of the nonlinear system (31), meaning that $\mathcal{A}m^*+Gf(p^*)+\mathcal{B}u^*+l=0$, and $Cp^*+Dm^*=0$. It will be shifted to the origin by letting $M(t)=m(t)-m^*$, $P(t)=p(t)-p^*$ [44]. Thus, we have

$$\begin{aligned}\dot{M}(t)&=\mathcal{A}M(t)+Gf\big(P(t)\big)+\mathcal{B}u(t),\\ \dot{P}(t)&=CP(t)+DM(t).\end{aligned} \quad (32)$$

where $M(t)$, $P(t)\in R^n$ and $f\big(P(t)\big)=g(P(t)+p^*)-g(p^*)$.The linearized control system at (m^*,p^*,u^*) is:

$$\begin{bmatrix}\dot{M}(t)\\ \dot{P}(t)\end{bmatrix}=\begin{bmatrix}\mathrm{A}_{11} & \mathrm{A}_{12}\\ \mathrm{A}_{21} & \mathrm{A}_{22}\end{bmatrix}\begin{bmatrix}M(t)\\ P(t)\end{bmatrix}+\begin{bmatrix}\mathcal{B}\\ 0\end{bmatrix}u(t)$$

where $A_{11}=\mathcal{A}$, $A_{12}=G\frac{\partial f}{\partial P}(m^*,p^*,u^*)$, $A_{21}=D$, $A_{22}=C$. So, an LTI model of GRNs with state variables $x(t)=\begin{bmatrix}M(t)\\ P(t)\end{bmatrix}$ is as follows:

$$\dot{x}(t)=Ax(t)+Bu(t) \quad (33)$$

where

$$A=\begin{bmatrix}\mathrm{A}_{11} & \mathrm{A}_{12}\\ \mathrm{A}_{21} & \mathrm{A}_{22}\end{bmatrix},\ B=\begin{bmatrix}\mathcal{B}\\ 0\end{bmatrix}.$$

The LTI system (33) is formed by applying control input to capture the dynamics of gene networks in an efficient way [45].

Author Contributions

S. Ozgoli and S. Bolouki served as the supervisor and advisor of the current research paper, respectively. They outlined the research framework and the roadmap and analyzed the results. Z. Ghassemi wrote the manuscript, proved the theoretical results, simulated the numerical case studies, and interpreted the results under the supervision and guidance of S. Ozgoli and S. Bolouki.

Acknowledgment

This work is entirely self-supported, and no financial support from any agency was involved.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. Furthermore, ethical issues including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been fully addressed by the authors.

Abbreviations

GRNs	Genetic Regulatory Networks
HJB	Hamilton-Jacobi-Bellman
LTI	Linear Time Invariant
WDS	Weighted Directed Signed

References

[1] P. H. Guzzi, S. Roy, Biological network analysis: trends, approaches, graphical theory and algorithms. Amsterdam: Elsevier, 2020.

[2] X. Chen, M. Xu, Y. An, "Identifying the essential nodes in network pharmacology based on multilayer network combined with random walk algorithm," J. Biomed. Inform., 114: 103666, 2021.

[3] J. Lü, P. Wang, "Identifying important nodes in bio-molecular networks," in Modeling and Analysis of Bio-molecular Networks, J. Lü and P. Wang, Eds., Singapore: Springer Singapore: 315-396, 2020.

[4] M. Ashtiani et al., "A systematic survey of centrality measures for protein-protein interaction networks," BMC Syst. Biol., 12(1), 2018.

[5] J. Wang, R. Peng, Z. Zhang, Y. Zhang, Y. Dai, Y. Sun, "Identification and validation of key genes in hepatocellular carcinoma by bioinformatics analysis," BioMed Res. Int., 6662114, 2021.

[6] H. B. Chakrapani, S. Chourasia, S. Gupta, T. Kumar D, G. P. Doss C, R. Haldar, "Effective utilisation of influence maximization technique for the identification of significant nodes in breast cancer gene networks," Comput. Biol. Med., 133, 2021.

[7] M. Newman, Networks: An Introduction. OUP Oxford, 2010.

[8] V. Latora, V. Nicosia, G. Russo, Complex networks: principles, methods and applications. Cambridge, United Kingdom; New York, NY: Cambridge University Press, 2017.

[9] S. Gómez, "Centrality in networks: Finding the most important nodes," in Business and Consumer Analytics: New Ideas, P. Moscato and N. J. de Vries, Eds., Cham: Springer International Publishing, 401-433, 2019.

[10] M. G. Everett, S. P. Borgatti, "The centrality of groups and classes," J. Math. Sociol., 23(3): 181-201, 1999.

[11] J. Tang et al., "Identification of top-k influential nodes based on enhanced discrete particle swarm optimization for influence maximization," Phys. Stat. Mech. Its Appl., 513: 477-496, 2019.

[12] G. Baggio, F. Pasqualetti, S. Zampieri, "Energy-Aware controllability of complex networks," Annu. Rev. Control Robot. Auton. Syst., 5(1): 465-489, 2022.

[13] G. Lindmark, C. Altafini, "Centrality measures and the role of non-normality for network control energy reduction," IEEE Control Syst. Lett., 5(3): 1013-1018, 2021.

[14] G. Lindmark, C. Altafini, "Combining centrality measures for control energy reduction in network controllability problems," in Proc. 2019 18th European Control Conference (ECC): 1518-1523, 2019.

[15] N. Bof, G. Baggio, S. Zampieri, "On the role of network centrality in the controllability of complex networks," IEEE Trans. Control Netw. Syst., 4(3): 643-653, 2017.

[16] G. Lindmark, C. Altafini, "Minimum energy control for complex networks," Sci. Rep., 8(1), 2018.

[17] L. Geris, D. Gomez-Cabrero, Uncertainty in Biology: A Computational Modeling Approach. in Studies in Mechanobiology, Tissue Engineering and Biomaterials. Springer International Publishing, 2015.

[18] F. Lin, Robust control design: an optimal control approach, John Wiley & Sons, 2007.

[19] Y. Lin, C. Lin, "Optimal control approach for robust control design of neutral systems," Optim. Control Appl. Methods, 30(1): 87-102, 2009.

[20] Y. Y. Liu, J. J. Slotine, A. L. Barabási, "Control centrality and hierarchical structure in complex networks," PLoS ONE, 7(9): e44459, 2012.

[21] U. Usman, A. Mahmood, L. Wang, "Robust control centrality," in Proc. 2019 Chinese Control Conference (CCC): 5486-5491, 2019.

[22] M. Siami, S. Bolouki, B. Bamieh, N. Motee, "Centrality measures in linear consensus networks with structured network uncertainties," IEEE Trans. Control Netw. Syst., 5(3): 924-934, 2018.

[23] Y. Ghaedsharaf, M. Siami, C. Somarakis, N. Motee, "Centrality in time-delay consensus networks with structured uncertainties," Automatica, 125: 109378, 2021.

[24] M. Amodio et al., "Generating hard-to-obtain information from easy-to-obtain information: Applications in drug discovery and clinical inference," Patterns, 2(7): 100288, 2021.

[25] G. Batt, B. Yordanov, R. Weiss, C. Belta, "Robustness analysis and tuning of synthetic gene networks," Bioinformatics, 23(18): 2415-2422, 2007.

[26] P. V. Chanekar, E. Nozari, J. Cortes, "Energy-Transfer edge centrality and its role in enhancing network controllability," IEEE Trans. Netw. Sci. Eng., 8(1): 331-346, 2021.

[27] L. Zhai, X. Yan, G. Zhang, "Bi-directional h-index: A new measure of node centrality in weighted and directed networks," J. Informetr., 12(1): 299-314, 2018.

[28] H. Ando, M. Bell, F. Kurauchi, K. I. Wong, K. F. Cheung, "Connectivity evaluation of large road network by capacity-weighted eigenvector centrality analysis," Transp. Transp. Sci., 17(4): 648-674, 2021.

[29] E. C. Baek, R. Hyon, K. López, E. S. Finn, M. A. Porter, C. Parkinson, "In-degree centrality in a social network is linked to coordinated neural activity," Nat. Commun., 13(1), 2022.

[30] X. Qi, E. Fuller, Q. Wu, Y. Wu, C. Q. Zhang, "Laplacian centrality: A new centrality measure for weighted networks," Inf. Sci., 194: 240-253, 2012.

[31] J. Laeuchli, Y. Ramírez-Cruz, R. Trujillo-Rasua, "Analysis of centrality measures under differential privacy models," Appl. Math. Comput., 412: 126546, 2022.

[32] Y. Li, Y. Sheng, X. Ye, "Group centrality algorithms based on the h-index for identifying influential nodes in large-scale networks," Int. J. Innovative Comput. Inf. Control, 16(4): 1183-1201, 2020.

[33] M. R. F. Mendonca, A. M. S. Barreto, A. Ziviani, "Approximating network centrality measures using node embedding and machine learning," IEEE Trans. Netw. Sci. Eng., 8,(1): 220-230, 2021.

[34] X. Li, Y. Liu, Y. Jiang, X. Liu, "Identifying social influence in complex networks: A novel conductance eigenvector centrality model," Neurocomput., 210: 141-154, 2016.

[35] Y. Yang, X. Wang, Y. Chen, M. Hu, C. Ruan, "A novel centrality of influential nodes identification in complex networks," IEEE Access, 8: 58742-58751, 2020.

[36] J. J. Molitierno, Applications of combinatorial matrix theory to Laplacian matrices of graphs. CRC Press, 2016.

[37] F. L. Lewis, D. L. Vrabie, V. L. Syrmos, Optimal control, 3rd ed. Hoboken: Wiley, 2012.

[38] V. V. Prasolov, S. Ivanov, Problems and theorems in linear algebra. in Translations of mathematical monographs, no. v. 134. Providence, R.I: American Mathematical Society, 1994.

[39] C. Tu, R. P. Rocha, M. Corbetta, S. Zampieri, M. Zorzi, S. Suweis, "Warnings and caveats in brain controllability," NeuroImage, 176: 83-91, 2018.

[40] M. Imani, U. Braga-Neto, "Optimal control of gene regulatory networks with unknown cost function," in Proc. 2018 Annual American Control Conference (ACC): 3939-3944, 2018.

[41] C. Li, L. Chen, K. Aihara, "Stability of genetic networks with SUM regulatory logic: Lur'e system and LMI approach," IEEE Trans. Circuits Syst. Regul. Pap., 53(11): 2451-2458, 2006.

[42] X. Zhang, Y. Wang, L. Wu, Analysis and Design of Delayed Genetic Regulatory Networks, vol. 207. in Studies in Systems, Decision and Control, vol. 207. Cham: Springer International Publishing, 2019.

[43] Y. Mori, Y. Kuroe, B. Ingalls, "A method for design of expression tracking controllers for gene regulatory networks," IFAC-PapersOnLine, 52(26): 135-142, 2019.

[44] Y. Y. Liu, A. L. Barabási, "Control principles of complex systems," Rev. Mod. Phys., 88, 2016.

[45] Y. J. Shin, L. Bleris, "Linear control theory for gene network modeling," PLOS ONE, 5(9), 2010.

Biographies

Zeinab Ghassemi Zahan received her B.Sc. degree in Computer Engineering from Ferdowsi University of Mashhad, Mashhad, Iran, and her M.Sc. degree in Electrical Engineering. She is currently a Ph.D. student in Electrical Engineering at Tarbiat Modares University, Tehran, Iran. Her experience includes being an international visiting graduate student in the Applied Mathematics department at the University of Waterloo, Canada. Her research interests include the area of control theory, complex networks, and systems biology.

- Email: zeinab.ghasemi@modares.ac.ir
- ORCID: 0000-0001-6995-8418
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Sadjaad Ozgoli received the B.S. and M.S. degrees in Electrical Engineering from Sharif University of Technology, Tehran, Iran Engineering Department. His current research interests include robust and nonlinear control, applied in systems biology and Mechatronic systems., in 1997 and 1999 respectively. His Ph.D. studies were in Electrical Engineering from K.N. Toosi University, Tehran, Iran, received in 2005. Since September 2006, he has been with the School of Electrical and Computer Engineering, Tarbiat Modares University in Tehran where he is currently an Associate Professor at the Control.

- Email: ozgoli@modares.ac.ir
- ORCID: 0000-0001-8567-1679
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

Sadegh Bolouki received the B.S. and Ph.D. degrees in Electrical Engineering from Sharif University of Technology, Tehran, Iran and Polytechnique Montréal, Montreal, QC, Canada in 2008 and 2014, respectively. Since September 2018, he has been an Assistant Professor of Electrical and Computer Engineering in Tarbiat Modares University, Tehran, Iran. He is also a Research Associate with the Department of Mechanical Engineering in Polytechnique Montréal, Montreal, Canada. His past experience includes postdoctoral tenures at the Department of Mechanical Engineering and Mechanics, Lehigh University, PA, USA and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, IL, USA. His research interests broadly include the areas of machine learning, network science, and game theory.

- Email: bolouki@modares.ac.ir
- ORCID: 0000-0002-4175-5894
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA

How to cite this paper:
Z. Ghassemi Zahan, S. Ozgoli, S. Bolouki, "A robust control centrality applicable to genetic regulatory networks with structured uncertainties," J. Electr. Comput. Eng. Innovations, 12(2): 511-524, 2024.

DOI: 10.22061/jecei.2024.10506.714

URL: https://jecei.sru.ac.ir/article_2150.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Automatic Cadastral Boundary Detection of Very High Resolution Images Using Mask R-CNN

N. Rahimpour Anaraki, A. Azadbakht, M. Tahmasbi *, H. Farahani, S. R. Kheradpishe, A. Javaheri

Department of computer and data sciences, Faculty of mathematical sciences, Shahid Beheshti University, Tehran, Iran.

Article Info

Article History:
Received 05 March 2024
Reviewed 15 April 2024
Revised 03 July 2024
Accepted 15 July 2024

Keywords:
Remote sensing
Mask R-CNN
Cadastral mapping
Instance segmentation

*Corresponding Author's Email Address:
m_tahmasbi@sbu.ac.ir

Abstract

Background and Objectives: Cadastral boundary detection deals with locating the boundary of the ownership and use of land. Recently, there has been high demand for accelerating and improving the automatic detection of cadastral mapping. As this problem is in its starting point, there are few researches using deep learning algorithms.

Methods: In this paper, we develop an algorithm with a Mask R-CNN core followed with geometric post-processing methods that improve the quality of the output. Many researches use classification or semantic segmentation but our algorithm employs instance segmentation. Our algorithm includes two parts, each of which consists of a few phases. In the first part, we use Mask R-CNN with the backbone of a pre-trained ResNet-50 on the ImageNet dataset. In the second part, we apply three geometric post-processing methods to the output of the first part to get better overall output. Here, we also use computational geometry to introduce a new method for simplifying lines which we call pocket-based simplification algorithm.

Results: We used 3 google map images with sizes 4963 × 2819, 3999 × 3999, and 5520 × 3776 pixels. And divide them to overlapping and non-overlapping 400×400 patches used for training the algorithm. Then we tested it on a google map image from Famenin region in Iran. To evaluate the performance of our algorithm, we use popular metrics Recall, Precision, and F-score. The highest Recall is 95%, which also maintains a high precision of 72%. This results in an F-score of 82%.

Conclusion: The idea of semantic segmentation to derive boundary of regions, is new. We used Mask R-CNN as the core of our algorithm, that is known as a very suitable tools for semantic segmentation. Our algorithm performs geometric post-process improves the f-score by almost 10 percent. The scores for a region in Iran containing many small farms is very good.

Introduction

One of the bases of land administration systems is recording the ownership and physical location of real properties, which are called cadastres [1]. Recently, cadastral mapping has received considerable attention. An effective cadastral system formalizes private property rights, which is very important to promote agricultural productivity, support national development, and secure an effective land market [2].

However, estimates suggest that about 75% of the world's population doesn't have access to a formal system to register and safeguard their land rights. Establishing a complete land cadaster and keeping it up-to-date is a contemporary challenge for many developing and developed countries [3], [4]. This lack of recorded

land rights increases insecure land tenure and fosters existence-threatening conflicts, particularly in developing countries. Recording land rights spatially, i.e., cadastral mapping by traditional field surveying approaches, is considered the most expensive, time-consuming, and labor-intensive part of a land administration system. Therefore, in order to speed up the process, we need innovative tools [5], [6].

Earth observation satellites provide very high-resolution (VHR) images. Unmanned aerial vehicle (UAV) images are also available in different areas. According to the Union of Concerned Scientists (UCS), there were 971 EO satellites in orbit on the 30th of April 2021. For context, when they did a similar report at the end of April 2018, there were only 684 satellites, so there has been a 41.95% increase over the three years [7].

Since the availability of VHR images, remote sensing has been used for mapping cadastral boundaries instead of field surveying and is advocated by fit-for-purpose (FFP) land administration [3]. In these images, visible cadastral boundaries are often marked by physical objects such as rivers, roads, water drainages, building walls, clusters of stones, fences, strips of uncultivated land, ditches, etc. [1]. These boundaries are apparent in remotely sensed images and can be automatically identified using image processing algorithms [6].

Recent studies have highlighted the effectiveness of deep learning techniques, such as Convolutional Neural Networks (CNNs), in extracting high-level representations crucial for detection and classification tasks, opening up new opportunities in cadastral boundary detection [8,9].

In deep learning, two primary approaches are commonly employed to train CNNs: starting from scratch or utilizing transfer learning [10]. In this study, we opt for the Mask R-CNN model with a pre-trained ResNet-50 backbone from the ImageNet dataset for instance segmentation to identify cadastral boundaries. This approach, initially used by Mayer et al. [11], focuses on delineating the boundaries of individual fields. Here we focus on finding the boundary of each individual farm and leveraging transfer learning for pre-trained ResNet-50 was used. Also, we introduce a new geometric method for simplifying lines (extracted boundary) which we call the Pocket-based simplification algorithm that achieves better performance than the commonly used Douglas-Peucker algorithm [12].

The Core of our algorithm is a Mask R-CNN with the backbone of a pre-trained ResNet-50 on the ImageNet dataset to produce the initial output, which is a probability map of fields or non-field pixels. Then, Otsu's binary method with a certain threshold is applied to the output to identify which pixels represent fields. Finally, Canny edge detection algorithm delineates the boundaries of these fields. In the second part, to enhance the results, we eliminate fields that have very small area. Then, we remove polygons inside other polygons, and finally, the extracted lines (boundaries) are simplified using a new geometric simplification algorithm. The third part involves accuracy assessment where we add a buffer to the ground-truth boundaries and calculate Recall, Precision, and F-score in various scenarios.

This paper is organized as follows: in section Literature Review we review recent results in this field. In the section Method, we first describe our algorithm in detail and evaluation methods. Details on the training data are provided in the section Experimental Study. Section Evaluation following by section results and discussion, focuses on assessing the model and comparing our proposed simplification algorithm with the state-of-the-art. Finally, in the last section, we present our final conclusion and open problems.

Literature Review

In this section, we review some studies in this field and discuss various image processing methods to address problems.

Some methods for cadastral mapping are based on image segmentation and edge detection. In Drăguţ et al. [13], they introduce a new automated approach called Multi-Resolution Segmentation (MRS) for parameterizing multi-scale image segmentation of multiple layers. This approach relies on the potential of local variance to detect scale transitions in geospatial data. Classical edge detection aims to identify sharp changes in image brightness through local measurements, including first-order (e.g., Prewitt or Sobel) and second-order (e.g., Gaussian or Laplacian) derivative-based detection [14].

In a study by Crommelinck et al. [15] they aimed to apply computer vision techniques to analyze remotely sensed UAV images for UAV-based cadastral mapping. Their approach involved a three-step process for identifying cadastral boundaries. The first step is image preprocessing, where the UAV orthoimage underwent resampling to lower resolutions and was divided into patches. Next, in boundary delineation, they utilized the Globalized Probability of Boundary (gPb) contour detection method, the state-of-the-art computer vision method, on each patch. This process generated contour maps with probabilities assigned to contours per pixel. Lastly, in image post-processing, all patches from the same image were combined to create a unified contour map and a binary boundary map, which was then converted into vector format. This systematic procedure allowed for the representation of cadastral boundaries for mapping applications.

Numerous studies have employed Convolutional Neural Network (CNN) tools for cadastral mapping, with Crommelinck et al. [16] presenting a three-step workflow. The first step involves image segmentation to extract

visible object outlines, followed by boundary classification to predict the likelihood of boundaries for the extracted segment lines. The final step is interactive delineation, connecting these lines based on the predicted boundary likelihood.

For image segmentation, Multiresolution Combinatorial Grouping (MCG) generates closed contours that show the outlines of objects. Then, a boundary classification is applied to the resulting post-processed MCG lines, using two machine learning approaches: Random Forest (RF) and pre-trained VGG19 through transfer learning, to achieve boundary likelihood per line. Interactive delineation facilitates the creation of final cadastral boundaries through various functions as a plugin in QGIS. All training images used in their study are obtained from UAV images.

In Fetai et al. [17], UAV images are utilized in their workflow, which comprises three main steps: Image pre-processing, boundary detection and extraction, and data post-processing. The initial step involves resampling the UAV orthoimage, followed by applying the ENVI feature extraction module [18], [19] to each down-sampled UAV orthoimage. In the final step, extracted objects are filtered and simplified.

Xia et al. [6] utilize deep Fully Convolutional Networks (FCNs) to detect cadastral boundaries using UAV images captured over urban and semi-urban areas. They approach boundary detection as a supervised pixel-wise image classification task to differentiate between boundary and non-boundary pixels. The network employed in their research is a modified version of the FCN with dilated kernel (FCN-DK) detailed in [20]. Other studies in this field can be found in [21]-[23].

Our solution is based on instance segmentation, which has become one of the relatively important, complex, and challenging areas in computer vision research. Hafiz et al. [24] review some advances in instance segmentation:

Girshick et al. [25] were among the first to explore CNNs for instance segmentation [26]. They developed the R-CNN technique, which integrated AlexNet [27] along with a region proposal using the selective search technique [28]. However, it has some drawbacks, such as having difficult and slow training. Therefore, Girshick [29] introduced Fast R-CNN, which addressed some of the issues of R-CNN and improved its object detection ability.

In Zagoruyko et al. [30], MultiPath Network was introduced by applying three modifications to the standard Fast R-CNN model. Initially, skip connections have been integrated to grant the object detector access to features from various network layers. Subsequently, a foveal component has been used to leverage the context of objects across varying resolutions. Lastly, an integral nature loss function has been included.

Although Fast R-CNN improved detection speed significantly, it still depended on external region proposals, which posed a computational bottleneck. To address this issue, Ren et al. [31] introduced the Faster R-CNN model with a Region Proposal Network (RPN) for efficient and accurate region proposal generation. This model utilized the same backbone network and extracted features from the last shared convolutional layer for both RPN-based region proposal and Fast R-CNN region classification.

Finally, He et al. [32] introduced Mask R-CNN as a straightforward and adaptable model for instance segmentation. This model efficiently achieves instance segmentation by combining object detection with the concurrent creation of precise masks. Mask R-CNN builds upon the foundation laid by Faster R-CNN. Typically, Faster R-CNN includes a branch dedicated to recognizing object bounding boxes. Mask R-CNN enhances this framework by introducing a parallel object mask prediction branch, thereby improving the overall performance of the model.

Method

In this section, we present our algorithm that consists of two main parts: A deep convolutional network that detects boundaries, and a geometric post-process that simplifies the boundary and cleans up the map.

A. R-CNN

Here, we propose an effective pipeline for detecting the boundaries of two fields in satellite images. The core module in this pipeline is a Mask R-CNN model. This model is responsible for detecting each field in small input images of size 400*400. The R-CNN is a two-stage detection algorithm. The first stage identifies a subset of regions in an image that might contain an object, i.e., a field. The second stage classifies the object in each region and returns a probability map of the presence of an object, i.e., a field over pixels.

The input images were raw images without additional labels or external data. In order to find the borders of fields in a supervised manner, we manually generated labels for some of the input images. Using the LabelMe [33] library, we carefully generated a mask for each input image, coloring each field area from 1 to 256 such that neighboring fields have different colors.

To train an instance segmentation model, we set the Mask R-CNN model with the backbone of a pre-trained ResNet-50 [34] on the Imagenet dataset [35]. ResNet-50 is selected as the backbone for its ability to convert input images into feature vectors of size 4096. These feature vectors play a crucial role in predicting the bounding box, class, and mask of each identified object. Without leveraging ResNet-50 as the backbone, the model would require a much larger amount of data for training purposes. Deeper ResNet networks require more computational power, and since we are working with super-large images, it is suitable choice. Finetuning a pre-

trained model would give better results in our experiments. Choosing the input size of this model is a tricky challenge. Small input images would cause the detected object to be more fragmented, while large input images would cause the model to ignore small features and consider small fields as noise. We chose an input size of 400*400 so that the models would detect enough small features and fields, and the output fields would not be too fragmented, while keeping the computational cost of training the model reasonable.

A dataset of 400*400 patches was extracted from the input images, and R-CNN models were trained on them. The model tries to detect each field as an individual object, and for each object, it returns a probability mask stating whether a pixel is in a field or not. In the challenge, we do not need each field alone, but we want some sense of the borders. So, if we could only identify the interior area of each field, we can extract the borders. To simplify the overall solution, we apply max pooling to all the detected object masks.

Our solution for this challenge consists of five steps: patchify, R-CNN model, unpatchify, edge and border detector, and border refiner. During inference time, we slide a window of size 400*400 with steps of 200 pixels on the input image. This way, the patches overlap. This step is called patchify. Each patch goes into the RCNN model, and the resulting masks are aggregated together to create the final large output. This step is called unpatchify. On the overlapping parts of patches, we apply different methods to aggregate the probabilities: max pooling, averaging, summation, and harmonic mean.

At the end of the unpatchify step, we obtain a large mask of the input model where values close to 1 are more likely to be fields. In order to extract borders from masks, we apply Otsu's binarizing method [36] to binarize probability masks adaptively; then the canny edge detection algorithm [37] applies to the output, and the resulting image contains only edges and borders of the fields.

Mask R-CNN is an advanced model for object detection and instance segmentation, building upon the foundations of Faster R-CNN. Faster R-CNN itself is designed to accurately identify and localize objects within an image by predicting both their bounding boxes and class scores. Mask R-CNN enhances this by introducing an additional branch that predicts segmentation masks for each identified object instance, providing a more detailed analysis of the image content. This capability is particularly useful in scenarios where precise object outlines are required, beyond mere detection.

In this study, we introduce a refined instance segmentation framework based on the integration of the Mask Region-based Convolutional Neural Network (Mask R-CNN) with a ResNet-50 architecture, pre-trained on the Comprehensive Object Detection, Localization, and Segmentation (COCO) dataset. This framework is engineered to perform dual functions: object detection and instance segmentation, enabling precise pixel-wise delineation of objects within an image. Utilizing the concept of transfer learning, the pre-trained model serves as a foundational backbone, facilitating accelerated training convergence and reducing the necessity for extensive domain-specific data.

The core adaptation of the model revolves around the customization of the Region of Interest heads, specifically the box predictor and the mask predictor, to cater to a user-defined number of object classes. The adjustment process begins with the box predictor, where the input features for the classifier are derived from the pre-existing model structure.

Subsequently, the original Fast R-CNN predictor is substituted with a novel Fast R-CNN Predictor, recalibrated to project the model's output to the specified number of classes. This alteration enables the adapted model to classify objects into a bespoke set of categories, diverging from the standard categorization learned during its initial training on the COCO dataset.

Parallel to the modifications in the box predictor, the mask predictor undergoes a similar transformation. The original mask predictor is replaced with a newly instantiated Mask R-CNN Predictor. This adjustment entails the configuration of input features alongside the introduction of an intermediate hidden layer with 256 nodes, culminating in the ability to generate class-specific segmentation masks.

Data augmentation is a critical technique in the field of machine learning and computer vision, particularly beneficial for enhancing the robustness and generalization capabilities of deep learning models. By artificially expanding the training dataset through various transformations, data augmentation introduces a diversity of perspectives, angles, and environmental conditions, simulating a more comprehensive range of real-world scenarios.

This process significantly mitigates the risk of overfitting, as the model is trained on a broader spectrum of data instances, improving its ability to generalize to unseen data. Moreover, data augmentation is applied. For this purpose, mirrored versions of the training images are added. Doing so ensures that the model is not biased towards the original orientation of objects within the images, fostering an ability to accurately detect and segment objects regardless of their horizontal alignment.

The model, leveraging a custom adaptation of the Mask R-CNN architecture with a ResNet-50 backbone, is trained using a Stochastic Gradient Descent (SGD) optimizer. This choice is motivated by SGD's proven efficacy in handling noise and its capacity for

generalization in large-scale data scenarios. The learning rate is set at 0.005, a value that balances the trade-off between training speed and the risk of overshooting minima in the loss landscape. To enhance the optimization process, a momentum of 0.9 is set, facilitating the acceleration of the optimizer in relevant directions and improving convergence speed.

Additionally, a weight decay of 0.0005 is applied as a regularization measure to prevent overfitting by penalizing large weights, ensuring the model's generalizability to unseen data.

The learning rate is reduced by a factor of 10 every 3 epochs. This approach, known as step decay, is instrumental in fine-tuning the model's performance by gradually decreasing the learning rate, allowing for more refined adjustments to the model weights as training progresses.

The training process spans 50 epochs, a duration determined to strike an optimal balance between achieving sufficient model convergence and avoiding excessive computational expenditure. The choice of 50 epochs is also influenced by empirical observations of model performance over time, ensuring that the model benefits from prolonged exposure to the training data without succumbing to overfitting.

B. Geometric Post-Process

Having a clean output as a binary PNG image, we need to transform it into a shapefile that can be used in Geographic Information Systems (GIS) like QGIS, ArcGIS, etc. Therefore, we extract polygons from the output image and after vectorizing it, a shapefile is created. The resulting polygons need further post-processing. First, we need to define some concepts, then we investigate post-process steps.

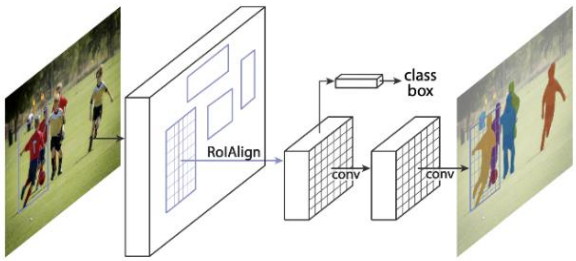


Fig. 1: Mask RCNN model.

A subset S of the plane is called convex if and only if for any pair of points $p, q \in S$ the line segment $\overline{pq}$ is completely contained in S. The *convex hull* of a set S (denoted by $CH(P)$) is the smallest convex set that contains S. To be more precise, it is the intersection of all convex sets that contain S [38].

The *pockets* of a simple polygon are the polygonal areas outside the polygon, but inside its convex hull [38]. In Fig. 2, three white polygons are pockets.

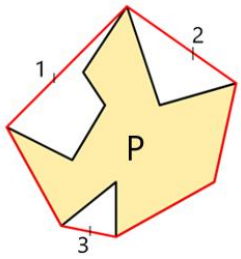


Fig. 2: Polygon P and its three pockets in white.

Our post-process performs three steps:

1. Deleting small polygons: Polygons with an area less than a fixed number with respect to Ground Sample Distance (GSD) of an image are removed because they have an area less than a real farm in that region.

2. Deleting polygons inside other polygons: We want to make sure no polygon contains any smaller polygon inside.

3. Simplifying over-segmentation: Over-segmentation can cause serrated edges in polygons. There are different simplification methods such as the Douglas-Peucker algorithm. We develop a new algorithm called the pocket-based simplification algorithm. Here is how this algorithm works:

Given a polygon P, we compute all of its pockets, $P_1 \dots P_k$. Now for each pocket, we calculate two distances:

1. $dist$: The length of the edge that belongs to the pocket but not the polygon P. We call this edge *probable edge* (numerized marked edges in Fig. 1).
2. d: The summation on the length of all edges belongs to pocket except the probable edge.

Then we choose certain threshold (denoted by t) and if $d < t \times dist$, all edges used for calculating d are added as edges of polygon P'. Else, only the probable edge is added as an edge of polygon P'. At the end of the process, polygon P' which is the simplified version of polygon P is added to the shapefile instead of the original polygon P itself. This process is repeated for all extracted polygons, resulting in a final shapefile with the simplified output. This approach can be beneficial for reducing the complexity of the shapefile while preserving important geometric characteristics of the original polygons.

Fig. 3 illustrates the impact of the three geometric post-processing steps on the raw shapefile. These steps include deleting small polygons, deleting polygons inside another one, and simplifying over-segmentation.

In Fig. 4, a comparison is made among the raw output, the output after applying the Douglas-Peucker algorithm, and the output after applying the Pocket-based simplification algorithm. The differences among these outputs showcase how each algorithm affects the shapefile in terms of simplification and preservation of geometric details. This comparison can provide insights

into the effectiveness of each algorithm in achieving the desired level of simplification while maintaining the essential characteristics of the original data.

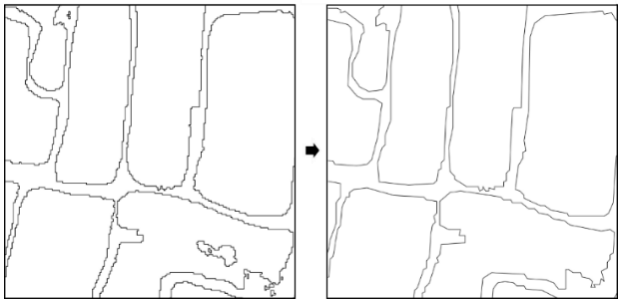


Fig. 3: Raw output (right) and the output after applying geometric post process (left), Simplified by Douglas-Peucker algorithm.

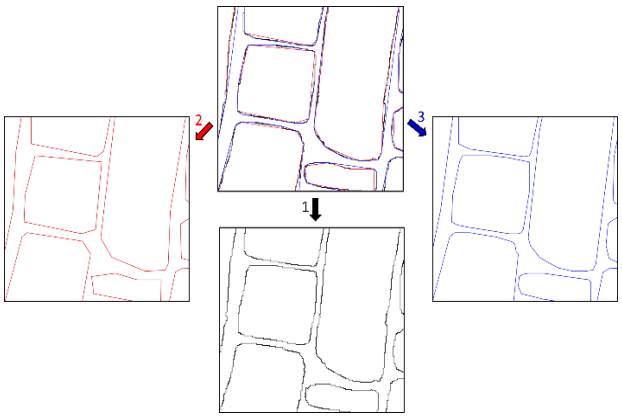


Fig. 4: Picture of boundary without simplification (1-black), simplified by Douglas-Peucker algorithm (2-red) and simplified by Pocket-based algorithm (3-blue).

Experimental Study

In this section, we present the data for training and how we prepare them to be ready for input to Mask R-CNN and the system process executed on.

A. Training Data

We use three images for training the network: an aerial image, an ortho image, and a Famenin Irrigated image. These images have dimensions of 4963 × 2819, 3999 × 3999, and 5520 × 3776 pixels respectively. The Famenin Irrigated image is a satellite image obtained from Google Earth, while the other two are UAV images taken in farmlands in Iran and Ethiopia. Since we are solving an instance segmentation problem, we need to create masks for each image to achieve our goal. We created the masks ourselves, as no online dataset satisfied our needs. To create the masks for the training data, we used LabelMe, a free graphical image annotation tool written in Python with a graphical interface built using Qt. In the final steps, each image was divided into tiles of 400 × 400 pixels, resulting in almost 300 tiles for the training data. Fig. 5 shows the result of a field together with its labeled mask. The only important requirement for the masks is that fields sharing the same boundary should not have the same color. Therefore, non-neighbor fields are free to have the same colors.

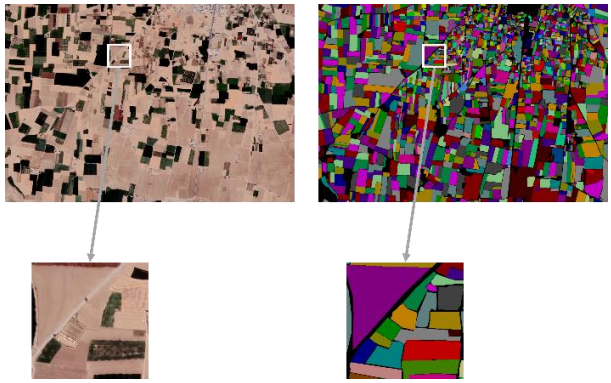


Fig. 5: Result of a training image with it's labeled mask.

B. System Config

The entire process of this study was implemented in Python. The training of the instance segmentation model was conducted using four K80 GPUs to optimize the balance between computational power and resource accessibility. This configuration facilitated a distributed training approach, enhancing efficiency and reducing training duration without compromising accuracy. The training time in this system takes approximately 2 hours.

Evaluation

In this section, we present the performance of our detected boundary.

A comparison of the accuracy assessment results obtained in this study with those from other studies cannot be done due to several reasons. Various UAVs and satellites may yield different quality images, and the study objects exhibit significant variations in terms of nature, size, location, and characteristics. To enable a reliable comparison of accuracy assessments across different feature extraction methods, each method needs to be individually studied and subsequently tested within the same study area(s) [17].

We calculate all accuracy assessments on one of the pictures from Famenin, which is 3432 × 3621 pixels. This picture isn't used for training and is only used for testing. For validating the predicted results with ground truth, a certain buffer is often considered for the reference boundary in cadastral mapping [19]. We consider both 5- and 6-pixel buffers for the reference boundary, which means the original reference boundary has a width of 1 pixel while in evaluation it is 5 or 6 pixels.

The measures we use to evaluate our work are common ones in this field: precision, recall, and F-score.

Precision represents whether the assigned boundary is valid. Recall shows the ability of the network to find all the boundaries, and F-score is the harmonic mean between precision and recall, which is a good overall measurement for final evaluation.

After overlapping our output boundary with the buffered reference, we calculate the mentioned measures using these formulas:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \times BF \quad (2)$$

$$F-score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (3)$$

If we consider binary classification for the confusion matrix, class positive stands for pixels labeled as "boundary," and class negative stands for pixels labeled as "non-boundary." Thus, we can create a confusion matrix and calculate the measures. BF in the recall formula stands for the buffer that we considered for the boundary reference in the Famenin test picture, and its quantification is different in each case. The rationale for the recall formula is that the sum of True Positives (TP) and False Positives (FP) represents the total number of detected boundary pixels, while TP + False Negatives (FN) indicates the total boundaries in the buffered reference, not the original reference. Therefore, to obtain the total boundaries in the original reference, which has a width of one pixel, we divide the sum of TP and FN by BF [6].

Results and Discussion

In this section, we present the performance of our detected boundary.

As we can see in Table 1, it is obvious that when we make the buffer thicker by 1 pixel, both precision and recall rise as more detected boundaries fall into the reference boundary buffer. A recall of 95% in a 6-pixel buffer shows that R-CNN finds almost all boundaries of the image. However, because the precision is less than 80%, we should note that some boundaries are detected that are not in the reference, making them extra.

If we fix the buffer, then the recall in both methods of simplifying polygons is almost equal, but precision is always better in the Pocket-based algorithm, indicating that more valid boundaries fall into the reference buffer in our method than in the Douglas-Peucker algorithm.

The important point is that precision achieved by a 5-pixel buffer and the paper's method is 67%, while this precision is achieved by the Douglas-Peucker algorithm only when the buffer is 6 pixels. This demonstrates how much stronger the Pocket-based algorithm works in finding valid boundaries compared to the Douglas-Peucker algorithm.

Based on better precision and almost equal recall in the Pocket-based simplification algorithm, our method consistently achieves a higher F-score. Therefore, based on these results, we can conclude that our simplification method, also known as the Pocket-based simplification algorithm, performs better in all tested situations.

Table 1: Accuracy comparison with respect to buffer and simplification method

Buffer	Method	Precision	Recall	F-Score
5 pixels	Douglas-Peucker	60	85	70
5 pixels	Pocket-based	67	87	76
6 pixels	Douglas-Peucker	66	95	78
6 pixels	Pocket-based	72	95	82

It is worth mentioning that our primary objective is to detect cadastral boundaries of farmlands. However, our presented method can also identify boundaries of buildings and other types of land in rural or urban areas. Additionally, there are many smallholders in our images, and detecting them is more challenging than identifying larger objects. One of the main issues with our detected boundaries is that the extracted polygons tend to be more rounded at the edges rather than straight. Therefore, geometric post-processing significantly impacts our network output and enhances its quality.

Conclusion

In this study, the Mask R-CNN model was used to solve instance segmentation for the automatic detection of cadastral boundaries in VHR images. The Mask R-CNN model is based on transfer learning, utilizing a pre-trained ResNet-50 backbone from the ImageNet dataset. The network can accept input at any desired resolution, and the output is a large mask of the input image, where values close to one represent the field. To extract the boundaries of each field from the masks, we apply Otsu's binarizing method to adaptively binarize the probability masks. Subsequently, the Canny edge detection algorithm is applied to the output, resulting in an image containing only cadastral boundaries. Since the instance segmentation method is used, the network's output contains individual boundaries for each field. Following the creation of a shapefile from the binary PNG output of the network, three geometric post-processing procedures are applied to enhance the raw output of the Mask R-CNN.

In the first step, polygons with an area smaller than a specific threshold based on the minimum area of farms in that region and the image's GSD are removed. In the second step, all polygons that are contained within another polygon are eliminated, as farms are non-overlapping. The final step involves using two methods to simplify the detected boundaries. While the Douglas-

Peucker algorithm, a well-known approach for line simplification, is utilized, a new method called the Pocket-based simplification algorithm is introduced and shown to outperform the Douglas-Peucker algorithm. This method is named after the pockets produced when creating the convex hull of each polygon. Based on a specific threshold, decisions are made regarding whether to add a pocket or connect edges between two endpoints of a polygon, resulting in a new simplified polygon. The Pocket-based simplification algorithm demonstrates higher accuracy in precision compared to the Douglas-Peucker algorithm, with nearly identical recall rates. This indicates that both methods can effectively identify cadastral boundaries, but the Pocket-based simplification algorithm excels at detecting valid boundaries. The final evaluation is based on the F-score, which is the harmonic mean between precision and recall and consistently favors the Pocket-based simplification algorithm over the Douglas-Peucker algorithm.

Instance segmentation for detecting automatic cadastral boundaries has shown promising results, suggesting that employing other networks optimized for instance segmentation may further improve outcomes.

There are some open problems:

Our algorithm does not detect and delete urban/rural areas, resulting in segmentation of those areas, which reduce the precision and recall of our algorithm. One solution is to use semantic segmentation to detect and delete urban/rural areas. Then, we use our algorithm to detect boundary of farmlands.

The result heavily depends on the texture and shape of farmland that differs in different geographical regions and type of farmlands.

In this research, we utilize only the three visible channels of images (RGB). However, incorporating invisible bands alongside RGB bands in satellite images can enhance the performance of this algorithm.

Abbreviations

VHR	Very Hight Resolution
UAV	Unmanned Aerial Vehicle
EO	Earth Observation
MRS	Multi Resolution Segmentation
CNN	Convolutional Neural Network
MCG	Multiresolution Combinatorial Grouping
FCN	Fully Convolutional Network

References

[1] X. Luo, R. M. Bennett, M. Koeva, C. Lemmen, "Investigating semi-automated cadastral boundaries extraction from airborne laser scanned data," Land, 6(3): 60, 2017.

[2] I. Williamson, "The justification of cadastral systems in developing countries," Geomatica, 51: 21-36, 1997.

[3] S. Enemark, K. C. Bell, C. Lemmen, R. McLaren, Fit-for-purpose land administration. International Federation of Surveyors (FIG), Copenhagen: International Federation of Surveyors (FIG), 2014.

[4] X. Luo, R. Bennett, M. Koeva, C. Lemmen, N. Quadros, "Quantifying the overlap between cadastral and visual boundaries: A case study from Vanuatu," Urban Sci., 1: 32, 2017.

[5] I. Williamson, S. Enemark, J. Wallace, A. Rajabifard, Land administration for sustainable development, Citeseer, 2010.

[6] X. Xia, C. Persello, M. Koeva, "Deep fully convolutional networks for cadastral boundary detection from UAV images," Remote Sens., 11(14): 1725, 2019.

[7] A. L. Dr Samantha Lavender, Trusted Earth Observation Experts, 2012.

[8] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," IEEE Geosci. Remote Sens. Mag., 5: 8-36, 2017.

[9] J. R. Bergado, C. Persello, C. Gevaert, "A deep learning approach to the classification of sub-decimetre resolution aerial images," in Proc. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016.

[10] D. Garcia-Gasulla, F. Parés, A. Vilalta, J. Moreno, E. Ayguadé, J. Labarta, U. Cortés, T. Suzumura, "On the behavior of convolutional nets for feature extraction," J. Artif. Intell. Res., 61: 563-592, 2018.

[11] L. Meyer, F. Lemarchand, P. Sidiropoulos, "A deep learning architecture for batch-mode fully automated field boundary detection," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 43: 1009-1016, 2020.

[12] S. Crommelinck, R. Bennett, M. Gerke, F. Nex, M. Y. Yang, G. Vosselman, "Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping," Remote Sens., 8(8): 689, 2016.

[13] L. Drăguţ, O. Csillik, C. Eisank, D. Tiede, "Automated parameterisation for multi-scale image segmentation on multiple layers," ISPRS J. Photogramm. Remote Sens., 88: 119-127, 2014.

[14] Y. Li, S. Wang, Q. Tian, X. Ding, "A survey of recent advances in visual feature detection," Neurocomputing, 149: 736-751, 2015.

[15] S. Crommelinck, R. Bennett, M. Gerke, M. Y. Yang, G. Vosselman, "Contour detection for UAV-based cadastral mapping," Remote Sens., 9: 171, 2017.

[16] S. Crommelinck, M. Koeva, M. Y. Yang, G. Vosselman, "Application of deep learning for delineation of visible cadastral boundaries from remote sensing imagery," Remote Sens., 11: 2505, 2019.

[17] B. Fetai, K. Oštir, M. Kosmatin Fras, A. Lisec, "Extraction of visible boundaries for cadastral mapping based on UAV imagery," Remote Sensing, 11: 1510, 2019.

[18] J. Wang, J. Song, M. Chen, Z. Yang, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," Int. Remote Sens., 36: 3144-3169, 2015.

[19] D. Poursanidis, N. Chrysoulakis. Z. Mitraka, "Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping," Int. J. Appl. Earth Obs. Geoinf., 35: 259-269, 2015.

[20] C. Persello, A. Stein, "Deep fully convolutional networks for the detection of informal settlements in VHR images," IEEE geoscience and remote sensing letters, 14: 2325-2329, 2017.

[21] Y. Xu, Z. Zhu, M. Guo, Y. Huang, "Multiscale edge-guided network for accurate cultivated land parcel boundary extraction from remote sensing images," IEEE Trans. Geosci. Remote Sens., 62, 2023.

[22] Z. Cai, Q. Hu, X. Zhang, J. Yang, H. Wei, J. Wang, Y. Zeng, G. Yin, W. Li, L. You, B. Xu, Z. Shi, "Improving agricultural field parcel delineation with a dual branch spatiotemporal fusion network by

integrating multimodal satellite data," ISPRS J. Photogramm. Remote Sens., 205: 34-49, 2023.

[23] M. T. Metaferia, R. M. Bennett, B. K. Alemie, M. Koeva, "Furthering automatic feature extraction for fit-for-purpose cadastral updating: Cases from Peri-Urban Addis Ababa, Ethiopia," Remote Sensing, 15: 4155, 2023.

[24] A. M. Hafiz, G. M. Bhat, "A survey on instance segmentation: state of the art," Int. J. Multimedia Inf. Retr., 9: 171-189, 2020.

[25] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[26] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, "Deep learning for generic object detection: A survey," Int. J. Comput. Vision, 128: 261-318, 2020.

[27] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012.

[28] K. E. A. Van de Sande, J. R. R. Uijlings, T. Gevers, A. W. M. Smeulders, "Segmentation as selective search for object recognition," in Proc. 2011 International Conference on Computer Vision, 2011.

[29] R. Girshick, "Fast R-CNN," in Proc. the IEEE International Conference On Computer Vision: 1440-1448, 2015.

[30] S. Zagoruyko, A. Lerer, T. Y. Lin, P. O. Pinheiro, S. Gross, S. Chintala, P. Dollár, "A multipath network for object detection," arXiv preprint arXiv: 1604.02135, 2016.

[31] S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, 28, 2015.

[32] K. He, G. Gkioxari, P. Doll'ar, R. Girshick, "Mask r-cnn," in Proc. the IEEE International Conference on Computer Vision, 2017.

[33] K. Wada, labelme: Image Polygonal Annotation with Python, GitHub, 2018.

[34] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[36] J. D. Yang, Y. S. Chen, W. H. Hsu, "Adaptive thresholding algorithm and its hardware implementation," Pattern Recognit. Lett., 15: 141-150, 1994.

[37] J. Canny, "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence, PAMI-8(6): 679-698, 1986.

[38] M. De Berg, Computational geometry: algorithms and applications, Springer Science & Business Media, 2000.

Biographies

Neda Rahimpour Anaraki received her B.Sc. degree from Alzahra University, and her M.S. degree from Amirkabir University of Technology in 2017 and 2019, respectively, both in Computer Science. She is currently Ph.D. candidate at Shahid Beheshti University. Her research interests are computational geometry, remote sensing, computer vision, machine learning and deep learning.

- Email: neda.rpa@gmail.com
- ORCID: 0009-0000-4382-7145
- Web of Science Researcher ID: NA
- Scopus Author ID : NA
- Hompage: NA

Alireza Azadbakht earned his B.Sc. in Computer Science from Shahid Beheshti University between 2016 and 2020, followed by an MSc, specializing in data mining, from the same institution between 2020 and 2022. His core research interests lie in computer vision, deep learning, remote sensing, data science, and machine learning.

- Email: ali.r.azadbakht@gmail.com
- ORCID: 0000-0002-9663-8388
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Hompage: NA

Maryam Tahmasbi received her Ph.D. in computer science from Amirkabir university of technology 2009. She is currently assistant professor in department of computer and data sciences in Shahid Beheshti university. Her research interest is geometric data analysis concentrating on image processing.

- Email: m_tahmasebi@sbu.ac.ir
- ORCID: 0000-0001-9513-0251
- Web of Science Researcher ID: NA
- Scopus Author ID: Maryam Tahmasbi - 56120848900
- Hompage: https://mathsci.sbu.ac.ir/~m_tahmasbi

Hadi Farahani is an assistant professor in computer science at Shahid Beheshti University. He received his Ph.D. in mathematics from Shahid Beheshti University in 2011. His research interests includes machine learning, deep learning, blockchain technology and applications of logic in computer science.

- Email: h_farahani@sbu.ac.ir
- ORCID: 0000-0002-8067-0383
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Hompage: https://mathsci.sbu.ac.ir/~h_farahani

Saeed Reza Kheradpishe is an assistant professor in computer science at Shahid Beheshti University. He received his Ph.D. in computer science from the University of Tehran in 2017. His research interests include machine learning, deep learning, spiking neural networks and computational neuroscience.

- Email: s_kheradpishe@sbu.ac.ir
- ORCID: 0000-0001-6168-4379
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Hompage: https://mathsci.sbu.ac.ir/~s_kheradpisheh

Alireza Javaheri received his B.Sc. degree and M.S. degree from Shahid Beheshti University, Tehran, Iran, respectively, both in Computer Science. His research interests are in the areas of Machine Learning, Deep Learning and Meta-Learning.

- Email: alireza.jvh98@gmail.com
- ORCID: 0000-0002-9663-8388
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Hompage: NA

How to cite this paper:
N. Rahimpour Anaraki, A. Azadbakht, M. Tahmasbi, H. Farahani, S. R. Kheradpishe, A. Javaheri, "Automatic cadastral boundary detection of very high resolution images using mask R-CNN," J. Electr. Comput. Eng. Innovations, 12(2): 525-534, 2024.

DOI: 10.22061/jecei.2024.10650.727

URL: https://jecei.sru.ac.ir/article_2151.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Society Deciling Process: A Socio-Inspired Meta-Heuristic Algorithm

E. Pira [*]***, A. Rouhi***

Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran.

Article Info

Article History:
Received 14 April 2024
Reviewed 28 May 2024
Revised 18 July 2024
Accepted 28 July 2024

Keywords:
Society deciles
Meta-heuristic
Test function
Cec 2019
Cec 2022
Convergence speed

[*]Corresponding Author's Email Address: *pira@azaruniv.ac.ir*

Abstract

Background and Objectives: The development of effective meta-heuristic algorithms is crucial to solve complex optimization problems. This paper introduces the Society Deciling Process (SDP), a novel socio-inspired meta-heuristic algorithm that simulates the social categorization into deciles based on metrics such as income, occupation, and education. The objective of this research is to introduce the SDP algorithm and evaluate its performance in terms of convergence speed and hit rate, comparing it with seven well-established meta-heuristic algorithms to highlight its potential in optimization tasks.

Methods: The SDP algorithm's efficacy was evaluated using a comprehensive set of 14 general test functions, including benchmarks from the CEC 2019 and CEC 2022 competitions. The performance of SDP was compared against seven established meta-heuristic algorithms: Artificial Hummingbird Algorithm (AHA), Dwarf Mongoose Optimization algorithm (DMO), Reptile Search Algorithm (RSA), Snake Optimizer (SO), Fick's Law Optimization (FLA), Prairie Dog Optimization (PDO), and Gazelle Optimization Algorithm (GOA). Statistical analysis was conducted using Friedman's rank and Wilcoxon signed-rank tests to evaluate the relative performance in terms of exploration, exploitation capabilities, and proximity to the optimum solution.

Results: The results demonstrated that the SDP outperforms its counterparts in terms of convergence speed and hit rate across the selected test functions. In statistical tests, SDP showed significantly better performance in exploration and exploitation, leading to a higher proximity to optimum compared to other algorithms. Furthermore, when applied to five complex engineering design problems, the SDP algorithm exhibited superior performance, outmatching the state-of-the-art algorithms in terms of effectiveness and efficiency.

Conclusion: The Society Deciling Process (SDP) algorithm introduces a novel and effective approach to optimization, inspired by societal structure dynamics. Its superior performance in convergence speed, exploration and exploitation capabilities, and application to complex engineering problems establishes SDP as a promising meta-heuristic algorithm. This research not only demonstrates the potential of socio-inspired algorithms in optimization tasks but also opens avenues for further enhancements in meta-heuristic algorithm designs.

Introduction

In the ever-evolving landscape of technology, optimization problems have become pervasive across diverse domains such as science, engineering, management, and economics. The complexity of these challenges, coupled with resource constraints, necessitates optimal solutions. The primary objective in solving optimization problems lies in determining the most suitable values for

variables to either maximize or minimize a given objective function [1]. However, conventional mathematical techniques and heuristic algorithms encounter a significant hurdle—local optima entrapment—when dealing with intricate problems featuring numerous candidate values and multiple local optima.

To address this challenge, meta-heuristic algorithms have emerged as a powerful class of stochastic optimization methods. Distinguished by their gradient-free, problem-independent, and local-optima-free nature, these algorithms offer a departure from traditional approaches [2]. By applying various operators iteratively, they navigate the search space based on an objective function [3], making them well-suited for a wide array of scientific and industrial applications [3].

Meta-heuristic algorithms can be broadly categorized into two groups: individual-solution-based and population-based [4]. The latter, starting with a randomly generated population of solutions, proves more popular due to its enhanced ability to explore and exploit the search space, targeting global optima [5]. Inspired by evolutionary processes, natural phenomena, and social behaviors, population-based meta-heuristic algorithms can be further classified into Evolutionary Algorithms (EAs), Natural Phenomenon (NP) algorithms, and Social Behaviors (SBs) algorithms:

1. Evolutionary Algorithms (EAs) that mimic the process of natural evolution. Genetic Algorithm (GA) [6], and Genetic Programming (GP) [7], Single Candidate Optimizer (SCO) [8], and Attack-Leave Optimizer (ALO) [9] are the well-known samples of this class.

2. *Natural Phenomenon (NP) algorithms* that exploit physical and chemistry principles. Simulated Annealing (SA) [10], Energy Valley Optimizer (EVO) [11], Water Cycle Algorithm (WCA) [12], Nutcracker optimizer (NOA) [13], Orchard Algorithm (OA) [14], Swarm Magnetic Optimizer (SMO) [15], Fusion–fission optimization (FuFiO) [16], and Fick's Law Optimization (FLA) [17], Geometric Mean Optimizer (GMO) [18] and Physics-Inspired Discriminative Classifier (PIDC) [19] can be considered as popular samples of this class.

3. *Social Behaviors (SBs) algorithms* which are divided into two subgroups: Swarm Intelligence (SI) algorithms and Human Behaviors (HB) algorithms. SI algorithms simulate the self-organized and collective behaviors in nature. Actually, these algorithms originate social behaviors of species, such as ants or bees. Ant Colony Optimization (ACO) inspired by foraging behaviors of ants is a popular algorithm in this group. Particle Swarm Optimization (PSO) [20], Snake Optimizer (SO) [21], Prairie Dog Optimization (PDO) [22], Aquila Optimizer (AO) [23], Red Fox Optimization (RFO) algorithm [24], Honey Badger Algorithm (HBA) [1], Reptile Search Algorithm (RSA) [25], Gazelle Optimization Algorithm (GOA) [26], Artificial Hummingbird Algorithm (AHA) [3], Dung Beetle Optimizer [27], Fox optimizer (FOX) [28], Giant Trevally Optimizer (GTO) [29], Mountain Gazelle Optimizer (MGO) [30], and Dwarf Mongoose Optimization (DMO) algorithm [31], Hippopotamus Optimization (HO) [32], Blood-Sucking Leech Optimizer (BSLO) [33], Greylag Goose Optimization (GGO) [34], Genghis Khan Shark Optimizer (GKSO) [35], Ladybug Beetle Optimization (LBO) [36], and Crayfish Optimization Algorithm (COA) [37] can be mentioned as other popular algorithms in this class.

HB algorithms imitate human behaviors in society. Social Group Optimization (SGO) [38] is a popular HB algorithm inspired by the social interaction of members. This algorithm has two phases: *improving* and *acquiring*. In the first phase, each group member increases his/her knowledge by interacting with the best member of the group. In the second phase, each member acquires knowledge from the best member of the group as well as other randomly selected members. Inspired from the council evolution, City Councils Evolution (CCE) [39] is another well-known HB algorithm. Here, councils evolve from the smallest neighbors to the largest ones, regions, and ultimately the whole city is considered. Due to the hierarchical manner of councils' evolution, CCE uses a hierarchical structure like a tree to model council members and bosses. Teaching–Learning-Based Optimization (TLBO) [40] is another popular HB algorithm inspired by the teaching and learning phenomenon in a classroom. TLBO has two phases named *teacher* and *student*. In the teacher phase, the fittest individual is selected as a teacher and the learning process is done by the teacher. While, in the student phase, all students play the teacher role and learning is carried out by the student interactions. Parliamentary Optimization Algorithm (POA) [41], Ideology Algorithm (IA) [42], Intelligent Clonal Optimizer (ICO) [43], Expectation Algorithm (ExA) [44], and Seasons Optimization (SO) [45] are the well-known HB algorithms.

The question that arises is why additional meta-heuristic optimization algorithms are necessary despite the existence of various ones. In response, it should be noted that technological advancements have unveiled new optimization problems. Based on the No Free Lunch (NFL) theorem [46], existing algorithms may not universally solve these problems. The NFL theorem asserts the absence of a single, all-encompassing meta-heuristic algorithm applicable to every optimization problem. Consequently, this paper introduces a novel socio-inspired meta-heuristic optimization algorithm known as the society deciling process (SDP). SDP emulates the societal deciling process based on factors

such as monthly income, occupation, and education.

To assess and compare the efficacy of SDP against AHA [3], DMO [31], RSA [25], SO [21], PDO [22], FLA [17], and GOA [26], experiments are conducted using 14 general test functions, 9 from CEC 2019, and 12 from CEC 2022. The evaluation criteria include the effectiveness in finding solutions closest to the optimum, early convergence, and hit rate (accuracy).

The key contributions of this paper include the introduction of the novel SDP algorithm, which demonstrates superior convergence speed and hit rate compared to existing algorithms. Our analysis, validated through Friedman's test, establishes the statistical superiority of SDP in terms of finding solutions closest to optima. Additionally, the paper extends the evaluation to three constrained engineering design problems, affirming the algorithm's effectiveness across diverse applications.

SDP proves particularly suitable for addressing optimization challenges in various fields, including but not limited to science, engineering, management, and economics.

Its robust performance, especially in scenarios with intricate and dynamic parameters, positions SDP as a versatile and effective tool for solving real-world optimization problems.

The remaining sections of the paper are outlined as follows. Initially, the inspiration and intricacies of the SDP algorithm will be presented. Following that, the experimental results of SDP across general, CEC 2019, and CEC 2022 test functions will be provided in the Experimental Results section. Subsequently, a statistical analysis of the results is conducted using the Friedman test. Moreover, the proposed algorithm (SDP) on five constrained engineering design problems is evaluated in this section. Finally, the conclusion section concludes the paper, summarizing the findings and offering directions for future research endeavors.

Society Deciling Process (SDP): The Presented Optimization Algorithm

This section delineates the Society Declining Process (SDP) algorithm, focusing on its inspiration and intricacies.

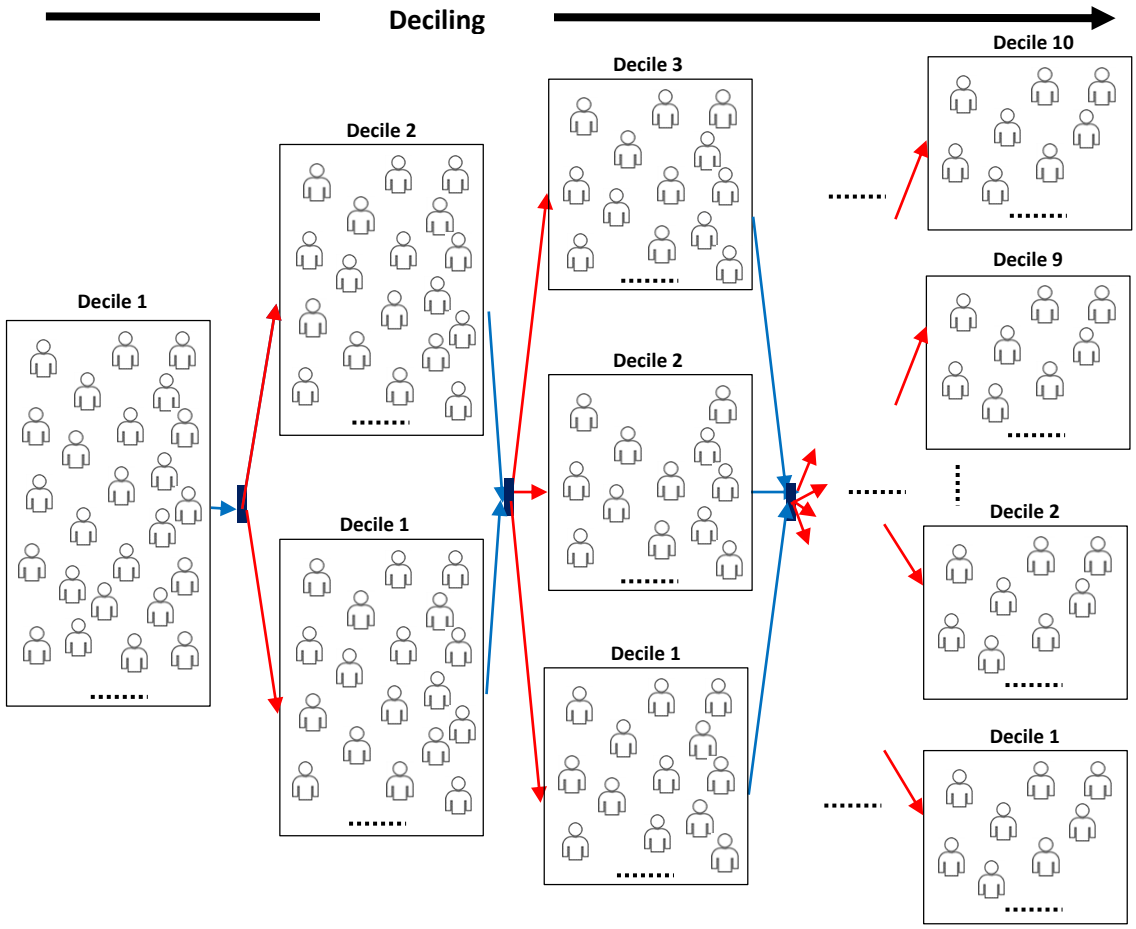


Fig. 1: The deciling process of all people in a society.

A. Inspiration

Adverse economic conditions, characterized by economic instability, high inflation, reduced purchasing power, and economic discrimination, significantly impact people's living standards, encompassing nutrition, health, and education. Recognizing vulnerable individuals and empowering them is pivotal. To achieve this, the population is categorized into deciles based on criteria like monthly income, occupation, and education. Regular evaluations, considering changes in these criteria, lead to modifications in the society's decile distribution.

Initially, all individuals are placed in the first decile. As some individuals become relatively wealthier, the society divides into two deciles (one and two), with this process continuing until all individuals are distributed across ten deciles. Fig. 1 illustrates the progressive deciling of society.

B. Details of SDP

In the SDP algorithm, akin to other meta-heuristic methods, the process begins with a randomly generated population X, illustrated in Fig. 2. This population consists of N vectors, each of length m, representing the individuals and their respective variables related to deciling criteria. The variables ($x_{i,j}$) are determined using the (1):

$$x_{i,j} = rand * (u_j - l_j) + l_j \tag{1}$$

Here, *rand* is a random value in the interval [0, 1] where l_j and u_j denote the lower and upper bounds of j-th variable. The fitness of each solution X_i for the test function *f* is computed using the (2):

$$\text{fitness} = f(X_i) = f(x_{i,1}, x_{i,2}, \dots, x_{i,m}) \tag{2}$$

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \cdots & \end{pmatrix}$$

Fig. 2: Population X with N candidate solutions.

In alignment with the societal deciling process (Fig. 1), the initial placement involves assigning all individuals to the first decile. After altering the average values of deciling criteria for some individuals, the society is iteratively divided into *two* deciles. Consequently, the entire initial population is initially placed in the *first* decile. Following fitness computation, solutions are sorted and distributed into *two* deciles (1 and 2), with the *N/2* solutions possessing the highest fitness placed in decile 2, and the rest in decile 1, establishing a 2-decile state.

In a general *d*-decile state ($2 \le d \le 10$), the fitness-based sorted population *X* is partitioned into *d* subsets of size *decN* (= *N/d*). These subsets are then allocated to the respective deciles (*d, d-1, ..., 1*). Fig. 3 visually represents the population *X* in a *d*-decile state. Denoting the maximum iteration number of the SDP algorithm as *MaxIter*, the number of iterations for each d-decile state ($2 \le d \le 10$) is *dIter* (= *MaxIter/9*). During each *d*-decile state iteration, the SDP algorithm follows these steps *dIter* times:

Step 1: Evaluate the population *X* using the test function *f* and sort *X* based on fitness values.

Step 2: Divide *X* into *d* partitions, each with the size of *decN*.

Step 3: For each deciling criterion, reposition all solutions in decile *k* (*k = d, d-1, ..., 1*) based on higher deciles *j* ($k \le j \le d$) using the *Repose* function. Algorithm 1 details the pseudo-code of the *Repose* function, which adjusts the position of a solution *P* (in decile *k*) based on the fittest solution in the same decile (B_k) and those in higher deciles (B_j). The weighted average (*wAvg*) of B_j in higher deciles is calculated using (3), emphasizing the importance of higher deciles. The final position is determined by combining the average of *wAvg* and B_k using a modified arithmetic crossover.

$$\boldsymbol{wAvg} = \frac{\sum_{j=k+1}^{d} (\boldsymbol{j} - \boldsymbol{k}) * \boldsymbol{Bj}}{\sum_{j=k+1}^{d} (\boldsymbol{j} - \boldsymbol{k})} \tag{3}$$

Algorithm 2 outlines the steps of the Society Declining Process (SDP) algorithm. In Line 2, a population *X* comprising candidate solutions is generated randomly. The fitness values of these candidates are then computed using the *calcFitness* function, and the entire population sort is accomplished according to these fitness values.

In Line 9, the algorithm proceeds to compute the first (*sIndex*) and last (*eIndex*) indices of the current decile. Subsequently, in Lines 10-13, the positions of all solutions within the specified range ($sIndex \le j \le eIndex$) are adjusted using the *Repose* function. Importantly, if the result of the *Repose* operation yields a solution with a higher fitness value, it replaces the original solution in the population *X*.

To better understand this algorithm, nested iteration loops are explained in more detail: 1) The first *for* loop in line 4: in this loop, the variable *d* controls the index of the *d*-decile state, which starts from 2 and ends in 10. 2) The second *for* loop in line 6: in this loop, the variable *iter* controls the number of algorithm repetition for the *d*-decile state, which starts from 1 and continues to *diter*

(=*MaxIter*/9). 3) The third *for* loop in line 7: in this loop, the variable *cr* controls the number of deciling criteria (the variable number of test functions), which starts from 1 and continues to *m*. 4) The fourth *for* loop in line 8: in this loop, the variable *k* controls the index of higher deciles, which starts from *d* and continues to 1. 5) The fifth *for* loop in line 10: in this loop, the variable *j* controls the indices of the current decile, which starts from *sIndex* = (*d*-*k*)**decN*+1 and continues to *eIndex* = (*d*-*k*+1)**decN*.

Additionally, Fig. 4 provides a visual representation of the algorithm's flowchart, offering a clear illustration of the sequential steps involved in the SDP algorithm. This visual aid enhances the understanding of the algorithm's execution and aids in visualizing the interplay of operations during each iteration.

Fig. 3: Population X in the d-decile state. Bk (1 ≤ k ≤ d) indicates the fittest solution in decile k.

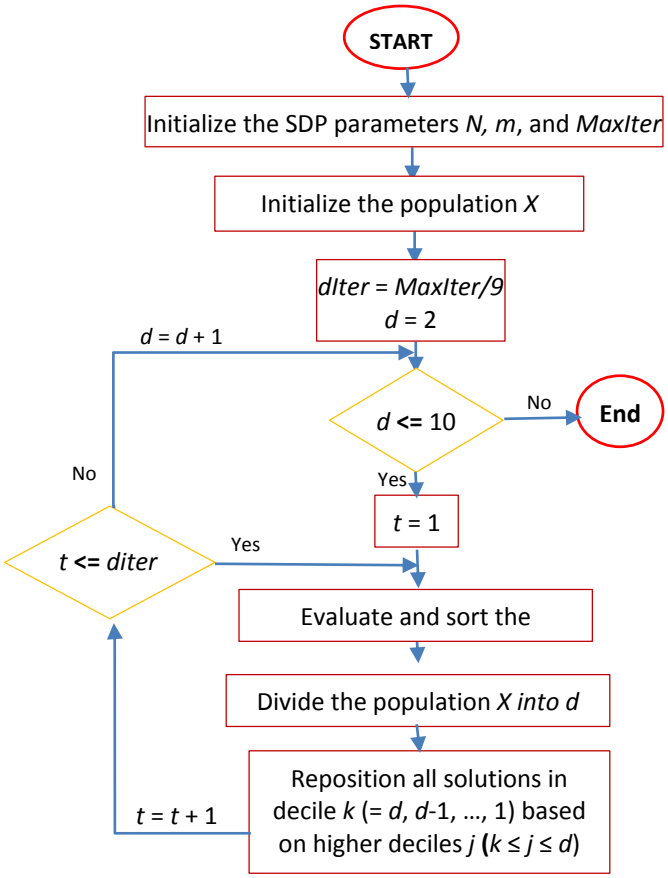


Fig. 4: The flowchart of the SDP algorithm.

Algorithm 1: The ***repose*** function

Input: *P*: a given solution, *k*,*d*;
Output: an individual with the highest fitness.
1: B_j = the fittest solution in decile *j*, ($k \le j \le d$);
2: Compute *wAvg* using Eq. (3);
3: *z* = a value is randomly selected as either 1 or 2;
4: *rnd* = a random value in the interval [0, 1];
5: $\alpha = (-1)^{k+1} * rnd$;
6: Y1 = α**P* +(1-α)*(B_k + *wAvg*)/2; Y2 = (1-α)**P* +α*(B_k + *wAvg*)/2;
7: **return** *max* (Y1, Y2);

Algorithm 2: The **SDP** algorithm

Input: *N*: the population size, *MaxIter*, *m*: the number of deciling criteria,
f : a test function.
Output: a solution with the best fitness.
1: Array [][] *X* = new Array [*N*][*m*]; Array [] *fit* = new Array [*N*];
2: *X* = createInitial (*N*, *m*); *fit* = calcFitness (*X, f*); Sort (*X* , *fit*); *best* = *X*[1]; *bestFit* = *fit*[1];
3: *diter* = *MaxIter/9*;
4: **for** *d* =2 to 10 **do**
5: *decN* = *N/d;*
6: **for** *iter* = 1 to *dIter* **do**
7: **for** *cr* = 1 to *m* **do**
8: **for** *k* = *d* downto 1 **do**
9: *sIndex*= (*d*-*k*)**decN*+1; *eIndex*= (*d*-*k*+1)**decN*;
10: **for** *j* = *sIndex* to *eIndex* **do**
11: *P* = *repose* (*X*[*j*], *k*, *d*);
12: **if** *f* (*P*) > *fit* [*j*] **then** M[*j*] =*P*; *fit*[*j*] = *f* (*P*);
13: **end for** *j*
14: **end for** *k*
15: *fit* = calcFitness (*X, f*); Sort (*X* , *fit*);
16: **end for** *cr*
17: **if** *fit* [1]>*bestFit* **then** *best* = *X*[1]; *bestFit* = *fit*[1];
18: **end for** *iter*
19: **end for** *d*
20: **return** best;

Experimental Results

To evaluate the proposed algorithm, i.e., SDP, four distinct experiment set were conducted. The obtained results were then compared with outcomes generated by seven established meta-heuristic algorithms, namely AHA [3], DMO [31], RSA [25], SO [21], PDO [22], FLA [17], and GOA [26]. The following experiments were carried out for comparison purposes:

1) The first experiment applies 14 general test functions to evaluate the exploration and exploitation capabilities of the underlying algorithms. Here, exploitation refers to the algorithm's capacity to enhance the quality of promising solutions through local search, while exploration ability pertains to the algorithm's capability to freely explore different areas of the landscape, thereby avoiding local optima.

2) The second experiment utilizes 9 test functions from CEC 2019 to evaluate the efficiency of algorithms in both exploration and exploitation aspects.

3) The third experiment encompasses 12 test functions from CEC 2022, serving as a basis for comparing algorithm efficiency in relation to exploration and exploitation.

4) The fourth experiment employs five challenging engineering design problems to validate the performance of SDP. The results are then compared with those of various algorithms from the existing literature.

Meta-heuristic algorithms involve various parameters that can be adjusted to suit different optimization problems. These parameters play a crucial role in striking a balance between exploration and exploitation capabilities. This equilibrium allows the algorithm to

navigate diverse regions within the state space. Additionally, when necessary, the algorithm can perform a local search around promising solutions to reach the optimal solution. Hence, determining appropriate values for these parameters is of utmost importance. Notably, algorithms with fewer parameters tend to perform better.

Fortunately, the SDP algorithm does not require dedicated parameter tuning. The parameter settings for SDP and other algorithms are presented in Table 1. To ensure a fair performance comparison among different algorithms, it is essential to set a consistent maximum number of fitness function evaluations (*MNFFE*), considering that the number of evaluations may vary between algorithms. Consequently, the maximum iteration number (*MaxIter*) can be derived from *MNFFE*. Assuming *NFFE* represents the fitness function calling number in each iteration of the algorithm, *MaxIter* can be calculated as *MNFFE*/*NFFE*.

For instance, in the SDP algorithm, *NFFE* is expressed as *2×m×N*, leading to *MaxIter* being *MNFFE*/(*2×m×N*). It's worth noting that the parameters *N* and *MNFFE* are universally set at 30 and 10E+5, respectively, for all algorithms.

The outcomes were derived from the execution of algorithms in the Matlab 2017a environment, performed 30 times using an Intel Core i5 CPU and 6GB RAM. The result tables present the best (*Best*), average (*Ave*), and standard deviation (*Std*) of the solutions identified as the best so far across all runs.

Table 1: Assumed parameter values of all algorithms (*N* = 30 and *MNFFE* = 10E+5)

Algorithm	Parameters and their appropriate values
AHA	*Migration coefficient* = 2×*N*
DMO	*Number of babysitters* = 3, *Alpha female vocalization* = 2
RSA	*Alpha* = 0.1, *Beta* = 0.005
SO	*Threshold* = 0.25, *Threshold2* = 0.6, *C1*= 0.5, *C2* = 0.05, *C3* = 2
PDO	*Rho* = 0.005, *epsPD* = 0.1
FLA	*C1* = 0.5, *C2* = 2, *C3* = 0.1, *C4* = 0.2, C5 = 2, *D* = 0.01
GOA	*PSRs*=0.34, *S*=0.88

C. Parameter Analysis

The effectiveness of the SDP algorithm is significantly influenced by certain parameters, such as *N* and *m* (representing the dimension of test functions). To assess this impact, the SDP is executed for various *N* values, considering F1 and F2 from the general test functions, F3 and F4 from the CEC 2019 functions, and F2 and F3 from the CEC 2022 functions. It is worth noting that the maximum number of fitness function evaluations (MNFFE) is set at 30000. As indicated in Table 2, the best effectiveness for SDP is observed when N is equal to 30.

To analyze the effect of *m* on SDP, the algorithm is executed for different values of this parameter across test functions F4, F5, F6, and F7 belonging to CEC 2022. It is important to note that general and CEC 2019 test functions are defined solely with dimension, necessitating the selection of CEC 2022 functions, which are specified for three different dimensions. Table 3 reveals that SDP exhibits optimal effectiveness when m is set to 2 in the majority of the considered test functions.

Furthermore, the table illustrates the impact of increasing the dimension of test functions on the algorithm's execution time. According to the table, as the dimension of the test functions increases, the algorithm's execution time rises due to the escalated number of fitness function evaluations, resulting in a reduction in the number of algorithm repetitions.

D. Experiment 1: 14 General Test Functions

In this experiment, 14 general test functions are used to analyze and compare SDP and others with regard to exploitation and exploration abilities. Table 4 [47] provides a more detailed descriptions of these functions. Functions F1-F5 are unimodal with single global optima, while functions F6-F9 are multimodal, featuring a global optimum and several local optima. Consequently, finding an optimal solution in test functions such as F6-F9 serves as a robust benchmark for evaluating and comparing the exploration and local optima avoidance capabilities of SDP and other algorithms. Additionally, functions F10-F14 exhibit shifted and rotation attributes, presenting increased complexity for a more precise evaluation of SDP and others. Table 5 presents the obtained results of SDP and other algorithms. Furthermore, the Friedman test is employed to rank the performance of SDP and others—a non-parametric statistical hypothesis test suitable for multiple comparisons of related data sets [48]. Here, first the rank of algorithms in each test

function is determined and then the mean ranks across all test functions are calculated. An algorithm achieving rank 1 will be the best indicating the least mean rank, while the highest rank signifies the worst performance. The results of this test, along with the hit rate based on the reported results in Table 5, are depicted in Fig. 5. As demonstrated in this figure, the first and second ranks are obtained by SO and SDP, respectively. Moreover, the highest hit rate (0.50) after RSA, PDO, and SO is achieved by SDP. In other words, in 7 test functions (i.e., 0.50*14), SDP generates the exact solution.

In addition to utilizing the Friedman test, researchers can employ the Wilcoxon signed-rank test to analyze the obtained results. This non-parametric statistical method, designed for comparing two samples [49], facilitates the determination of cases where Algorithm X outperforms, underperforms, or demonstrates similar performance to Algorithm Y. To fulfill this purpose, three key test statistics, namely R^-, R^+, and $R^=$, are utilized. Fig. 6 illustrates the Wilcoxon signed-rank test outcomes for the pairwise comparison of SDP versus AHA, DMO, RSA, SO, PDO, FLA, and GOA.

The analysis depicted in Fig. 6 reveals that, in the majority of pairwise comparisons, the values of R^- surpass those of R^+. This observation suggests that SDP exhibits superior effectiveness compared to most other algorithms under consideration.

Another criterion that can be employed for comparing the efficiency of our considered algorithms is convergence speed. The optimal solution is reached sooner with a higher convergence speed of an algorithm. To achieve this, test functions F11, F12, F13, and F14, characterized by being shifted and rotated, are considered, and all algorithms are executed up to the number of fitness function evaluations equal to 10E+5. The convergence curve of SDP and others in these test functions is depicted in Fig. 7. In most of these functions, the fastest convergence is observed with SDP.

Table 2: The results of executing SDP for different *N*

Type	Function	Metric	*N* = 10	*N* = 20	*N* = 30	*N* = 40	*N* = 50
General CEC 2019	F1	Best	3.1824e-43	7.0545e-44	**1.7861e-65**	1.0526e-32	7.8267e-26
		Ave	5.522e-43	5.7028e-43	6.2165e-65	2.4259e-32	1.719e-25
		Std	2.5488e-43	8.0896e-43	6.5017e-65	1.1896e-32	8.9457e-26
	F2	Best	1.3689e-16	**6.3971e-33**	5.6104e-22	1.0191e-16	1.7863e-13
		Ave	1.0695e-15	1.0718e-32	7.2167e-22	1.6129e-16	2.1227e-13
		Std	1.5509e-15	4.3078e-33	2.5795e-22	5.9981e-17	5.3591e-14
	F3	Best	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**
		Ave	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01
		Std	5.7793e-11	1.4992e-09	2.3484e-09	1.4422e-09	1.5322e-09
	F4	Best	12.775	12.585	**8.792**	13.325	12.313
		Ave	16.369	14.728	13.841	14.703	13.57
		Std	5.7466	14.728	6.6563	1.299	1.0942
CEC 2022	F2	Best	400.01	**400**	**400**	400.01	400.01
		Ave	400.05	400.07	400.04	400.10	400.12
		Std	0.063384	0.090989	0.06685	0.15129	0.15441
	F3	Best	**600**	**600**	**600**	**600**	**600**
		Ave	600	600	600	600	600
		Std	8.0389e-14	6.00435e-14	0.0	0.0	0.0

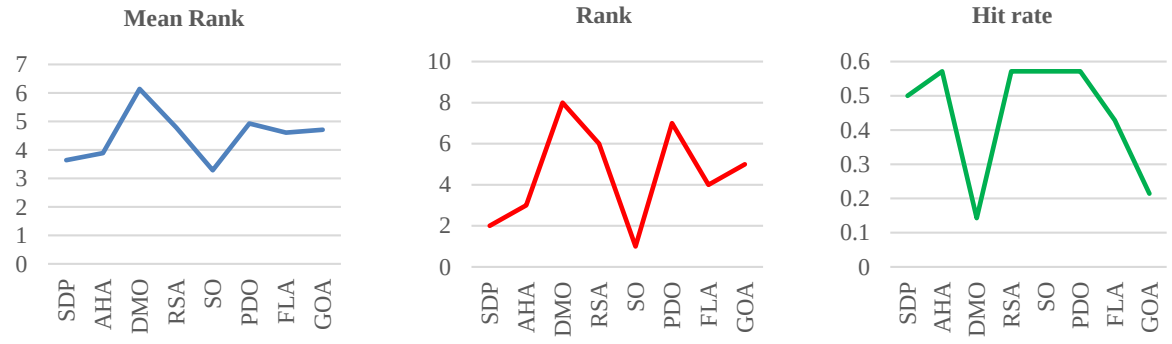


Fig. 5: Results of the Friedman's rank test along with hit rate for general test functions.

Table 3: The results of executing SDP for different *m* (the dimension of test functions)

Type	Function	Metric	*m* = 2	*m* = 10	*m* = 20
CEC 2022	F4	Best	**800**	811.96	898.77
		Ave	800	814.3	910.4
		Std	5.5482e-06	4.0129	10.674
		Time	2.0811	1.8952	1.7664
	F5	Best	**900**	910.74	2942.9
		Ave	900	994.73	3458.2
		Std	0.0	73.803	634.74
		Time	2.2686	1.9223	1.9048
	F6	Best	**1800**	1851.3	4520.3
		Ave	1800	1994.1	6590.2
		Std	0.0	244.61	1927.8
		Time	2.1512	2.0323	2.0964
	F7	Best	2463.5	**2000.5**	2024.3
		Ave	2463.5	2001	2031.5
		Std	0.0	0.42429	7.7289
		Time	3.4253	3.0333	3.9496

Table 4: The general test function details

Name	Dimension	Range	Optimum value	Type
F1 (Sphere)	30	[-100, 100]	0	Unimodal
F2 (Schwefel 2.22)	30	[-10, 10]	0	
F3 (Schwefel 1.2)	30	[-100, 100]	0	
F4 (Rosenbrock)	30	[-10, 10]	0	
F5 (Step)	30	[-100, 100]	0	
F6 (Schwefel's)	30	[-500, 500]	0	
F7 (Rastrigin)	30	[-.5.12, 5.12]	0	Multimodal
F8 (Ackley)	30	[-32, 32]	0	
F9 (Griewank)	30	[-600, 600]	0	
F10 (Shifted Schwefel's P1.2)	30	[-100, 100]	-450	Shifted and Rotated
F11 (Shifted Rotated High Conditioned Elliptic)	30	[-100, 100]	-450	
F12 (Shifted Rosenbrock's)	30	[-100, 100]	390	
F13 (Shifted Rotated Rastrigin's)	30	[-5, 5]	-330	
F14 (Shifted Rotated Weierstrass)	30	[-0.5, 0.5]	90	

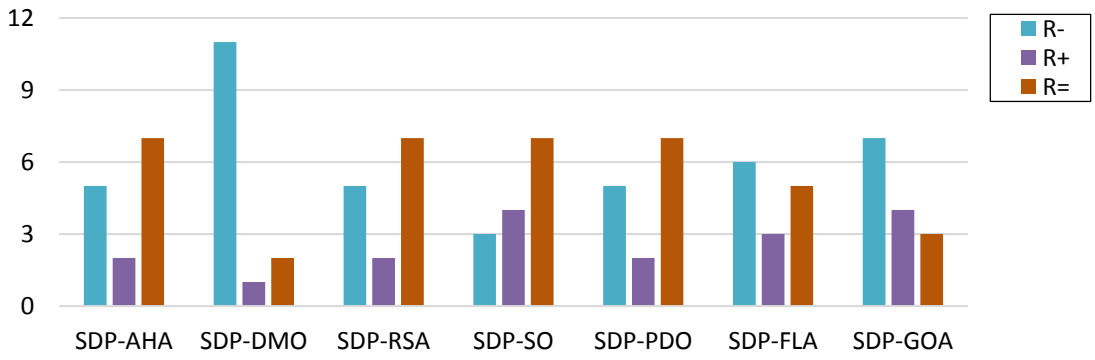


Fig. 6: Results of the Wilcoxon Signed-Rank Test on General Test Functions.

Table 5: Comparative results for general test functions

Function	Metric	SDP	AHA	DMO	RSA	SO	PDO	FLA	GOA
	Best	**0.00E+00**	**0.00E+00**	1.89E-24	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	1.81E-203
F1	Ave	0.00E+00	0.00E+00	4.20E-24	0.00E+00	0.00E+00	0.00E+00	0.00E+00	5.89E-101
	Std	0.00E+00	0.00E+00	2.15E-24	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.32E-100
	Best	**0.00E+00**	**0.00E+00**	1.64E-18	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	6.11E-122
F2	Ave	0.00E+00	0.00E+00	1.51E-17	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.02E-43
	Std	0.00E+00	0.00E+00	1.29E-17	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.52E-43
	Best	**0.00E+00**	**0.00E+00**	2.07E-22	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	1.33E-206
F3	Ave	0.00E+00	0.00E+00	6.09E-22	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.95E-119
	Std	0.00E+00	0.00E+00	3.66E-22	0.00E+00	0.00E+00	0.00E+00	0.00E+00	8.83E-119
	Best	1.15E+01	2.35E+01	2.14E+01	**1.06E-28**	2.29E-01	8.14E-02	6.82E-01	2.23E+01
F4	Ave	1.51E+01	2.35E+01	2.33E+01	**1.45E+01**	1.46E+00	3.29E+00	1.30E+00	2.32E+01
	Std	3.44E+00	4.83E-03	1.73E+00	**2.05E+01**	1.74E+00	4.54E+00	8.47E-01	5.64E-01
	Best	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
F5	Ave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Std	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Best	8.66E+04	**0.00E+00**	3.37E+05	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	1.75E-35
F6	Ave	1.28E+05	0.00E+00	5.33E+05	0.00E+00	0.00E+00	0.00E+00	0.00E+00	9.80E-11
	Std	3.75E+04	0.00E+00	1.70E+05	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.19E-10
	Best	**0.00E+00**	**0.00E+00**	6.71E+01	**0.00E+00**	**0.00E+00**	**0.00E+00**	9.95E-01	**0.00E+00**
F7	Ave	0.00E+00	0.00E+00	8.97E+01	0.00E+00	0.00E+00	0.00E+00	1.35E+00	0.00E+00
	Std	0.00E+00	0.00E+00	3.23E+01	0.00E+00	0.00E+00	0.00E+00	6.15E-01	0.00E+00
	Best	**0.00E+00**	**0.00E+00**	1.77E-12	**0.00E+00**	**0.00E+00**	**0.00E+00**	3.55E-15	3.55E-15
F8	Ave	0.00E+00	0.00E+00	2.03E-12	0.00E+00	1.78E-15	0.00E+00	3.55E-15	3.55E-15
	Std	0.00E+00	0.00E+00	2.27E-13	0.00E+00	2.51E-15	0.00E+00	0.00E+00	0.00E+00
	Best	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
F9	Ave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Std	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Best	5.04E+03	6.24E+01	2.71E+04	3.37E+04	**1.66E+01**	4.35E+04	1.57E+02	1.37E+02
F10	Ave	7.37E+03	1.65E+02	3.27E+04	3.76E+04	1.89E+01	6.08E+04	1.95E+02	4.66E+02
	Std	3.05E+03	1.45E+02	5.04E+03	5.46E+03	3.27E+00	2.44E+04	3.28E+01	2.39E+02
	Best	**1.84E+06**	2.49E+06	1.36E+08	2.86E+08	4.33E+06	4.01E+08	9.82E+06	8.12E+05
F11	Ave	1.07E+07	4.11E+06	2.25E+08	6.51E+08	6.42E+06	6.27E+08	1.36E+07	3.58E+06
	Std	1.61E+06	2.29E+06	7.85E+07	5.17E+08	2.95E+06	3.19E+08	3.44E+06	2.19E+06
	Best	**1.16E+01**	1.98E+02	2.49E+01	1.37E+10	2.03E+02	1.01E+10	2.40E+02	4.26E+02
F12	Ave	4.16E+01	2.01E+02	8.86E+01	2.29E+10	2.16E+02	1.06E+10	3.87E+02	1.92E+03
	Std	2.18E+01	4.62E+00	1.02E+02	1.31E+10	1.87E+01	6.88E+08	1.74E+02	2.39E+03
	Best	2.30E+02	2.83E+02	2.04E+02	6.07E+02	**1.16E+02**	5.01E+02	3.28E+02	9.44E+01
F13	Ave	2.67E+02	4.15E+02	2.14E+02	6.15E+02	1.51E+02	6.25E+02	4.05E+02	1.15E+02
	Std	3.64E+01	1.87E+02	1.08E+01	1.19E+01	5.06E+01	1.76E+02	9.61E+01	1.31E+01
	Best	**1.88E+01**	3.05E+01	3.82E+01	4.03E+01	2.08E+01	4.07E+01	3.05E+01	1.96E+01
F14	Ave	2.70E+01	3.13E+01	3.98E+01	4.12E+01	2.67E+01	4.07E+01	3.23E+01	2.37E+01
	Std	5.67E+00	1.18E+00	1.46E+00	1.32E+00	8.29E+00	1.09E-01	1.53E+00	3.72E+00

E. Experiment 2: CEC 2019 Test Functions

In this experiment, the performance of SDP is compared with that of the considered algorithms in terms of both exploration and exploitation for solving the test functions designed for CEC 2019 [50]. These functions are of a multimodal type with one global minimum. Furthermore, functions F4 to F10 possess few movement and rotation attributes, whereas functions F1 to F3 have default attributes. The details of these functions, such as name, dimension, range, and optimum value, are presented in Table 5. According to this table, functions F1 to F3 have different dimensions and ranges, whereas F4 to F10 are 10-dimensional and fall within the interval [-100, 100]. Additionally, the optimum value for all functions is set to 1. Some algorithms, such as SDP and AHA, exhibit peculiar behavior when solving the F7 test function, leading to the exclusion of this test function from the test suite. The results of all algorithms on solving CEC 2019 test functions are reported in Table 7.

The statistical comparison of SDP and other algorithms in solving the test functions of CEC 2019 involves the use of the Friedman test. The results of this test, presented in Fig. 8 along with the hit rate in Table 5, confirm that SDP exhibits the least mean rank, signifying superior performance in identifying solutions closest to an optimum.

This observation implies that powerful balancing of exploitation and exploration abilities is achieved by SDP compared to other algorithms. Furthermore, the highest hit rate (0.11) among the considered algorithms is attained by SDP, DMO, and RSA, as depicted in the third diagram of Fig. 8.

Furthermore, alongside the Friedman test, the Wilcoxon signed-rank test highlights that across all paired comparisons, the values of R^- consistently surpass R^+, suggesting that SDP exhibits superior effectiveness compared to the majority of other algorithms, as depicted in Fig. 9.

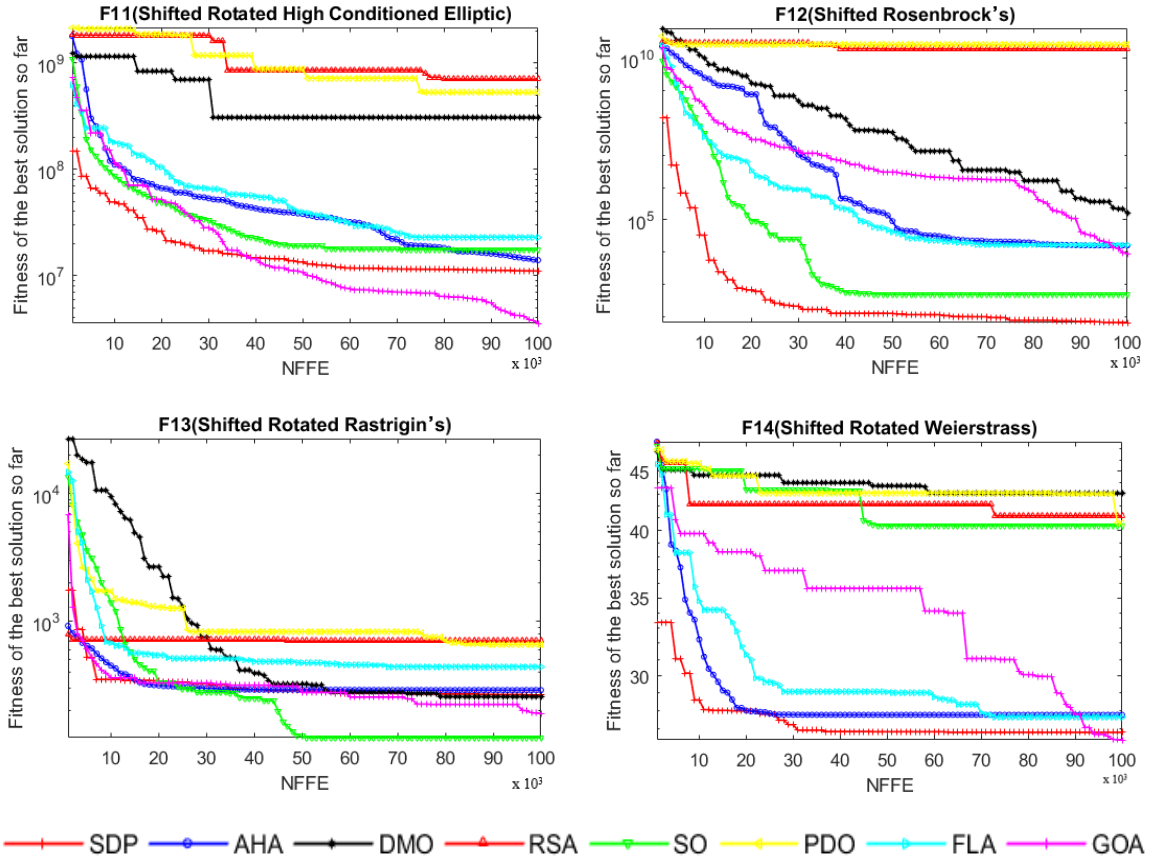


Fig. 7: Convergence curve of all algorithms for general test functions.

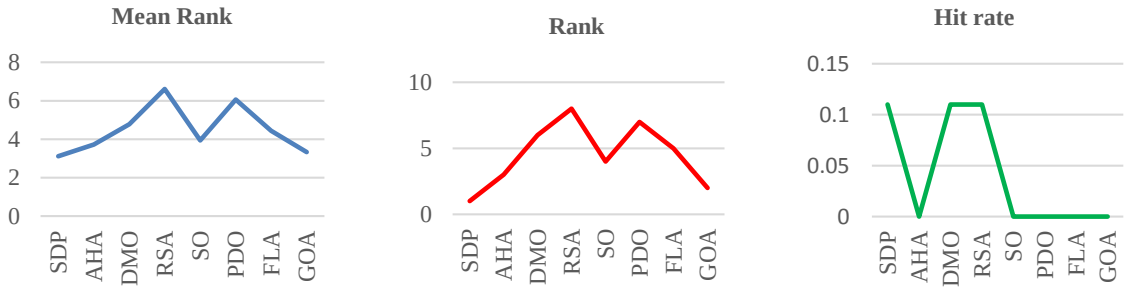


Fig. 8: Results of the Friedman's rank test along with hit rate for CEC 2019 test functions.

Fig. 9: Results of the Wilcoxon signed-rank test to the CEC 2019 test functions.

Table 6: The details of CEC 2019 test functions

Function	Name	D	Range	Optimum value
F1	Storn's Chebyshev Polynomial Fitting Problem	9	[-8192,8192]	1
F2	Inverse Hilbert Matrix Problem	16	[-16384,16384]	1
F3	Lennard-Jones Minimum Energy Cluster	18	[-4,4]	1
F4	Rastrigin's Function	10	[-100,100]	1
F5	Griewank's Function	10	[-100,100]	1
F6	Weierstrass Function	10	[-100,100]	1
F8	Expanded Schaffer's F6 Function	10	[-100,100]	1
F9	Happy Cat Function	10	[-100,100]	1
F10	Ackley Function	10	[-100,100]	1

Table 7: Comparative results for CEC 2019 test functions

Function	Metric	SDP	AHA	DMO	RSA	SO	PDO	FLA	GOA
	Best	3.23E+10	3.74E+04	2.34E+09	4.36E+04	2.58E+05	4.62E+04	4.62E+04	**3.14E+04**
F1	Ave	4.07E+10	3.76E+04	5.70E+09	1.22E+05	2.60E+06	5.65E+04	5.65E+04	3.14E+04
	Std	1.19E+10	1.80E+02	3.72E+09	1.11E+05	3.15E+06	1.45E+04	1.45E+04	3.64E-12
	Best	**1.73E+01**	**1.73E+01**	**1.73E+01**	1.80E+01	**1.73E+01**	1.74E+01	1.74E+01	**1.73E+01**
F2	Ave	1.73E+01	1.73E+01	1.73E+01	1.80E+01	1.73E+01	1.74E+01	1.74E+01	1.73E+01
	Std	0.00E+00	4.35E-15	0.00E+00	1.70E-05	0.00E+00	1.25E-02	6.15E-03	0.00E+00
	Best	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**	**1.27E+01**
F3	Ave	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01
	Std	1.04E-09	1.26E-15	0.00E+00	6.93E-06	6.86E-14	1.35E-07	1.76E-08	0.00E+00
	Best	**1.02E+01**	3.68E+01	8.45E+00	7.86E+03	4.98E+00	4.34E+03	2.40E+01	3.62E+01
F4	Ave	1.62E+01	5.88E+01	1.55E+01	8.82E+03	1.02E+01	1.16E+04	4.34E+01	4.80E+01
	Std	5.29E+00	2.41E+01	6.60E+00	1.35E+03	5.38E+00	1.03E+04	2.64E+01	1.66E+01
	Best	**1.00E+00**	1.13E+00	**1.00E+00**	3.73E+00	1.06E+00	3.31E+00	1.08E+00	1.02E+00
F5	Ave	1.02E+00	1.13E+00	1.06E+00	3.83E+00	1.08E+00	3.37E+00	1.13E+00	1.02E+00
	Std	9.71E-03	6.25E-03	5.56E-02	1.41E-01	1.87E-02	8.57E-02	4.13E-02	3.45E-03
	Best	3.94E+00	3.75E+00	9.06E+00	9.34E+00	8.99E+00	8.74E+00	**3.63E+00**	5.07E+00
F6	Ave	4.42E+00	4.63E+00	9.58E+00	9.82E+00	1.00E+01	8.88E+00	4.27E+00	5.60E+00
	Std	3.80E-01	9.61E-01	5.75E-01	6.81E-01	6.77E-01	1.87E-01	6.28E-01	7.50E-01
	Best	3.01E+00	3.44E+00	6.44E+00	5.51E+00	**2.89E+00**	5.21E+00	5.04E+00	3.05E+00
F8	Ave	3.90E+00	4.35E+00	6.48E+00	6.05E+00	3.54E+00	5.52E+00	5.52E+00	3.15E+00
	Std	5.27E-01	9.44E-01	3.58E-02	7.61E-01	4.22E-01	4.41E-01	4.83E-01	9.32E-01
	Best	**2.31E+00**	2.35E+00	2.34E+00	1.48E+03	2.34E+00	9.17E+02	2.47E+00	2.34E+00
F9	Ave	2.42E+00	2.36E+00	2.34E+00	1.94E+03	2.34E+00	9.97E+02	2.63E+00	2.36E+00
	Std	9.39E-02	1.25E-02	2.08E-03	6.55E+02	2.84E-03	1.14E+02	1.38E-01	1.96E-02
	Best	1.86E+00	**2.78E-13**	2.03E+01	2.02E+01	2.03E+01	2.03E+01	2.00E+01	2.00E+01
F10	Ave	1.64E+01	7.05E+00	2.03E+01	2.04E+01	2.04E+01	2.03E+01	2.00E+01	2.00E+01
	Std	8.12E+00	1.12E+01	2.97E-02	1.81E-01	5.38E-02	8.96E-02	1.76E-03	8.07E-03

To assess the convergence speed of SDP and other algorithms, they were executed on the test functions of F2 and F3 (including default attributes), as well as F4, F5, F8, and F9 (including a few movement and rotation attributes), up to the number of fitness function evaluations equal to 10E+5. The convergence speed of all algorithms in these test functions is depicted in Fig. 10. In all of these functions, the fastest convergence speed is exhibited by SDP.

F. Experiment 3: CEC 2022 Test Functions

In this experiment, the performance of SDP was compared with that of AHA, DMO, RSA, SO, PDO, FLA, and GOA in terms of both exploration and exploitation capabilities on the designed test functions of CEC 2022. The suite used includes 12 functions categorized into different types: *one unimodal* (F1), *four basic* (F2-F5), *three hybrids* (F6-F8), and *four composition functions* (F9-F12). As detailed in Table 8, each function is characterized by similar dimensions, uniformly set at 10, and a consistent range of [-100, 100]. The comparative results for all algorithms on the CEC 2022 test functions are presented in Table 9.

The evaluation of the performance of SDP and others in addressing these test functions involves the utilization of the Friedman test. The results of this test, along with the hit rate from Table 7, are presented in Fig. 11. In the first and second diagrams of this figure, confirmation is provided that the first rank in discovering the closest solutions to optima is achieved by SDP. It is thereby inferred that a powerful balance between exploitation

and exploration abilities is maintained by SDP when compared to other algorithms. In essence, SDP is capable of freely exploring different parts of the search space, if necessary, and enhancing the quality of current promising solutions through local searching around them. Consequently, SDP can effectively avoid local optima.

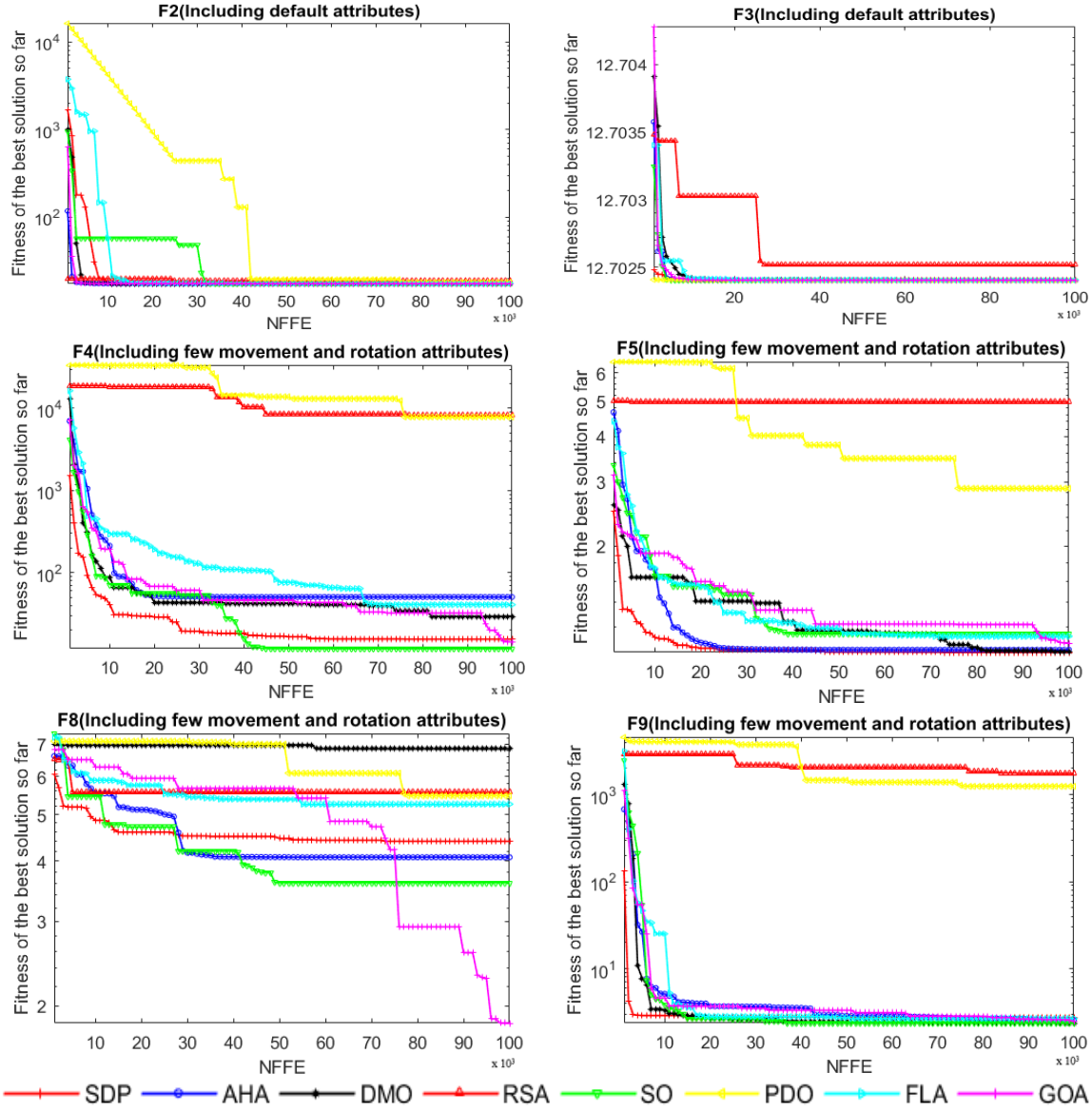


Fig. 10: Convergence curve of all algorithms on solving the CEC 2019 test functions.

Moreover, the third diagram in Fig. 11 illustrates that, following GOA, which boasts the highest hit rate (0.5), SDP, together with AHA and SO, attains the second-highest hit rate (0.41) among the algorithms under consideration. Additionally, beyond the Friedman test, the Wilcoxon signed-rank test results in Fig. 12 indicate that the values of R- surpass R+ in all pair comparisons. It is thereby concluded that SDP exhibits superior effectiveness compared to the majority of other algorithms. Given that the test suite of CEC 2022 comprises four distinct types of test functions, the evaluation and comparison of algorithm convergence speed involve the selection of specific functions of varying types.

The chosen functions include unimodal (F1), basic (F2 and F3), hybrid (F6, F7, and F8), and composition (F9 and F10). The convergence speed of all algorithms in these selected test functions is depicted in Fig. 13. In the majority of these functions, the fastest convergence speed is observed with SDP.

G. Experiment 4: Five Constrained Engineering Design Problems

In this experiment, five constrained engineering design problems are utilized to evaluate the performance of SDP, and the results are compared with various algorithms presented in the literature. Further details regarding these problems, including

mathematical formulas, can be found in [3]. To assess the effectiveness of SDP in optimizing the specified constrained problems, the constraint-handling mechanism introduced in [51] is applied to SDP. In this mechanism, also known as the penalty method, a penalty function is defined to transform the constrained problem into an unconstrained problem. The outcomes from each problem are presented in Tables 10-14, reflecting 30 independent runs for each.

Speed reducer design: The objective of this problem is to minimize the weight of a speed reducer. Seven decision variables need to be defined to meet eleven constraints. Table 10 presents the outcomes from SDP and various notable algorithms from the literature, including SC [52], PSO-DE [53], DELC [54], DEDS [55], HEAA [56], MDE [57], and ABC [58]. It is confirmed by this table that the closest solution to the optimal is jointly achieved by SDP, DELC, and DEDS, each attaining the highest rank.

Pressure vessel design*:* The goal here is to minimize the fabrication cost of a pressure vessel, involving four design variables and four constraints. The results from SDP and several reputable algorithms like GA2 [59], GA3 [60], QPSO [61], and PSO [53] are shown in Table 11. According to the table, the first rank in achieving the closest solution to the optimal is maintained by SDP.

Tension/compression spring design*:* This problem focuses on designing a tension/compression spring with the aim of minimizing its weight, involving four design variables and constraints concerning minimum deflection, shear stress, and surge frequency. Table 12 showcases results from SDP and other recognized algorithms in the literature such as GA2, GA3, CAEP [62], CSPSO [63], HPSO [64], DE [65], SC, and ABC. Based on this table, the closest solution to the optimum is found by SDP, which is ranked first.

Multiple disc clutch brake (MDCB): The optimization of a multiple disc clutch brake, aimed at minimizing its overall mass, is the goal of this problem. Five key design variables are manipulated in this optimization process: the inner radius, outer radius, disc thickness, actuating force, and the number of friction surfaces. The results are displayed in Table 13, where the efficacy of SDP in comparison with other notable algorithms such as JAYA [66], TLBO [40], ABC [40], MVO [67], and CMVO [68] is demonstrated. It is confirmed by this table that the closest solution to the optimum is obtained by SDP, which ranks first.

Welded beam design (WBD): The development of an economical welded beam design through cost minimization is the objective of this problem, which involves four critical variables: weld thickness, length of the welded segment, beam height, and beam width. The outcomes of various algorithms, including SDP, TEO [69], SCA [52], CDE [70], HAS-GA [71], and CAEP [62], are presented in Table 14. It is shown by this table that SDP achieves the top rank, confirming its effectiveness in securing a solution that closely aligns with the optimum.

Table 8: The details of CEC 2022 test functions (For all test functions: D = 10 and Range = [-100, 100])

Function	Name	Optimum value	Type
F1	Shifted and full Rotated Zakharov Function	300	Unimodal
F2	Shifted and full Rotated Rosenbrock's Function	400	
F3	Shifted and full Rotated Expanded Schaffer's *f*6 Function	600	Basic
F4	Shifted and full Rotated Non-Continuous Rastrigin's Function	800	
F5	Shifted and full Rotated Levy Function	900	
F6	HF 1 (*N* = 3)	1800	
F7	HF 2 (N = 6)	2000	Hybrid
F8	HF 3 (*N* = 5)	2200	
F9	CF 1 (N = 5)	2300	
F10	CF 2 (*N* = 4)	2400	Composition
F11	CF 3 (N = 5)	2600	
F12	CF 4 (N = 6)	2700	

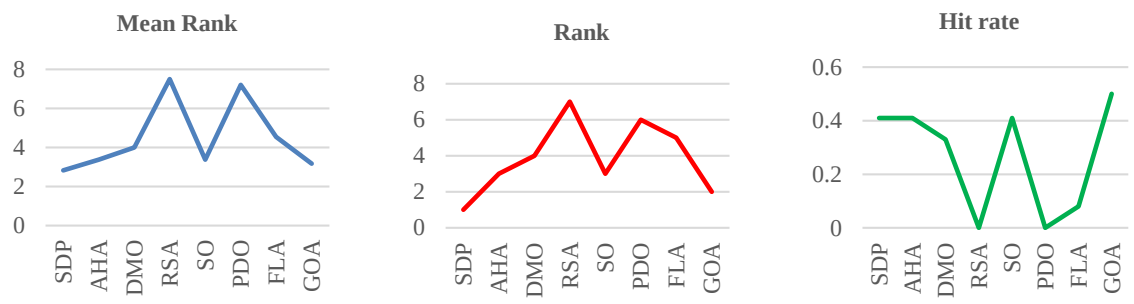


Fig. 11: Results of the Friedman's rank test along with hit rate for CEC 2022 test functions.

Table 9: Comparative results for CEC 2022 test functions

Function	Metric	SDP	AHA	DMO	RSA	SO	PDO	FLA	GOA
	Best	2.01E+03	**3.00E+02**	**3.00E+02**	5.73E+03	**3.00E+02**	1.16E+04	3.03E+02	**3.00E+02**
F1	Ave	5.58E+03	3.00E+02	3.00E+02	7.13E+03	3.00E+02	1.53E+04	3.09E+02	3.00E+02
	Std	2.91E+03	8.04E-14	0.00E+00	1.97E+03	2.84E-14	5.12E+03	7.73E+00	0.00E+00
	Best	**4.00E+02**	**4.00E+02**	4.01E+02	5.32E+02	**4.00E+02**	7.75E+02	4.07E+02	**4.00E+02**
F2	Ave	4.00E+02	4.02E+02	4.02E+02	5.36E+02	4.02E+02	8.31E+02	4.08E+02	4.00E+02
	Std	2.70E-01	2.53E+00	1.72E+00	4.70E+00	2.55E+00	7.99E+01	9.79E-01	0.00E+00
	Best	**6.00E+02**	**6.00E+02**	**6.00E+02**	6.35E+02	**6.00E+02**	6.40E+02	**6.00E+02**	**6.00E+02**
F3	Ave	6.00E+02	6.00E+02	6.00E+02	6.38E+02	6.00E+02	6.42E+02	6.00E+02	6.00E+02
	Std	8.04E-14	5.81E-02	0.00E+00	4.26E+00	6.00E-01	3.86E+00	3.65E-02	1.51E-08
	Best	8.13E+02	8.18E+02	8.21E+02	8.39E+02	8.13E+02	8.34E+02	8.19E+02	**8.02E+02**
F4	Ave	8.18E+02	8.23E+02	8.25E+02	8.40E+02	8.14E+02	8.36E+02	8.33E+02	8.05E+02
	Std	6.62E+00	5.01E+00	3.20E+00	7.46E-01	1.30E+00	1.91E+00	1.28E+01	4.22E+00
	Best	9.37E+02	9.01E+02	**9.00E+02**	1.22E+03	**9.00E+02**	1.23E+03	9.01E+02	**9.00E+02**
F5	Ave	1.03E+03	9.23E+02	9.00E+02	1.26E+03	9.00E+02	1.38E+03	9.54E+02	9.00E+02
	Std	1.10E+02	2.69E+01	0.00E+00	5.42E+01	8.23E-09	2.13E+02	8.56E+01	0.00E+00
	Best	1.83E+03	1.84E+03	6.31E+03	4.43E+07	1.93E+03	1.20E+07	1.98E+03	**1.80E+03**
F6	Ave	1.87E+03	1.87E+03	8.70E+03	4.78E+07	2.90E+03	5.18E+07	3.46E+03	1.80E+03
	Std	3.59E+01	2.77E+01	2.76E+03	4.97E+06	1.05E+03	5.63E+07	1.77E+03	7.43E-02
	Best	**2.00E+03**	**2.00E+03**	2.02E+03	2.10E+03	2.02E+03	2.08E+03	2.02E+03	**2.00E+03**
F7	Ave	2.00E+03	2.01E+03	2.02E+03	2.10E+03	2.03E+03	2.08E+03	2.02E+03	2.00E+03
	Std	3.11E-01	1.36E+01	4.36E-01	7.51E-01	1.45E+01	2.47E+00	8.47E-02	9.44E-03
	Best	**2.21E+03**	2.22E+03	**2.21E+03**	2.24E+03	2.22E+03	2.23E+03	2.22E+03	2.22E+03
F8	Ave	2.22E+03	2.22E+03	2.22E+03	2.25E+03	2.22E+03	2.24E+03	2.22E+03	2.22E+03
	Std	2.67E+00	4.55E-01	6.80E+00	1.13E+01	4.05E-01	4.55E+00	3.19E-01	3.81E-01
	Best	2.53E+03	2.53E+03	**2.49E+03**	2.70E+03	2.53E+03	2.73E+03	2.53E+03	2.63E+03
F9	Ave	2.54E+03	2.53E+03	2.49E+03	2.70E+03	2.53E+03	2.76E+03	2.53E+03	2.63E+03
	Std	4.93E+00	5.57E-13	0.00E+00	2.18E+00	2.27E-13	3.58E+01	1.18E-03	0.00E+00
	Best	**2.40E+03**	2.50E+03	2.50E+03	2.69E+03	2.50E+03	2.52E+03	2.61E+03	2.50E+03
F10	Ave	2.42E+03	2.54E+03	2.51E+03	2.70E+03	2.61E+03	2.52E+03	2.62E+03	2.50E+03
	Std	4.39E+01	6.17E+01	1.75E+01	3.86E+00	9.97E+01	1.59E+00	4.05E+00	4.29E-02
	Best	**2.60E+03**	**2.60E+03**	**2.60E+03**	2.90E+03	**2.60E+03**	2.84E+03	2.60E+03	**2.60E+03**
F11	Ave	2.60E+03	2.60E+03	2.63E+03	3.40E+03	2.72E+03	3.03E+03	2.78E+03	2.60E+03
	Std	8.51E-13	2.52E-11	6.73E+01	7.13E+02	1.64E+02	2.77E+02	2.02E+02	0.00E+00
	Best	**2.87E+03**	**2.87E+03**	2.90E+03	2.89E+03	**2.87E+03**	2.88E+03	2.86E+03	2.88E+03
F12	Ave	2.87E+03	2.87E+03	2.90E+03	2.89E+03	2.87E+03	2.88E+03	2.86E+03	2.88E+03
	Std	1.64E+00	5.68E+00	1.48E-04	5.38E+00	2.70E+00	1.75E-01	2.44E+00	2.79E+00

Fig. 12: Results of the Wilcoxon signed-rank test for CEC 2022 test functions.

H. Discussion

The intense exploitation and exploration capabilities of SDP are evident from the aforementioned four experiments. Its performance in solving unimodal/multimodal functions affirms its exploitation/exploration capability. The competitive edge of SDP over alternative meta-heuristic algorithms is demonstrated by the test outcomes.

Superiority is exhibited by SDP, particularly in effectively addressing high-dimensional functions, surpassing most other algorithms in this context. This result signifies a crucial finding, reinforcing the no-free-lunch theory, as SDP does not excel in solving all 40 functions; in some cases, other algorithms prove more effective than the SDP algorithm.

Conclusion

This paper introduces the innovative Society Deciling Process (SDP), a meta-heuristic algorithm inspired by the societal practice of categorizing individuals into deciles based on factors like monthly income, occupation, and education. The evaluation of SDP against seven established meta-heuristic algorithms (AHA, DMO, RSA, SO, PDO, FLA, and GOA) using three sets of test functions reveals its superior performance in terms of convergence speed and hit rate. Friedman's rank analysis consistently positions SDP with the best mean rank, indicating its proximity to optimal solutions and heightened exploration and exploitation capabilities compared to other algorithms.

Moreover, the potential expansion of the SDP algorithm to include other quantiles, such as quartiles, is discussed. This expansion introduces flexibility in population division and repositioning, with implications for granularity, sensitivity, computational efficiency, exploration and exploitation balance, adaptability to problem characteristics, convergence speed, and algorithm robustness. The choice of quantiles becomes crucial and should align with the specific properties of the optimization problem at hand.

In summary, the inclusion of other quantiles enhances the adaptability of the SDP approach to different problem characteristics and optimization requirements. The careful consideration of quantile choice is essential for optimizing the algorithm's performance. Additionally, recognizing SDP as a single-objective algorithm prompts future considerations for developing a multi-objective version or exploring its binary iteration. Combining SDP with other optimization techniques also offers promising avenues for future research.

To assess the overall performance of the presented SDP algorithm, a statistical analysis of the experiment results is conducted using the Friedman test. The results of the Friedman test, as depicted in Fig. 14, include the hit rate for all reported outcomes. Due to the test result (i.e., Fig. 14), the least mean rank (i.e., first rank) for finding the closest solutions to an optimum is attributed to SDP. Furthermore, the highest hit rate (0.40) among the considered algorithms is confirmed by the third diagram in this figure, signifying that SDP achieves an optimum solution in 40% of the considered test functions.

Table 10: Results of different methods on the speed reducer problem

Method	Best	Worst	Ave	Std	Rank (Ave)
SDP	2994.471066	2994.471066	2994.471066	0.000000	**1**
SC	2994.744241	3009.964736	3001.758264	4.0	**6**
PSO-DE	2996.348167	2996.348204	2996.348174	6.4E-06	**3**
DELC	2994.471066	2994.471066	2994.471066	1.9E-12	**1**
DEDS	2994.471066	2994.471066	2994.471066	3.6E-12	**1**
HEAA	2994.499107	2994.752311	2994.613368	7.0E-02	**2**
MDE	2996.356689	NA	2996.367220	8.2E-03	**4**
ABC	2997.058000	NA	2997.058000	0.0	**5**

Table 11: Results of different methods on the pressure vessel problem

Method	Best	Worst	Ave	Std	Rank (Ave)
SDP	6047.5862	6262.3625	6156.4525	101.4904	**1**
GA2	6288.7445	6308.4970	6293.8432	7.4133	**3**
GA3	6059.9463	6469.3220	6177.2533	130.9297	**2**
QPSO	6059.7209	8017.2816	6440.3786	6059.7209	**4**
PSO	6693.7212	14076.3240	8756.6803	1492.5670	**5**

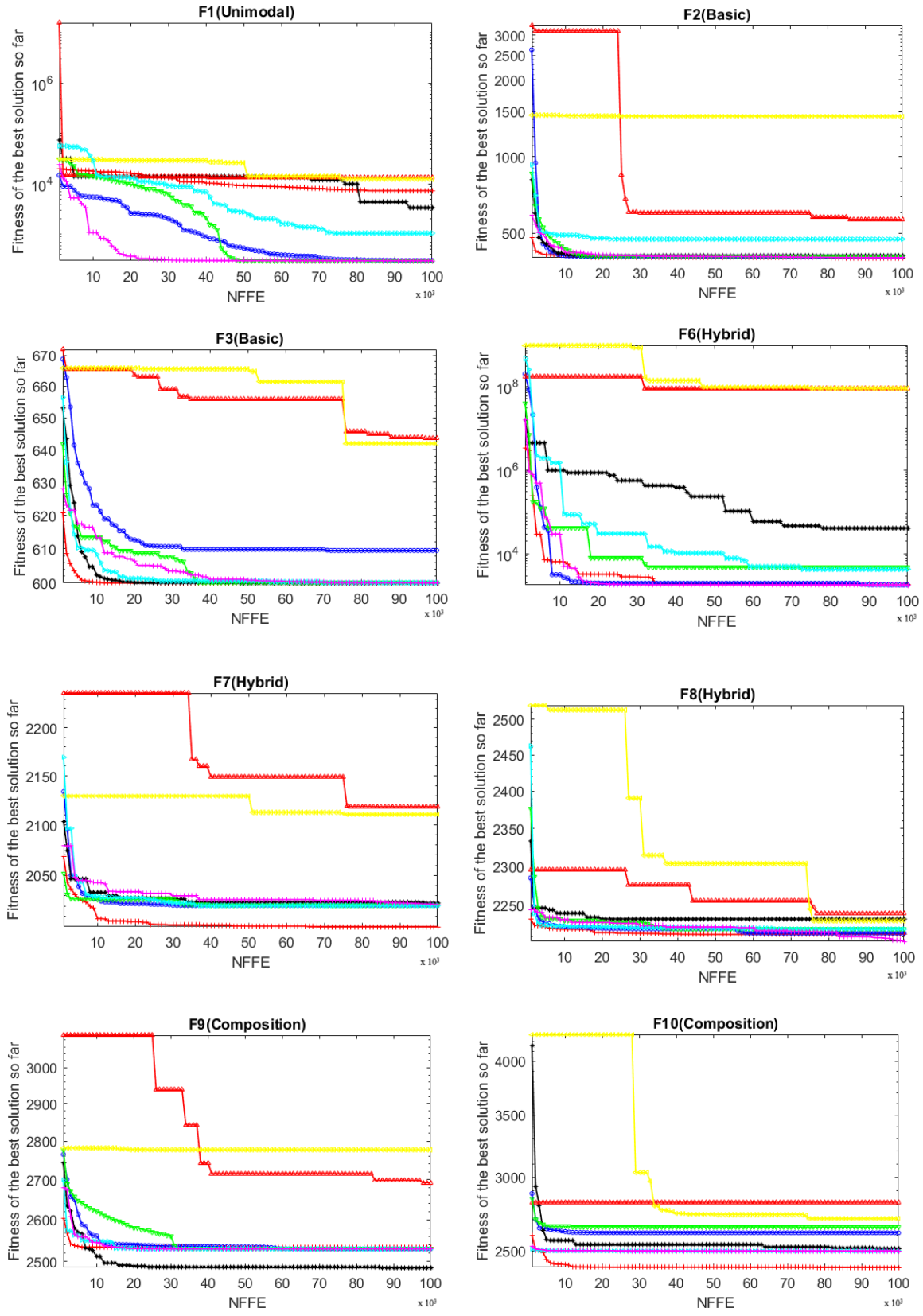


Fig. 13: Convergence curve of all algorithms on solving the CEC 2022 test functions.

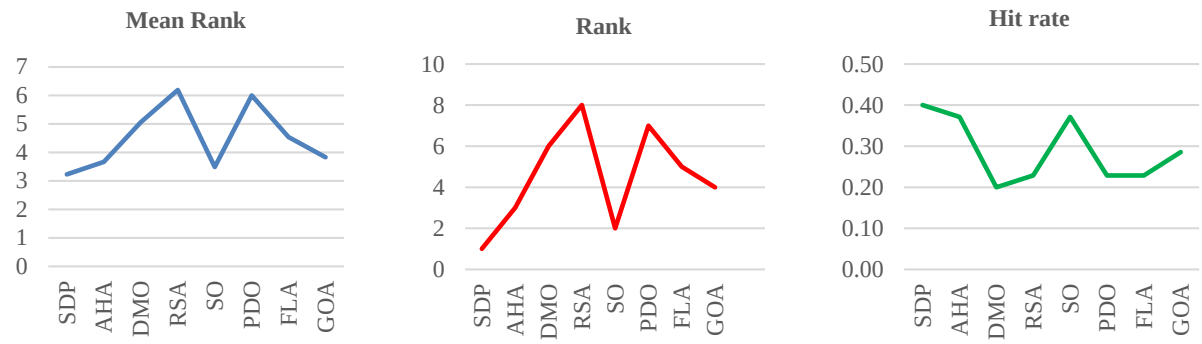


Fig. 14: Results of the Friedman's rank test along with hit rate for all test functions.

Table 12: Comparative analysis of various approaches for tension/compression spring design

Method	Best	Worst	Ave	Std	Rank (Ave)
SDP	0.012668	0.012712	0.012680	0.000018	**1**
GA2	0.012704	0.012822	0.012769	3.94E-05	**7**
GA3	0.012681	0.012973	0.012742	5.90E-05	**6**
CAEP	0.012721	0.015116	0.013568	8.42E-04	**9**
CPSO	0.012674	0.012924	0.012730	5.20E-04	**5**
HPSO	0.012665	0.012719	0.012707	1.58E-05	**3**
DE	0.012670	0.012790	0.012703	2.7E-05	**2**
SC	0.012669	0.016717	0.012922	1.2E-08	**8**
ABC	0.012665	NA	0.012709	0.012813	**4**

Table 13 Comparative analysis of various approaches to addressing the multiple disc clutch brake issue

Algorithm	Best	Worst	Ave	Std	Rank(Ave)
SDP	0.313657	0.42573	0.321232	0.421	**1**
JAYA	0.313657	0.3867384	0.324425	0.69	**3**
TLBO	0.313657	0.392071	0.3271662	0.67	**5**
ABC	0.313657	0.324751	0.324751	0.54	**4**
MVO	0.313656	0.40120	0.351241	2.31E-05	**6**
CMVO	0.313656	0.32316	0.320104	3.22E−07	**2**

Table 14 Comparative analysis of various approaches for solving the welded beam design challenge

Algorithm	Best	Worst	Ave	Std	Rank(Ave)
SDP	1.5621	1.8394	1.7332	0.2671	**1**
TEO	1.7252	1.9311	1.7680	0.0581	**3**
SCA	2.3854	6.3996	3.2551	0.9590	**6**
CDE	1.7335	1.8241	1.7681	0.0221	**4**
HAS-GA	2.2500	2.2800	2.2600	0.0078	**5**
CAEP	1.7248	3.1797	1.9718	0.4430	**2**

Author Contributions

Both authors of the paper, namely E. Pira and A. Rouhi, developed the metaheuristic algorithm, executed the experiments, analyzed the results, and authored the manuscript.

Acknowledgment

We would like to express our sincere gratitude to the editor and anonymous reviewers for their valuable time, insightful feedback, and constructive comments, which greatly contributed to the improvement and quality of

this paper.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

EA	Evolutionary Algorithm
GA	Genetic Algorithm
GP	Genetic Programming
NP	Natural Phenomenon
SA	Simulated Annealing
EVO	Energy Valley Optimizer
WCA	Water Cycle Algorithm
NOA	Nutcracker optimizer
OA	Orchard Algorithm
SMO	Swarm Magnetic Optimizer
FuFiO	Fusion–fission optimization
FLA	Fick's Law Optimization
SBs	Social Behaviors
SCO	Single Candidate Optimizer
GMO	Geometric Mean Optimizer
HO	Hippopotamus Optimization
GGO	Greylag Goose Optimization
LBO	Ladybug Beetle Optimization
PDO	Prairie Dog Optimization
AO	Aquila Optimizer
RFO	Red Fox Optimization
HBA	Honey Badger Algorithm
RSA	Reptile Search Algorithm
IA	Ideology Algorithm
ICO	Intelligent Clonal Optimizer
EXA	Expectation Algorithm
SO	Seasons Optimization
NFL	No Free Lunch
SDP	Society Deciling Process
GOA	Gazelle Optimization Algorithm
AHA	Artificial Hummingbird Algorithm
DBO	Dung Beetle Optimizer
FOX	Fox optimizer
GTO	Giant Trevally Optimizer
MGO	Mountain Gazelle Optimizer
DMO	Dwarf Mongoose Optimization
ALO	Attack-Leave Optimizer
PIDC	Physics-Inspired Discriminative Classifier
BSLO	Blood-Sucking Leech Optimizer
GKSO	Genghis Khan shark Optimizar
COA	Crayfish Optimization Algorithm
MaxIter	Maximum iteration number
wAvg	Weighted average
MNFFE	Maximum number of fitness function evaluations
NFFE	The number of fitness function calls
Ave	Average
m	The dimension of test functions

References

[1] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk, W. Al-Atabany, "Honey badger algorithm: New metaheuristic algorithm for solving optimization problems," Math. Comput. Simul., 192: 84-110, 2022.

[2] E. Pira, "City councils evolution: a socio-inspired metaheuristic optimization algorithm," J. Ambient Intell. Hum. Comput., 14(9): 12207-12256, 2023.

[3] W. Zhao, L. Wang, S. Mirjalili, "Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications," Comput. Methods Appl. Mech. Eng., 388: 114194, 2022.

[4] M. Abdel-Basset, L. Abdel-Fatah, A. K. Sangaiah, "Metaheuristic algorithms: A comprehensive review," Comput. Intell. multimedia Big Data Cloud Eng. Appl., 2018: 185-231, 2018.

[5] I. Boussaïd, J. Lepagnot, P. Siarry, "A survey on optimization metaheuristics," Inf. Sci., 237: 82-117, 2013.

[6] J. H. Holland, "Genetic algorithms," Sci. Am., 267(1): 66-73, 1992.

[7] J. R. Koza, "Evolution of subsumption using genetic programming," in Proc. the First European Conference on Artificial Life: 110-119, 1992.

[8] T. M. Shami, D. Grace, A. Burr, P. D. Mitchell, "Single candidate optimizer: a novel optimization algorithm," Evol. Intell., 17(2): 863-887, 2024.

[9] P. D. Kusuma, F. C. Hasibuan, "Attack-Leave optimizer: A new metaheuristic that focuses on the guided search and performs random search as alternative," Int. J. Intell. Eng. Syst., 16(3), 2023.

[10] S. Kirkpatrick, C. D. Gelatt Jr, M. P. Vecchi, "Optimization by simulated annealing," science, 220(4598): 671-680, 1983.

[11] M. Azizi, U. Aickelin, H. A. Khorshidi, M. Baghalzadeh Shishehgarkhaneh, "Energy valley optimizer: a novel metaheuristic algorithm for global and engineering optimization," Sci. Rep., 13(1): 226, 2023.

[12] H. Eskandar, A. Sadollah, A. Bahreininejad, M. Hamdi, "Water cycle algorithm–A novel metaheuristic optimization method for solving constrained engineering optimization problems," Comput. Struct., 110: 151-166, 2012.

[13] M. Abdel-Basset, R. Mohamed, M. Jameel, M. Abouhawwash, "Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems," Knowledge-Based Syst., 262: 110248, 2023.

[14] M. Kaveh, M. S. Mesgari, B. Saeidian, "Orchard Algorithm (OA): A new meta-heuristic algorithm for solving discrete and continuous optimization problems," Math. Comput. Simul., 208: 95-135, 2023.

[15] P. D. Kusuma, F. C. Hasibuan, "Swarm magnetic optimizer: A new optimizer that adopts magnetic behaviour," Int. J. Intell. Eng. Syst., 16(4), 2023.

[16] B. Nouhi, N. Darabi, P. Sareh, H. Bayazidi, F. Darabi, S. Talatahari, "The fusion–fission optimization (FuFiO) algorithm," Sci. Rep., 12(1): 12396, 2022.

[17] F. A. Hashim, R. R. Mostafa, A. G. Hussien, S. Mirjalili, K. M. Sallam, "Fick's law algorithm: A physical law-based algorithm for numerical optimization," Knowledge-Based Syst., 260: 110146, 2023.

[18] F. Rezaei, H. R. Safavi, M. Abd Elaziz, S. Mirjalili, "GMO: geometric mean optimizer for solving engineering problems," Soft Comput., 27(15): 10571-10606, 2023.

[19] M. Monemizadeh, S. R. Samareh Hashemi, M. Sheikh-Hosseini, H. Fehri, "A new physics-inspired discriminative classifier," AUT J. Electr. Eng., 2024.

[20] R. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory," in MHS'95. Proc. the sixth International Symposium on Micro Machine And Human Science: 39-43, 1995.

[21] F. A. Hashim, A. G. Hussien, "Snake optimizer: A novel meta-heuristic optimization algorithm," Knowledge-Based Syst., 242: 108320, 2022.

[22] A. E. Ezugwu, J. O. Agushaka, L. Abualigah, S. Mirjalili, A. H. Gandomi, "Prairie dog optimization algorithm," Neural Comput. Appl., 34(22): 20017-20065, 2022.

[23] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-Qaness, A. H. Gandomi, "Aquila optimizer: a novel meta-heuristic optimization algorithm," Comput. Ind. Eng., 157: 107250, 2021.

[24] D. Połap, M. Woźniak, "Red fox optimization algorithm," Expert Syst. Appl., 166: 114107, 2021.

[25] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, A. H. Gandomi, "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer," Expert Syst. Appl., 191: 116158, 2022.

[26] J. O. Agushaka, A. E. Ezugwu, L. Abualigah, "Gazelle optimization algorithm: A novel nature-inspired metaheuristic optimizer," Neural Comput. Appl., 35: 4099-4131, 2022.

[27] J. Xue, B. Shen, "Dung beetle optimizer: A new meta-heuristic algorithm for global optimization," J. Supercomput., 79(7): 7305-7336, 2023.

[28] H. Mohammed, T. Rashid, "FOX: a FOX-inspired optimization algorithm," Appl. Intell., 53(1): 1030-1050, 2023.

[29] H. T. Sadeeq, A. M. Abdulazeez, "Giant trevally optimizer (GTO): A novel metaheuristic algorithm for global optimization and challenging engineering problems," IEEE Access, 10: 121615-121640, 2022.

[30] B. Abdollahzadeh, F. S. Gharehchopogh, N. Khodadadi, S. Mirjalili, "Mountain gazelle optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems," Adv. Eng. Software, 174: 103282, 2022.

[31] J. O. Agushaka, A. E. Ezugwu, L. Abualigah, "Dwarf mongoose optimization algorithm," Comput. Methods Appl. Mech. Eng., 391: 114570, 2022.

[32] M. H. Amiri, N. Mehrabi Hashjin, M. Montazeri, S. Mirjalili, N. Khodadadi, "Hippopotamus optimization algorithm: A novel nature-inspired optimization algorithm," Sci. Rep., 14(1): 5032, 2024.

[33] J. Bai, H. Nguyen-Xuan, E. Atroshchenko, G. Kosec, L. Wang, M. A. Wahab, "Blood-sucking leech optimizer," Adv. Eng. Software, 195: 103696, 2024.

[34] E. S. M. El-kenawy, N. Khodadadi, S. Mirjalili, A. A. Abdelhamid, M. M. Eid, A. Ibrahim, "Greylag goose optimization: Nature-inspired optimization algorithm," Expert Syst. Appl., 238: 122147, 2024.

[35] G. Hu, Y. Guo, G. Wei, L. Abualigah, "Genghis Khan shark optimizer: a novel nature-inspired algorithm for engineering optimization," Adv. Eng. Inf., 58: 102210, 2023.

[36] S. Safiri, A. Nikoofard, "Ladybug Beetle optimization algorithm: Application for real-world problems," J. Supercomput., 79(3): 3511-3560, 2023.

[37] H. Jia, H. Rao, C. Wen, S. Mirjalili, "Crayfish optimization algorithm," Artif. Intell. Rev., 56(Suppl 2): 1919-1979, 2023.

[38] S. Satapathy, A. Naik, "Social group optimization (SGO): A new population evolutionary optimization technique," Complex Intell. Syst., 2(3): 173-203, 2016.

[39] E. Pira, "City councils evolution: a socio-inspired metaheuristic optimization algorithm," J. Ambient Intell. Hum. Comput., 14: 12207-12256, 2022.

[40] R. V. Rao, V. J. Savsani, D. Vakharia, "Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems," Comput.- Aided Des., 43(3): 303-315, 2011.

[41] A. Borji, "A new global optimization algorithm inspired by parliamentary political competitions," in Proc. Mexican International Conference on Artificial Intelligence: 61-71, 2007.

[42] T. T. Huan, A. J. Kulkarni, J. Kanesan, C. J. Huang, A. Abraham, "Ideology algorithm: A socio-inspired optimization methodology," Neural Comput. Appl., 28(1): 845-876, 2017.

[43] V. Sahargahi, V. Majidnezhad, S. T. Afshord, Y. Jafari, "An intelligent chaotic clonal optimizer," Appl. Soft Comput., 115: 108126, 2022.

[44] A. S. Shastri, A. Jagetia, A. Sehgal, M. Patel, A. J. Kulkarni, "Expectation algorithm (ExA): a socio-inspired optimization methodology," in Socio-cultural Inspired Metaheuristics: Springer, pp. 193-214, 2019.

[45] H. Emami, "Seasons optimization algorithm," Eng. Comput., 38: 1845-1865, 2020.

[46] D. H. Wolpert, W. G. Macready, "No free lunch theorems for optimization," IEEE Trans. Evol. Comput., 1(1): 67-82, 1997.

[47] A. Kaveh, T. Bakhshpoori, "Water evaporation optimization: a novel physically inspired optimization algorithm," Comput. Struct., 167: 69-85, 2016.

[48] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," Ann. Math. Stat., 11(1): 86-92, 1940.

[49] R. F. Woolson, "Wilcoxon signed-rank test," Wiley Encyclopedia of Clinical Trials: 1-3, 2007.

[50] G. Hu, J. Zhong, B. Du, G. Wei, "An enhanced hybrid arithmetic optimization algorithm for engineering applications," Comput. Methods Appl. Mech. Eng., 394: 114901, 2022.

[51] A. H. Gandomi, X. S. Yang, A. H. Alavi, S. Talatahari, "Bat algorithm for constrained optimization tasks," Neural Comput. Appl., 22: 1239-1255, 2013.

[52] T. Ray, K. M. Liew, "Society and civilization: An optimization algorithm based on the simulation of social behavior," IEEE Trans. Evol. Comput., 7(4): 386-396, 2003.

[53] H. Liu, Z. Cai, Y. Wang, "Hybridizing particle swarm optimization with differential evolution for constrained numerical and engineering optimization," Appl. Soft Comput., 10(2): 629-640, 2010.

[54] L. Wang, L. p. Li, "An effective differential evolution with level comparison for constrained engineering design," Struct. Multidiscip. Optim., 41(6): 947-963, 2010.

[55] M. Zhang, W. Luo, X. Wang, "Differential evolution with dynamic stochastic selection for constrained optimization," Inf. Sci., 178(15): 3043-3074, 2008.

[56] Y. Wang, Z. Cai, Y. Zhou, Z. Fan, "Constrained optimization based on hybrid evolutionary algorithm and adaptive constraint-handling technique," Struct. Multidiscip. Optim., 37(4): 395-413, 2009.

[57] E. Mezura-Montes, C. C. Coello, J. Velázquez-Reyes, "Increasing successful offspring and diversity in differential evolution for engineering design," in Proc. the Seventh international Conference on Adaptive Computing in Design and Manufacture (ACDM 2006): 131-139, 2006.

[58] D. Karaboga, B. Basturk, "Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems," in Proc. International Fuzzy Systems Association World Congress: 789-798, 2007.

[59] C. A. C. Coello, "Use of a self-adaptive penalty approach for engineering optimization problems," Comput. Ind., 41(2): 113-127, 2000.

[60] C. A. C. Coello, E. M. Montes, "Constraint-handling in genetic algorithms through the use of dominance-based tournament selection," Adv. Eng. Inf., 16(3): 193-203, 2002.

[61] L. dos Santos Coelho, "Gaussian quantum-behaved particle swarm optimization approaches for constrained engineering design problems," Expert Syst. Appl., 37(2): 1676-1683, 2010.

[62] C. A. Coello Coello, R. L. Becerra, "Efficient evolutionary optimization through the use of a cultural algorithm," Eng. Optim., 36(2): 219-236, 2004.

[63] Q. He, L. Wang, "An effective co-evolutionary particle swarm optimization for constrained engineering design problems," Eng. Appl. Artif. Intell., 20(1): 89-99, 2007.

[64] Q. He, L. Wang, "A hybrid particle swarm optimization with a feasibility-based rule for constrained optimization," Appl. Math. Comput., 186(2): 1407-1422, 2007.

[65] J. Lampinen, "A constraint handling approach for the differential evolution algorithm," in Proc. the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600), 2: 1468-1473, 2002.

[66] R. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," Int. J. Ind. Eng. Comput., 7(1): 19-34, 2016.

[67] S. Mirjalili, S. M. Mirjalili, A. Hatamlou, "Multi-verse optimizer: A nature-inspired algorithm for global optimization," Neural Comput. Appl., 27: 495-513, 2016.

[68] G. I. Sayed, A. Darwish, A. E. Hassanien, "A new chaotic multi-verse optimization algorithm for solving engineering optimization problems," J. Exp. Theor. Artif. Intell., 30(2): 293-317, 2018.

[69] A. Kaveh, A. Dadras, "A novel meta-heuristic optimization algorithm: thermal exchange optimization," Adv. Eng. Software, 110: 69-84, 2017.

[70] F. Z. Huang, L. Wang, Q. He, "An effective co-evolutionary differential evolution for constrained optimization," Appl. Math. Comput., 186(1): 340-356, 2007.

[71] S. Korkmaz, N. B. H. Ali, I. F. Smith, "Configuration of control system for damage tolerance of a tensegrity bridge," Adv. Eng. Inf., 26(1): 145-155, 2012.

Biographies

Einollah Pira received his B.Sc. degree in Computer Engineering (Software) from the University of Kharazmi, Tehran, Iran [1996–2000], the M.Sc. degree in Computer Engineering (Software) from the Sharif University of Technology, Tehran, Iran [2000–2002], and Ph.D. degree in Computer Engineering (Software) from Arak University, Iran [2013-2017]. Currently, he is an Associate Professor with Department of Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran. His research interests include model checking, formal methods, software testing, evolutionary computation, and machine learning.

- Email: pira@azaruniv.ac.ir
- ORCID: 0000-0001-9010-6113
- Web of Science Researcher ID: NA
- Scopus Author ID 55941352000
- Homepage: http://pajouhesh.azaruniv.ac.ir/_Pages/ResearcherEn.aspx?ID=6617

Alireza Rouhi received his B.Sc. at Kharazmi University of Tehran in September, 2000; M.Sc. at Sharif University of Technology in June, 2004; and Ph.D. at University of Isfahan in September 2017, all in Software Engineering field. He rewarded as outstanding researcher of Ph.D. students at Faculty of Computer Engineering, University of Isfahan in 2017. Currently, he is an Associate Professor at Azarbaijan Shahid Madani University, Tabriz, Iran. He is interested in Software Engineering in general and Formal Specification, Model Transformation, Metaheuristics, and Social Networks in particular.

- Email: rouhi@azaruniv.ac.ir
- ORCID: 0000-0003-1494-3467
- Web of Science Researcher ID: L-2209-2018
- Scopus Author ID 57189992181
- Homepage: http://pajouhesh.azaruniv.ac.ir/_Pages/ResearcherEn.aspx?ID=5384

How to cite this paper:
E. Pira, A. Rouhi, "Society deciling process: A socio-inspired meta-heuristic algorithm," J. Electr. Comput. Eng. Innovations, 12(2): 535-556, 2024.

DOI: 10.22061/jecei.2024.10831.740

URL: https://jecei.sru.ac.ir/article_2152.html

Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

A Probabilistic Framework for Active Distribution Network Optimal Energy Management Considering Correlated Uncertain Variables

S. Abbasi, D. Nazarpour, S. Golshannavaz *

Electrical Engineering Department, Faculty of Electrical and Computer Engineering, Urmia University, Urmia, Iran.

Article Info

Article History:
Received 16 April 2024
Reviewed 05 June 2024
Revised 07 July 2024
Accepted 24 July 2024

Keywords:
Energy management
Renewable energy resources
Expected energy not supplied
Uncertainty
Correlation

*Corresponding Author's Email Address:
s.golshannavaz@urmia.ac.ir

Abstract

Background and Objectives: Distributed generations (DGs) based on renewable energy, such as PV units, are becoming more prevalent in distribution networks due to technical and environmental benefits. However, the intermittency and uncertainty of these sources lead to technical and operational challenges. Energy storage application, uncertainty analysis, and network reconfiguration are apt therapies to resist these challenges.

Methods: Energy management of modern, smart, and renewable-penetrated distribution networks is tailored here considering the uncertainties correlations. Network operation costs including switching operations, the expected energy not served (EENS) index as the reliability objective, and the node voltage deviation suppression as the technical objective are mathematically modeled. Multi-objective particle swarm optimization (MOPSO) is considered as the optimization engine. Scenario generation method and Nataf transformation are used in probabilistic evaluations of the problem. Moreover, the technique for Order Preference by Similarity to the Ideal Solution (TOPSIS) is deployed to make a final balance between different objectives to yield a unified solution.

Results: To show the effectiveness of the proposed approach, the IEEE 33-node distribution network is put under extensive simulations. Different cases are simulated and interrogated to assess the performance of the proposed model.

Conclusion: For different objectives dealing with different aspects of the network, remarkable achievements are attained. In brief, the final solution shows 4.50% decrease in operation cost, 13.07% improvement in reliability index, and 18.85% reduction in voltage deviation compared to the initial conditions.

Introduction

With the reduction of fossil fuel resources and global concerns about environmental pollution caused by these resources, the trend towards distributed generation (DGs) based on renewable energy sources (RESs) has been increased [1]-[3]. The utilization of RESs has significantly contributed to positive environmental, technical and economic benefits. However, it has some challenges too. Due to the uncertainty in the power delivered by RESs to the network, the exploitation of these resources in distribution networks has faced with several challenges [4].

Solar energy based on photovoltaics (PVs), as one of the most common types of RESs, has attracted the attention of distribution network operators. These sources provide their output electrical power by receiving solar radiation. Changes in the intensity of sunlight during the day and the dependency to the generation power of these units cause changes in the generation of power in the distribution network nodes. On the other hand, the mismatch between generation and consumption would cause dumping of available energy. Under these

conditions, the utilization and non-utilization of the available energy of these units lead to challenges such as voltage deviation in the nodes or an increase in costs in case of increased power losses [5].

Meanwhile of getting smart and moldering the networks, one of the available solutions to alleviate these conditions as well as balancing the share of RESs generation and the share of consumed load would be adoption of energy storage systems (ESS) alongside RESs [6]-[8]. For instance, using ESSs together with PV units provides the possibility of obtaining maximum power from these units and a more uniform output profile. With the help of ESSs, the output power of PVs can be stored until the time of need for their consumption. Balancing the power through ESS separates generation and consumption times, effectively. Alongside, by ESS deployment, the stability of the transmission and distribution network and the overall security of the energy system increases as well [9]-[11]. The operation strategy in networks with high penetration of RESs along with ESSs has changed during the past decades to overwhelm the probable problems. For example, in peak load conditions, the operator must reduce energy costs and improve energy efficiency; in other operating states, the reduction of greenhouse gases is considered. On the other hand, the positive effect of RESs and ESSs on the distribution network from a technical point of view depends on several factors and various considerations should be made.

The mentioned cases and sources of uncertainties have turned optimal energy management in distribution networks into a challenge task. Energy management is a set of methods and actions carried out in different systems to use energy correctly and maximize benefits or minimize costs without reducing the quality of services. In other words, energy management is a method to ensure the rational use of energy in a system to improve the efficiency of that system [12]. Efficient and dynamic energy management is essential in active distribution networks to function adequately. This helps to improve operating conditions regarding cost optimization and technical capabilities. It also enhances reliability indices and minimizes voltage deviations at network nodes.

As mentioned, renewable resources are aligned with uncertainties and if the effects of the uncertainties in the problem are not considered, the uncertain parameters would heavily influence the optimization goals. Here, the optimization plans might deviate from the desired goals and lead to the inefficiencies. Besides, various correlated uncertainties should be considered to provide reliable and durable solutions.

Several studies have been conducted on the optimal management of charging and discharging of EES besides RESs [13]-[15]. However, these studies have not modeled network reconfiguration capability. Meanwhile, some studies have used intelligent evolutionary optimization algorithms for optimal management of distribution networks considering reconfiguration. In reference [16], the enhanced gravitational search algorithm (GSA) is proposed to improve transient stability, reduce total operating costs, and reduce losses. In [17], a combined method using particle swarm optimization (PSO) and the Nelder-Mead simplex search algorithm is proposed to implement reconfiguration for reducing active power losses. Also, in [18], genetic algorithm (GA) is proposed for reconfiguring distribution networks by considering a variable population size. In the presented studies, the changes in daily load curve have been ignored and a predetermined period is considered. Therefore, these conditions cannot produce accurate results and provide an optimal solution for daily 24-hours scheduling of distribution networks. In [19], the objective function minimizes operation and reliability costs with Tabu search. ESSs serve multiple objectives including peak shaving, voltage regulation, and reliability enhancement. A method for optimal scheduling of active distribution networks is tailored in [20]. This method is a two-stage process that considers the uncertainty risk associated with RESs, load, electricity price, and system component failure. This study focuses on the optimal dispatching of active distribution networks with ESSs under these uncertain conditions.

Although these studies have addressed uncertainties, possible correlations between them were ignored. Few studies have paid attention to this gap. In [21], probabilistic energy management of an active distribution network is proposed which includes plug-in hybrid electric vehicles and power electronics devices like soft open point and smart transformer with an objective function of the average voltage deviation to be minimized, improving voltage stability, and maximizing daily profits. The correlation between uncertain input variables is modelled by modifying the “Hong’s 2m point estimate method”. Multi-objective DGs planning in distribution networks by considering correlations among uncertainties, i.e., wind speed, light intensity and load demand, is considered in [22]. Here, the objective function minimizes the annual total costs and risks.

The studies mentioned have analyzed the impact of correlation among uncertain input variables on the energy management problem. However, there are still gaps in terms of utilizing the probabilistic evaluation method that is independent of the problem's dimensions and the number of uncertain variables, as well as modelling the correlation between them. Additionally, there is a need to develop effective objective functions and consider the possible technical tools such as reconfiguration of distribution networks. The main

contributions of this paper are summarized as follows:

- Providing a probabilistic framework for optimal reconfiguration of the distribution network, optimal scheduling of ESSs and optimal reactive power setting of the compensator for each period. In this view, the cost of purchasing energy from the upstream network and the switching cost is reduced, reliability is increased, and voltage deviation index (VDI) is suppressed;
- Proposing a scenario generation approach in probabilistic evaluation of the problem;
- Developing a multi-objective PSO (MOPSO) method to optimally schedule of the network and provide a set of optimal solutions based on the Pareto front concept;
- Deploying the technique for Order Preference by Similarity to the Ideal Solution (TOPSIS) for a trade-off between the objectives.

Modelling of Uncertainties and Their Correlations

They are different sources of uncertainties affecting the deterministic model results. Effective solutions should be considered for this issue. In this paper, three uncertain variables say as the load, the intensity of solar irradiance, and the price of electricity supplied by the substation are considered.

A. Uncertainty Modelling

- Active Power Load

For each time period, the samples of active power loads are generated by the Normal distribution function as (1) [23].

$$f(x) = \frac{1}{\sigma \times \sqrt{2\pi}} \times e^{\frac{-(x-\mu)^2}{2\times\sigma^2}} \quad (1)$$

where, σ and μ are the standard deviation and mean values, respectively. It should be mentioned that x denotes the related uncertain variable.

- Solar Radiation

The beta distribution function generates the solar irradiance samples for each time period as (2) [24].

$$f(R) = \frac{\Gamma(\alpha_\beta + \beta_\beta)}{\Gamma(\alpha_\beta)\Gamma(\beta_\beta)} \times R^{\alpha_\beta - 1} \times (1-R)^{\beta_\beta} \quad (2)$$

where, α_β and β_β are the parameters of the beta distribution function. It should be mentioned that the output power of PV units is the function of solar irradiance. Therefore, the related characteristic is considered as (3).

$$P^{PV}(R) = \begin{cases} P_r^{PV} \times \left(\frac{R^2}{R_{STD} R_C}\right) & 0 \le R < R_C \\ P_r^{PV} \times \frac{R}{R_{STD}} & R_C \le R < R_{STD} \\ P_r^{PV} & R_{STD} \le R \end{cases} \quad (3)$$

where, R denotes the solar radiation, R_C denotes the certain radiation point, R_{STD} denotes the solar radiation in the standard conditions, and P_r^{PV} denotes the power output of the PV unit.

- The Purchased Electricity Price

For each time period, the samples of the price of electricity supplied by the substation are generated by the Normal distribution function.

B. Correlation Modelling

The Nataf transformation, or the Nataf correlation transformation, is a mathematical technique to model correlated input variables in engineering and risk analysis. It is beneficial in situations where traditional methods assume that input variables are independent; but in reality, they exhibit correlations [25]. The steps to perform this method are summarized as follows:

- Correlation Matrix (R)

The first step is to characterize the correlations between the input variables. This is done by specifying a correlation matrix, denoted by R, which quantifies the pairwise correlations between the variables.

- Cholesky Decomposition

Once the correlation matrix R is defined, it is decomposed using the Cholesky decomposition method. The Cholesky decomposition factors a positive definite matrix into a product of a lower triangular matrix and its transpose. This decomposition transforms the correlated variables into a set of uncorrelated variables.

- Transformation Function

After obtaining the Cholesky decomposition, a transformation function is applied to the original correlated variables to obtain uncorrelated variables. This transformation involves multiplying the Cholesky factor with the vector of correlated variables.

- Inverse Transformation

Once the analysis is performed on the uncorrelated variables, the results need to be transformed back to the original correlated space. This is achieved by applying the inverse of the transformation function to the results obtained from the uncorrelated variables. The matrix M, as a matrix of uncertain variables of the problem, is represented by (4).

$$M = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad (4)$$

The covariance matrix of the M is equal to (5).

$$C_M = E(MM^T) = R \quad (5)$$

In (5), R is a symmetric matrix and it can be written according to the Cholesky decomposition method to obtain the value of L, where L is the Cholesky factor of a

lower triangular matrix. To construct the correlation vector M, assume that W is a vector with N dimensions of an uncertain variable whose values are independent and whose variance is equal to one. Therefore, its covariance matrix is given by (6).

$$C_W = E(WW^T) = I \quad (6)$$

Using (6), the covariance matrix M can be written by (7) and the correlation matrix M can be obtained.

$$C_M = E(\mathrm{MM}^T) = E(LWW^T L^T) = R \quad (7)$$

Therefore, the correlation between the values of the uncertain vector M is applied.

C. Scenario Selection Method for Probabilistic Evaluation

In this approach, a large number of scenarios are generated to create a precise model of the system. However, developing more scenarios results in a higher computational burden. Therefore, it is essential to select a number of scenarios that reduces the computational burden while maintaining a good approximation of the uncertain parameters. The backward method eliminates duplicate scenarios or scenarios with minimum distance and helps to reduce the number of scenarios [26].

Problem Formulation

In active distribution operation scheduling, there are different time frames to be investigated. The most important one, which is the case of this study too, is the day-ahead operation scheduling of the network. This scheduling platform is usually in hourly basis, as the case of this study. The time frame resolution effects the system on computational burden and the accuracy of the obtained results which are the inputs for the next scheduling smaller time frames, such the hour-ahead and etc. This section discusses decision variables, objective functions, and all equal and unequal constraints of the developed model.

A. Decision Variables

The decision variables in this study include the state of normally open (NO) and normally close (NC) switches, the amount of charge and discharge of ESSs, and the amount of injected reactive power of capacitors at each operation interval. These variables are shown in (8).

$$Z = [Z_{SW}\ \ Z_{ESS}\ \ Z_{Cap}] \quad (8)$$

where, Z_{SW} is the opened switches, Z_{ESS} is the amount of charge and discharge of ESSs, and Z_{Cap} is the amount of reactive power injected by capacitors.

The switches opened in the problem can be expressed according to (9).

$$Z_{SW} = [\bar{Z}_{SW_1}, \bar{Z}_{SW_2}, \dots, \bar{Z}_{SW_{N_{SW}}}] \quad (9)$$

where, Z_{SW_i} denotes the i^{th} opened switch and N_{SW} denotes the number of opened switches. However, Z_{SW_i} is also in accordance by (10).

$$\bar{Z}_{SW_i} = \left[\bar{Z}^1_{SW_i}, \bar{Z}^2_{SW_i}, \dots, \bar{Z}^{N_T}_{SW_i}\right] \quad i\epsilon\{1,2,\dots,N_{SW}\} \quad (10)$$

where, $\bar{Z}^j_{SW_i}$ denotes the i^{th} opened switch in t^{th} time interval and N_T denotes the number of time intervals.

The charge/discharge amount of ESSs in the problem can be expressed by (11).

$$Z_{ESS} = [\bar{Z}_{ESS_1}, \bar{Z}_{ESS_2}, \dots, \bar{Z}_{ESS_{N_{ESS}}}] \quad (11)$$

where, $\bar{Z}_{ESS_i}$ is the charge/discharge amount of ith ESSs and N_{ESS} is the number of ESSs. However, the Z_{ESS_i} is also in accordance by (12).

$$\bar{Z}_{ESS_i} = \left[\bar{Z}^1_{ESS_i}, \bar{Z}^2_{ESS_i}, \dots, Z^{N_T}_{ESS_i}\right] \quad i\epsilon\{1,2,\dots,N_{ESS}\} \quad (12)$$

where, $\bar{Z}^t_{ESS_i}$ denotes the charge/discharge amount of i^{th} ESSs in the t^{th} time period.

The amount of reactive power injected by capacitors in the problem can be expressed according to (13).

$$Z_{Cap} = [\bar{Z}_{Cap_1}, \bar{Z}_{Cap_2}, \dots, \bar{Z}_{Cap_{N_{Cap}}}] \quad (13)$$

where, $\bar{Z}_{Cap_i}$ is the amount of reactive power injected by the i^{th} capacitor and N_{Cap} is the number of capacitors. It should be mentioned that $\bar{Z}_{Cap_i}$ is denoted by (14).

$$\bar{Z}_{Cap_i} = \left[\bar{Z}^1_{Cap_i}, \bar{Z}^2_{Cap_i}, \dots, \bar{Z}^{N_T}_{Cap_i}\right] \quad i\epsilon\{1,2,\dots,N_{Cap}\} \quad (14)$$

where, $\bar{Z}^j_{Cap_i}$ denotes the amount of reactive power injected by the i^{th} capacitor in the t^{th} time period.

B. Objective Functions

This study considers the total energy cost purchased from the upstream network, switching cost, reliability index based on expected energy not served (EENS), and VDI as objective functions.

Each objective function is introduced in the following and their mathematical relations are stated. It should be noted that according to the developed probabilistic framework, the expected value of each objective function is considered and the relationships are presented accordingly.

- Overall Operation Cost

This objective function is given by (15).

$$Cost = \sum_{t=1}^{N_T}\sum_{k=1}^{K}\rho_k \times \left(\sum_{n=1}^{N_{Sub}} C^{ss}_{n,k,t} \times P^{ss}_{n,k,t}\right) + \sum_{t=1}^{N_T}\sum_{i=1}^{N_{SW}} C^{SW} \times |S^t_i - S^t_{0,i}| \quad (15)$$

where, K is the number of scenarios, ρ_k is the probability of k^{th} scenario, N_{Sub} is the number of substations, $C^{ss}_{n,k,t}$ is the price of energy from n^{th} substation, $P^{ss}_{n,k,t}$ is the amount of active power received from n^{th} substation belonging to k^{th} scenario in t^{th} time period, C^{SW} is the switching cost, and S^t_i and $S^t_{0,i}$ are the new and old state of the i^{th} key in t^{th} period, respectively.

- EENS as Reliability Index

To increase the reliability, EENS index is minimized in the developed model. The mentioned index can be expressed according to (16) [27]-[28].

$$EENS = \sum_{t=1}^{N_T}\sum_{k=1}^{K}\rho_k \times \left(\sum_{i=1}^{N_{Nodes}} P_{i,k,t} \times \left(\sum_{l\in H_i} U_l + \sum_{s\in H_i'} U_s'\right)\right) \quad (16)$$

where, N_{Nodes} denotes the number of nodes, $P_{i,k,t}$ denotes the sum of the active power generation and consumption of the i^{th} node belonging to the kth scenario in t^{th} period of the operation period, U_l denotes the amount of unavailability related to the repair time of the l^{th} branch down of the i^{th} node, U_s' denotes the unavailability related to the recovery time s of the branch forward of the i^{th} node, respectively. Also, H_i and H_i' are the sets of branches downstream and upstream of the i^{th} line, respectively. It should be noted that U_l and U_s' are expressed by (17) and (18), respectively.

$$U_l = \beta_l \times t_l \quad (17)$$

$$U_s' = \beta_s \times t_s' \quad (18)$$

where, β_l and β_s are the failure rates related to l^{th} and s^{th} branches, respectively. Also, t_l and t_s' are the average repair time of the l^{th} branch and the average recovery time of the s^{th} branch (distribution line), respectively.

- VDI

This objective function is defined based on difference between the network nodes' voltage magnitude and the distribution substation's voltage magnitude (usually assumed to be equal to 1 per unit) expressed by (19).

$$VDI = \frac{1}{N_T}\sum_{t=1}^{N_T}\sum_{k=1}^{K}\rho_k \times \left(\sum_{i=1}^{N_{Nodes}}(1-|V_i|)^2\right)_{k,t} \quad (19)$$

where, $|V_i|$ denotes the voltage magnitude of i^{th} node.

C. Equal and Unequal Constraints

At each time interval, the constraints related to the proposed optimization problem are considered as follows. In the probabilistic environment, the value of network output variables, such as the magnitude of the node voltage or current in distribution lines, is replaced by the expected value of these variables [29]-[30].

$$P^{ss} + \sum_{i=1}^{N_{PV}} P_{PV_i} + \sum_{i=1}^{N_{ESS}} \pm P_{ESS_i} = \sum_{i=1}^{N_{Load}} P_{L_i} + \sum_{i=1}^{N_{Lines}} P_{Losses_i} \quad (20)$$

$$Q^{ss} + \sum_{i=1}^{N_{PV}} Q_{PV_i} + \sum_{i=1}^{N_{Cap}} Q_{Cap_i} = \sum_{i=1}^{N_{Load}} Q_{L_i} + \sum_{i=1}^{N_{Lines}} Q_{Losses_i} \quad (21)$$

$$E_{i,t} = E_{i,t-1} \pm P_{ESS_i} \times \Delta t \quad (22)$$

$$E_{ESS}^{min} \le E_{ESS_i} \le E_{ESS}^{max} \quad i \in N_{ESS} \quad (23)$$

$$-P_{ESS}^{max} \le P_{ESS_i} \le +P_{ESS}^{max} \quad i \in N_{ESS} \quad (24)$$

$$0 \le Q_{Cap_i} \le Q_{Cap}^{max} \quad i \in N_{Cap} \quad (25)$$

$$|V|^{min} \le \mathrm{E}[|V_i|] \le |V|^{max} \quad i \in N_{Nodes} \quad (26)$$

$$0.9 \le |V_i| \le 1.1 \quad i \in N_{Nodes} \quad (27)$$

$$\mathrm{P}(|V_i| \ge |V|^{max}) \le 0.05 \quad i \in N_{Nodes} \quad (28)$$

$$\mathrm{P}(|V_i| \le |V|^{min}) \le 0.05 \quad i \in N_{Nodes} \quad (29)$$

$$\mathrm{E}[|I_i|] \le |I_{rate_i}| \quad i \in N_{Lines} \quad (30)$$

$$|I_i| \le 1.25 \times |I_{rate_i}| \quad i \in N_{Lines} \quad (31)$$

$$\mathrm{P}(|I_i| \ge |I_{rate_i}|) \le 0.05 \quad i \in N_{Lines} \quad (32)$$

where, P^{ss} and Q^{ss} are the active and reactive power received from the upstream network, respectively, N_{PV} is the number of PV units, P_{PV_i} and Q_{PV_i} are the output active and reactive power of i^{th} PV, respectively, Q_{Cap_i} is the amount of reactive power injected by i^{th} capacitor, P_{ESS_i} is the amount of charging power (with a negative sign) or the amount of discharging power (with a positive sign) corresponding to the i^{th} ESS, N_{Load} is the load number, P_{Losses_i} and Q_{Losses_i} are the active and reactive power losses of the i^{th} line, respectively, Δt is the time frame from operation period, E_{ESS_i} is the amount of energy of the i^{th} ESS, $E_{i,t}$ and $E_{i,t-1}$ are the amount of energy available in the i^{th} ESS in the t^{th} and t-1^{th} time periods, respectively, E_{ESS}^{min} and E_{ESS}^{max} are the lower and upper limits of the amount of energy available in ESSs, respectively. ESS is a device which provides a solution for energy storing and then discharging. In this way, to manage the uncertainties and especially when there are fluctuations in energy generation such as wind and photovoltaics and also energy consumption such as consumer behaviors, ESS provides a place where the energy shortfall could be compensated or extra energy generation could be stored. In this way, the developed operation scheduling model should determine the energy charging/discharging of ESS beside the other decision variables. P_{ESS}^{max} is the maximum active power that can be charged and discharged by ESSs, Q_{Cap}^{max} is the maximum reactive power injected by the i^{th} capacitor, $|V|^{min}$ and $|V|^{max}$ are the lower and the upper limits of the nodes voltage magnitude, respectively. In operation scheduling problems, maintaining a proper voltage quality is of great

importance for a proper service provision for customers. To do so, technical constraints are considered which keep the voltage within the operation standards in the developed model. In this study, this requirement is met by constraints (26)-(29). $P(\ )$ is the probability operator, $E[\ ]$ is the expected value operator, and $\left|I_{rate_i}\right|$ is the maximum rate of i^{th} line.

Simulation Results

A. Case Study

In order to show the effectiveness of the proposed method, the IEEE 33-node test network is used. This network has 33 nodes, 32 branches, and 5 tie lines. The nominal voltage of this network is 12.66 kV [22].

It should be noted that the failure rate of the lines is considered such that the line with the lowest and highest impedance has a failure rate of 0.1 and 0.4 per year, respectively. The failure rate of the rest of the lines is obtained from the interpolation method [23]. Fig. 1 shows the single-line diagram of the IEEE standard 33-node network.

B. Assumptions

The load curve for the study case is shown in Fig. 2. It is reminded that the operation period in this study contains 24 hours by 3-hours-duration step intervals. Fig. 3 also shows the curve of variations in the purchase price of energy from the upstream network for each time period of the study period.

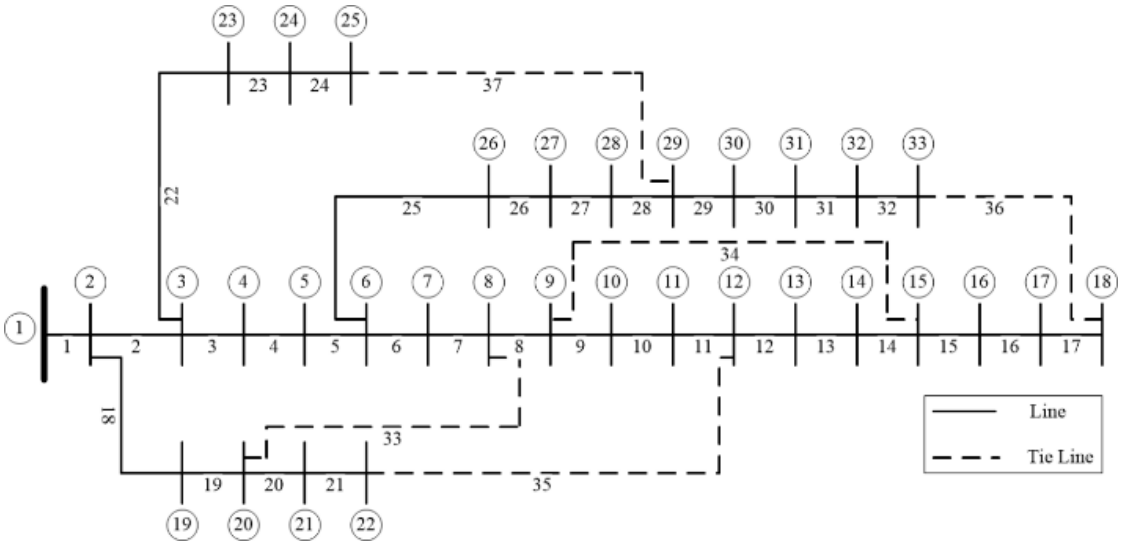


Fig. 1: The IEEE 33-node test network.

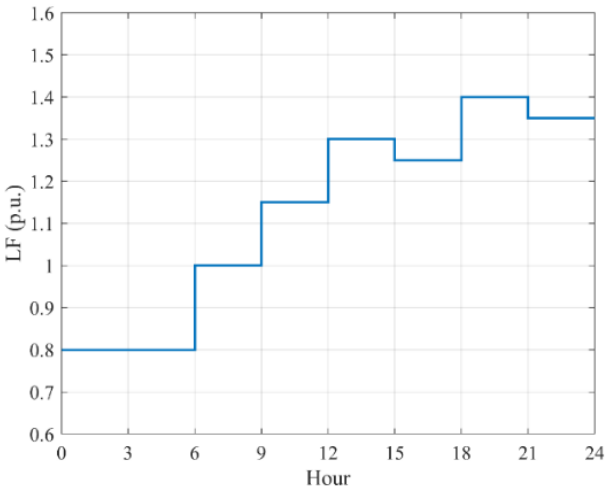


Fig. 2: Load variations curve.

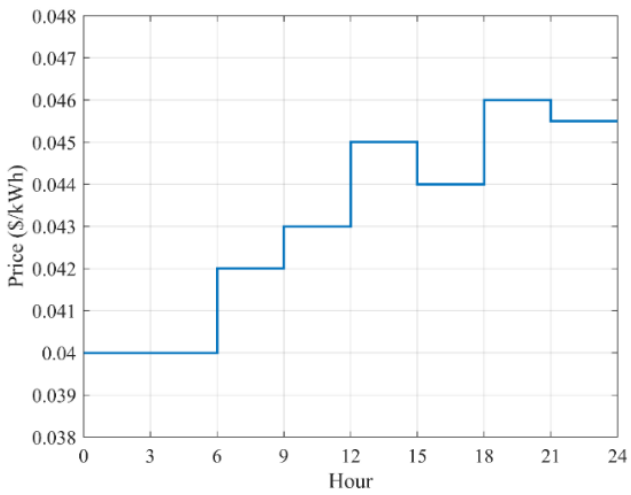


Fig. 3: Electricity price variations.

It should be mentioned that the correlation between uncertain input variables is modeled by Nataf Transformation method. For this purpose, the correlation coefficient between loads in different nodes is assumed to be 0.2, the correlation coefficient between the intensity of sunlight in different nodes is 0.7, and the correlation coefficient between energy consumption and the intensity of sunlight in the corresponding nodes is assumed to be -0.2. The charging and discharging of ESSs is assumed to be between 20% and 80% of their total capacity. For the equal or unequal constraints, the following assumptions are also considered:

$$E_{ESS}^{min} = 60\ kWh$$
$$E_{ESS}^{max} = 240\ kWh$$
$$P_{ESS}^{max} = 30\ kW$$
$$|V|^{min} = 0.95$$
$$|V|^{max} = 1.05$$

Also, it is assumed that three PV units with nominal capacity equal to 500 kW, 300 kW, and 400 kW are installed at nodes 8, 12, 28, respectively. It should be mentioned that the power factor of these units is considered equal to 1.

In Table 1, shape parameters of the beta distribution function in different time periods is presented. It should be noted that these parameters are used for all three PV units. For example, the probability distribution function related to the intensity of sunlight in fourth and sixth periods is presented in Fig. 4.

Table 1: Shape parameters of the beta distribution function

Period	$\alpha_\beta\ (\frac{kW}{m^2})$	$\beta_\beta\ (\frac{kW}{m^2})$
1	0	0
2	0	0
3	2.1440	0.4440
4	2.1440	0.4440
5	1.0820	0.3860
6	1.0820	0.3860
7	0.0714	0.6040
8	0	0

C. Simulation results

Table. 2 presents the statistical information related to the objective functions considered in this study in the initial evaluation conditions. These values are used to show the effectiveness of the proposed solution. It should be noted that Monte Carlo simulation (MCS) scenario generation is used to extract the information in Table 2.

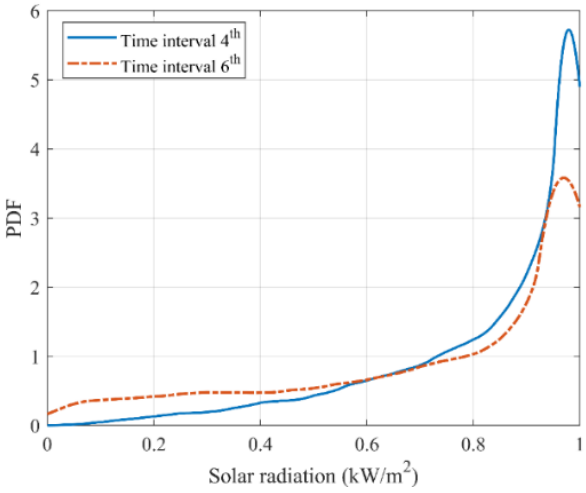


Fig. 4: Probable distribution related to the intensity of sunlight in the 4th and 6th periods.

Table 2: The values of the objective functions in the initial evaluation conditions

Objective functions	Probabilistic methods			
	MCS		Scenario generation	
	E[ ]	σ[ ]	E[ ]	σ[ ]
$Cost\ (\$)$	4143.58	396.81	4146.87	206.89
$EENS\ (\frac{kWh}{year})$	1027768	54088.96	1024829	41245.35
$VDI\ (p.u.)$	0.1278	0.0209	0.1289	0.0127
$Time\ (s)$	1300.12		3.45	

As it is clear from Table 2, scenario generation method has good accuracy in extracting the expected value of the objective functions; this is while, its computational time is much less than the MCS method. It is essential to mention that the standard deviation value error of the scenario generation method is higher than the expected value error of the MCS method. For example, the error of the scenario generation method compared to the MCS method in calculating the expected value and standard deviation of cost is 0.0793% and 47.86%, respectively. Since the expected value of the objective functions is used in this study, the scenario generation performance is very suitable.

The MOPSO method is used to solve the proposed problem and the TOPSIS is contemplated to establish a logical compromise between different objective functions to reach to the final solution. Table 3 shows the value of the objective functions for this set of obtained solutions. As can be seen from this table, the value of the objective functions is obtained for each solution. Among the

obtained solutions, solution number 1 is the best solution from the cost point of view. Under these conditions, the value of this objective function is decreased from \$4146.87 in the initial conditions to \$3718.82. Meanwhile, solution number 15 is the best solution from the viewpoint of EENS index. This solution has achieved 19.48% improvement in this objective function.

Table 3: The value of the objective functions for the set of optimal solutions obtained based on the Pareto front using MOPSO algorithm

Solution No.	$Cost$ (\$)	$EENS$ ($\frac{kWh}{year}$)	VDI ($p.u.$)
1	3718.82	945710	0.1045
2	3774.89	930216	0.1013
3	3793.68	925520	0.1026
4	3865.81	907274	0.1019
5	3919.62	893886	0.1049
6	3960.12	890841	0.1046
7	3962.17	888589	0.1075
8	4046.62	869990	0.1068
9	4073.78	861792	0.1065
10	4128.61	851891	0.1086
11	4155.60	843368	0.1120
12	4219.15	832810	0.1136
13	4222.10	829960	0.1152
14	4247.13	825250	0.1174
15	4250.20	825160	0.1170

Also, solution number 2 is known as the best solution from the VDI point of view. By applying this solution, the value of this objective function decreases from 0.1289 to 0.1013. Fig. 5 also shows the three-dimensional compromise space between different objective functions obtained for the problem per set of optimal solutions based on the Pareto front.

The mentioned solutions are the best from the point of view of each objective function. The situation of an objective function for the best solution from the point of view of another objective function may be even worse than the initial conditions (for example, solution number 15 has worsened the value of cost compared to the first evaluation). In this regard, TOPSIS is used to choose an optimal solution that can make a compromise between all the objective functions and improve all these functions to an appropriate extent. The ranking results of this method for the set of optimal solutions are obtained in Table 3 are presented in Table 4.

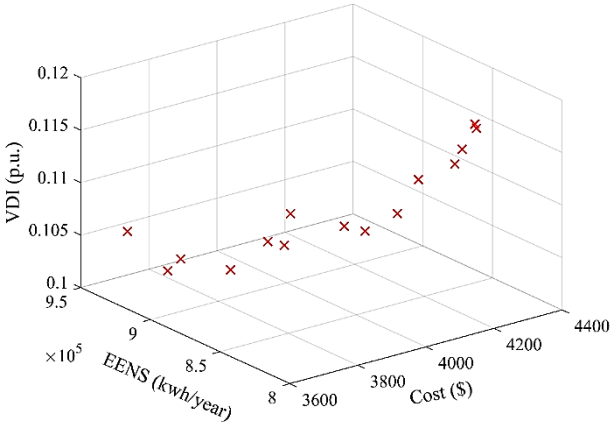


Fig. 5: Three-dimensional compromise space between the objective functions based on the set of optimal solutions obtained from the Pareto front.

Table 4: Decision-making results with TOPSIS

Scenario 1			**Scenario 2**		
$\omega_1 = 0.5, \omega_2 = 0.25, \omega_3 = 0.25$			$\omega_1 = 0.34, \omega_2 = 0.33, \omega_3 = 0.33$		
Rank	**Solution No.**	Cl^+	**Rank**	**Solution No.**	Cl^+
1	2	0.6967	**1**	6	0.6190
2	3	0.6210	**2**	4	0.6000
3	1	0.6786	**3**	5	0.5970

As can be seen from Table 4, it is presented for two decision-making modes. In the first case, the importance of the cost objective function is more significant than that of the other objective functions (the weight coefficient of this objective function is higher). However, the importance of the EENS and VDI index is assumed to be the same. Under these conditions, solution number 2 is chosen as the preferred solution. In the second case, the weight coefficients of all objective functions are considered the same. In other words, importance of all objective functions for the network operator is the same. Under these conditions, solution 6 is selected as the final solution. Table 5 shows this solution's optimal values.

As can be seen from Table 5, each of the decision variables is optimally determined in each period of the network operation. This solution is known as a solution that has made a reasonable compromise between different objective functions. On the other hand, it is resistant to any uncertainty in the network and can be dynamically used by the network operator.

Table 5: optimal values of decision-making variables

Decision variables	Study period							
	1th	2th	3th	4th	5th	6th	7th	8th
Z_{SW_1}	6	7	6	6	5	7	6	7
Z_{SW_2}	11	10	13	14	10	10	14	14
Z_{SW_3}	13	13	21	21	14	13	21	21
Z_{SW_4}	17	26	26	26	25	25	27	26
Z_{SW_5}	26	30	31	30	30	30	31	30
Z_{ESS_1} (kW)	16.56	10.49	-23.61	-5.18	-21.18	38.18	-3.05	14.98
Z_{ESS_2} (kW)	2.94	5.51	-15.60	-9.80	10.05	16.00	-4.60	-4.65
Z_{ESS_3} (kW)	10.18	-19.40	10.12	0.5489	10.18	2.56	-26.9	20.16
Z_{Cap_1} $(kVAr)$	34.30	41.90	39.35	70.14	35.50	56.80	65.10	32.65
Z_{Cap_2} $(kVAr)$	51.94	59.10	53.26	54.48	93.26	50.68	64.30	26.90

Fig. 6 shows the charging and discharging state of storage for each operation period. In this study, the charge state is assumed to be a negative sign and the discharge state is considered a positive sign. This figure shows that storage is often charged during low energy price hours. Meanwhile, as energy prices become more expensive, storage devices are usually in a state of discharge. Table 6 presents statistical information related to cost in different periods, which is very important in risk management and knowing the number of changes in the objective function due to the uncertainties above.

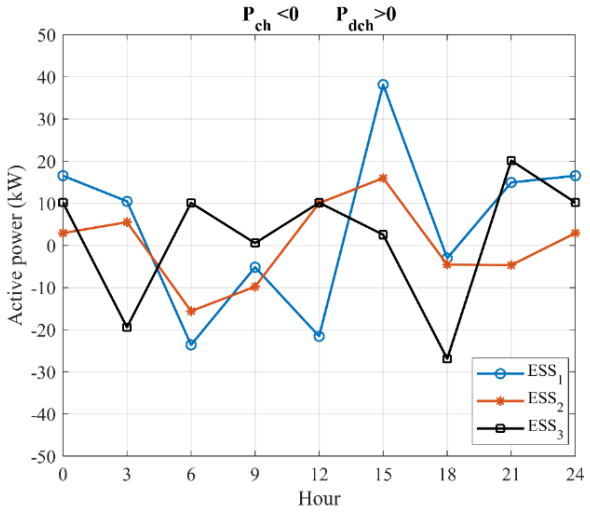


Fig. 6: The charging and discharging state pf storage for each operation period.

Table 6: Operation cost statistical information

Period	$E[\,]$	$\sigma[\,]$
1th	328.91	14.36
2th	335.45	14.61
3th	325.08	18.90
4th	404.68	23.40
5th	510.95	27.90
6th	468.80	25.62
7th	689.59	30.48
8th	658.12	30.01

In general, the results obtained in this section showed that the proposed study method in solving the problem of energy management along with the optimal rearrangement of distribution networks with the goals of reducing the cost of operation, including the cost of purchasing energy from the upstream network and reducing switching costs, improving the reliability index in the form of an index EENS and VDI enhancements are very effective. Considering a wide range of uncertainties and including correlations between uncertain input variables leads to providing more reliable solutions. Meanwhile, providing a set of optimal solutions gives the network operator more flexibility in decision-making.

Conclusion

In this study, the MOPSO method was shown to provide a good performance in solving the energy management problem along with the rearrangement of distribution networks. This method, with the ability to provide a set of solutions based on the Pareto front, gives more decision-making power to the network operator. The solutions provided by this method can be summarized in four modes. The solution chosen as the best solution from the cost of operating point of view reduced the value of this objective function by 10.32%. At the same time, the EENS and VDI indices improved by 7.72% and 18.93%, respectively. The solution obtained as the best solution from the reliability improvement point of view was able to improve the value of this objective function by 19.48%. Meanwhile, network operation cost for this situation has increased by 2.57% and VDI has also improved by 9.23%. The solution chosen as the best solution from the VDI point of view reduced the value of this objective function by 21.41%. Meanwhile, this solution improved the operating cost and EENS indices by 8.97% and 9.23%, respectively. By applying TOPSIS to the set of solutions obtained for the problem in question and considering the same weighting factor for all objective functions, a solution was obtained that can be said to be

a balanced solution for the situation. Because it guarantees the relative improvement of all objective functions, this solution has improved the operating cost value by 4.50%, the EENS index value by 13.07% and the VDI value by 18.85% compared to the initial conditions. The effectiveness of the scenario generation method was proved as one of the probabilistic evaluation methods in this problem. This method has a much higher speed than methods like MCS. Meanwhile, its relative error rate in comparison of statistical moments compared to the MCS method as the reference one is appropriate and acceptable. The correlation between non-deterministic input variables affects the distribution and extraction of non-deterministic samples. Therefore, it can be said that the problems' uncertainties are more severe, and the obtained solutions are closer to reality. Changing the extraction of input samples effectively solves the problem. Therefore, all the obtained solutions based on the Pareto front in this problem are more resistant

Author Contributions

All of the authors carried out the theoretical and simulation results and wrote the manuscript.

Acknowledgment

The authors would like to thank the respected referees for their accurate reviewing of this paper.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

References

[1] F. Sher, O. Curnick, M. T. Azizan, “Sustainable conversion of renewable energy sources,” Sustainability, 13(5): 2940, 2021.

[2] A. Hamidi, D. Nazarpour, S. Golshannavaz, “Multiobjective scheduling of microgrids to harvest higher photovoltaic energy,” IEEE Trans. Ind. Inf., 14(1): 47-57, 2018.

[3] A. Hamidi, S. Golshannavaz, D. Nazarpour, “D-FACTS cooperation in renewable integrated microgrids: A linear multiobjective approach,” IEEE Trans. Sustainable Energy, 10(1): 355-363, 2019.

[4] M. S. Alam, F. S. Al-Ismail, A. Salem, M. A. Abido, “High-level penetration of renewable energy sources into grid utility: Challenges and solutions,” IEEE Access, 8: 190277-190299, 2020.

[5] A. Rabiee, S. M. Mohseni-Bonab, “Maximizing hosting capacity of renewable energy sources in distribution networks: A multi-objective and scenario-based approach,” Energy, 120: 417-430, 2017.

[6] S. Golshannavaz, S. Afsharnia, F. Aminifar, “Smart distribution grid: Optimal day-ahead scheduling with reconfigurable topology,” IEEE Trans. Smart Grid, 5(5): 2402-2411, 2014.

[7] T. Shekari, S. Golshannavaz, F. Aminifar, “Techno-economic collaboration of PEV fleets in energy management of microgrids,” IEEE Trans. Power Syst., 32(5): 3833-3841, 2017.

[8] S. Golshannavaz, M. Esmaeeli, F. Aminifar, M. Shahidehpour, “Cloud-based energy storage systems: A shared pool of benefits in distributed electric power systems,” IEEE Electrif. Mag., 10(2): 82-91, 2022.

[9] A. S. Awad, T. H. El-Fouly, M. M. Salama, “Optimal ESS allocation and load shedding for improving distribution system reliability,” IEEE Trans. Smart Grid, 5(5): 2339-2349, 2014.

[10] T. Sattarpour, D. Nazarpour, S. Golshannavaz, “A multi-objective HEM strategy for smart home energy scheduling: A collaborative approach to support microgrid operation,” Sustainable Cities Soc., 37: 26-33, 2018.

[11] T. Sattarpour, S. Golshannavaz, D. Nazarpour, P. Siano, “A multi-stage linearized interactive operation model of smart distribution grid with residential microgrids,” Int. J. Electr. Power Energy Syst., 108: 456-471, 2019.

[12] Y. Cao, D. Li, Y. Zhang, Q. Tang, A. Khodaei, H. Zhang, Z. Han, “Optimal energy management for multi-microgrid under a transactive energy framework with distributionally robust optimization,” IEEE Trans. Smart Grid, 13(1): 599-612, 2021.

[13] A. Azizivahed, M. Barani, S. E. Razavi, S. Ghavidel, L. Li, J. Zhang, “Energy storage management strategy in distribution networks utilized by photovoltaic resources,” IET Gener., Transmiss. Distrib., 12(21): 5627–5638, 2018.

[14] R. Li, W. Wang, M. Xia, “Cooperative planning of active distribution system with renewable energy sources and energy storage systems,” IEEE Access, 6: 5916-5926, 2018.

[15] W. Yi, Y. Zhang, Z. Zhao, Y. Huang, “Multiobjective robust scheduling for smart distribution grids: Considering renewable energy and demand response uncertainty,” IEEE Access, 6: 45715-45724, 2018.

[16] M. R. Narimani, A. A. Vahed, R. Azizipanah-Abarghooee, M. Javidsharifi, “Enhanced gravitational search algorithm for multi-objective distribution feeder reconfiguration considering reliability, loss and operational cost,” IET Gener., Transmiss. Distrib., 8(1): 55-69, 2014.

[17] T. Niknam, E. Azadfarsani, M. Jabbari, "A new hybrid evolutionary algorithm based on new fuzzy adaptive PSO and NM algorithms for distribution feeder reconfiguration," Energy Convers. Manage., 54(1): 7-16, 2012.

[18] M. Abdelaziz, "Distribution network reconfiguration using a genetic algorithm with varying population size," Electr. Power Syst. Res., 142: 9-11, 2017.

[19] M. Sedghi, A. Ahmadian, E. Pashajavid, M. Aliakbar-Golkar, “Storage scheduling for optimal energy management in active distribution network considering load, wind, and plug-in electric vehicles uncertainties,” J. Renewable Sustainable Energy, 7(3): 2015.

[20] H. Ma, Z. Liu, M. Li, B. Wang, Y. Si, Y. Yang, M. A. Mohamed, “A two-stage optimal scheduling method for active distribution networks considering uncertainty risk,” Energy Rep., 7: 4633-4641, 2021.

[21] A. Singh, A. Maulik, “Energy management of an active distribution network considering correlation between uncertain input variables,” Arabian J. Sci. Eng., 48(5): 6377-6398, 2023.

[22] S. Zhang, H. Cheng, K. Li, N. Tai, D. Wang, F. Li, “Multi-objective distributed generation planning in distribution network considering correlations among uncertainties,” Appl. Energy, 226: 743-755, 2018.

[23] S. Rezaeian-Marjani, S. Galvani, V. Talavat, M. Farhadi-Kangarlu, “Optimal allocation of D-STATCOM in distribution networks including correlated renewable energy sources,” Int. J. Electr. Power Energy Syst., 122: 106178, 2020.

[24] M. Aien, M. Fotuhi-Firuzabad, M. Rashidinejad, "Probabilistic optimalpower flow in correlated hybrid wind–photovoltaic power systems," IEEE Trans. Smart Grid, 5(1): 130–138, 2014.

[25] X. Lin, Y. Jiang, S. Peng, H. Chen, J. Tang, W. Li, "An efficient Nataf transformation based probabilistic power flow for high-dimensional correlated uncertainty sources in operation," Int. J. Electr. Power Energy Syst., 116: 105543, 2020.

[26] L. Wu, M. Shahidehpour, T. Li, "Stochastic security-constrained unit commitment," IEEE Trans. Power Syst., 22(2): 800-811, 2007.

[27] R. Billinton, R. N. Allan, Reliability Evaluation of Engineering Systems. New York, NY, USA: Springer, 1992.

[28] M. Gitizadeh, A. A. Vahed, J. Aghaei, "Multistage distribution system expansion planning considering distributed generation using hybrid evolutionary algorithms," Appl. Energy, 101: 655-666, 2013.

[29] S. Rezaeian-Marjani, S. Galvani, V. Talavat, "A generalized probabilistic multi-objective method for optimal allocation of soft open point (SOP) in distribution networks," IET Renewable Power Gener., 16(5): 1046-1072, 2022.

[30] A. Kavousi-Fard, M. Akbari-Zadeh, "Reliability enhancement using optimal distribution feeder reconfiguration," Neurocomputing, 106: 1-11, 2013.

Biographies

Safar Abbasi was born in Varzegan, East Azarbayjan, Iran, in 1980, received the B.Sc. degrees in Electronic Engineering from, Faculty of Engineering, Azad University, Tabriz, Iran, in 2006. Received the M.Sc. degrees in Power Electrical Engineering from Department of Electrical Engineering, Faculty of Engineering, Shahid Madani Azarbayjan University, Tabriz, Iran. Currently he is Ph.D. student in Power Electrical Engineering at Department of Electrical Engineering, Faculty of Computer, Electrical and Advanced Technologies, Urmia University, Urmia, Iran.

- Email: s.abbassi@urmia.ac.ir
- ORCID: 0009-0004-7604-017X
- Web of Science Researcher ID: 000900047604017X
- Scopus Author ID:
- Homepage: https://scholar.google.com/citations?hl=en-US&user=-LFaNd4AAAAJ

Daryoush Nazarpour was born in Urmia, Iran, in 1958. He received the B.Sc. degree from the Iran University of Science and Technology, Tehran, Iran, in 1982 and the M.Sc. and Ph.D. degrees from the Faculty of Engineering, University of Tabriz, Tabriz, Iran, in 1988 and 2005, respectively, all in Electrical Power Engineering. Currently, he is an Full Professor in Urmia University. His research interests include power electronics and flexible ac transmission system.

- Email: s.abbassi@urmia.ac.ir
- ORCID: 0000-0002-2327-6013
- Web of Science Researcher ID: 35409932600
- Scopus Author ID: 35409932600
- Homepage: https://scholar.google.com/citations?user=iAfvpYEAAAAJ&hl=en

Sajjad Golshannavaz was born in Urmia, Iran, in 1986. He received the B.Sc. (Honors) and M.Sc. (Honors) degrees in Electrical Engineering from Urmia University, Urmia, Iran, in 2009 and 2011, respectively. He received his Ph.D. degree in Electrical Power Engineering from School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran, in 2015. Currently, he is an Associate Professor in Electrical Engineering Department, Urmia University, Urmia, Iran. Since 2014 he has been collaborating with the smart electric grid research laboratory, Department of Industrial Engineering, University of Salerno, Salerno, Italy. His research interests are in smart distribution grid operation and planning studies, design of distribution management system (DMS), demand side management (DSM) concepts and applications, microgrid design and operation studies, design of energy management system (EMS), application of FACTS Controllers in Power systems, application of intelligent controllers in power systems. He can be contacted at email: s.golshannavaz@urmia.ac.ir

- Email: s.golshannavaz@urmia.ac.ir
- ORCID: 0000-0003-4999-8281
- Web of Science Researcher ID: AAB-5779-2020
- Scopus Author ID: 36677225300
- Homepage: https://scholar.google.com/citations?user=YzezRFUAAAAJ&hl=en

How to cite this paper:
S. Abbasi, D. Nazarpour, S. Golshannavaz, "A probabilistic framework for active distribution network optimal energy management considering correlated uncertain variables," J. Electr. Comput. Eng. Innovations, 12(2): 557-567, 2024.

DOI: 10.22061/jecei.2024.10837.741

URL: https://jecei.sru.ac.ir/article_2153.html