SRTTU

Journal of Electrical and Computer Engineering Innovations (JECEI)

Journal homepage: http://www.jecei.sru.ac.ir

JECEI

Research Paper

Mining Student Opinions from MOOC Discussions Using a Multi-Output BERT-Based Deep Learning Approach

Mujtaba Sultani [1], ***Negin Daneshpour*** [2,*]

[1] *Faculty of Computer Science, Kabul Polytechnic University (KPU), Kabul, Afghanistan.*

[2] *Faculty of Computer Engineering, Shahid Rajaee Teacher Training University (SRTTU), Tehran, Iran.*

Article Information

Article History:
Received 30 April 2025
Reviewed 19 June 2025
Revised 07 August 2025
Accepted 16 August 2025

Keywords:
BERT
CBiLSTM
Opinion Extraction
MOOC discussion forums

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

Abstract

Background and Objectives: Massive Open Online Courses (MOOCs) face unique challenges in extracting student feedback from large, asynchronous student discussion forums. While traditional survey methods are commonly used, they struggle with scalability and real-time analysis in the MOOC context. This study aims to address these limitations and focus on automated extraction and classification of student opinions and their urgency. The study bridges the gap between suggestion mining in commercial applications and educational domains.
Methods: We presented a novel deep learning approach using a BERT-based hybrid Convolutional Neural Network (CNN) – Bidirectional Long Short-term Memory (BiLSTM) multi-output model, named CBiLSTM. The model was trained to classify student posts into opinions and further categorize them by urgency. Performance metrics such as F1-weighted scores, Precision-Recall curves, and Area Under the Curve (AUC) were used to evaluate the model's efficacy, particularly in handling imbalanced datasets.
Results: The presented CBiLSTM model got an F1-weighted score of 87.3% for opinion classification and 81.1% for urgency classification, which represents an improvement of 1.3% and 1.8% over the best-performing baseline model. Precision-Recall curves and AUC metrics highlight the model's strength in balancing precision and recall. These findings demonstrate the model's capacity to accurately classify and prioritize student feedback in the educational domain.
Conclusion: This study offers a robust framework to enhance decision-making processes in MOOCs through effective feedback analysis. The CBiLSTM model provides a scalable, data-driven solution that empowers instructors, course designers, and policymakers to make targeted improvements and improves student engagement and course quality.

How to cite this paper:
M. Sultani, N. Daneshpour, "Mining student opinions from MOOC discussions using a multi-output BERT-Based deep learning approach," J. Electr. Comput. Eng. Innovations, 14(1): 145-162, 2026.

DOI: 10.22061/jecei.2025.11499.810

URL: https://jecei.sru.ac.ir/article_2393.html

Introduction

In recent years, online learning has gained more popularity among students and instructors globally, particularly with the rise of MOOCs. MOOCs are designed to present extensive access to open online resources on a global scale. MOOCs enroll a wide range of students, and this capacity for scalable instruction is a fundamental benefit they offer [1].

The collection and analysis of students' feedback regarding their learning experiences represent a foundational strategy for determining the quality of educational processes. In the context of traditional educational institutions, the practice of mandating mid-term or end-term surveys for students is prevalent. These surveys serve the purpose of soliciting students' perspectives on various aspects such as reaching course learning objectives and outcomes, the structure and presentation of the course, and the teaching methods and effectiveness of the instructors. This evaluative approach empowers both educators and institutional leaders to incorporate students' viewpoints into the ongoing monitoring and enhancement of the educational and learning process [2].

Academic institutions prioritize quantitative feedback that can be easily summarized and analyzed using statistical methods. Surveys usually contain closed-ended questions, commonly presented as Likert-scale items with varying rating scales, in order to capture students' opinions. While free-text comments are routinely collected, they are frequently underutilized, despite their potential to offer valuable and insightful perspectives on various aspects [3]. Incorporating open-ended questions enables the capture of automatic expressions of personal thoughts and emotions, granting students a platform to voice their perspectives and fostering a sense of value in their contributions. However, conventional feedback methods are impractical in MOOCs due to the high student-to-teacher ratios [4], [5]. An innovative approach is needed for effective course management, including real-time monitoring of student progress and feedback analysis. Detecting and comprehending student feedback is critical, given the reported attrition rates. Real-time feedback and adjustments are valuable to reduce disengagement. Low MOOC completion rates mean final evaluations may lack representativeness, and the voices of dropouts may be overlooked. Additionally, questionnaire wording can introduce bias [6], making natural discourse or interaction more effective for gathering student opinions.

The discussion forum within a MOOC has emerged as a promising aspect for gaining insights into course dynamics and tracking student progress [7]. These forums enable learner-instructor interactions as well as peer-to-peer communication [8]. They play a vital role in supporting diverse learning processes driven by the cognitive variances among MOOC participants. Additionally, these forums provide an essential platform for students to voice their questions and immediate concerns [9]. However, discussion forums have limitations due to their high volume of unstructured posts, which hinder instructors from effectively tracking and utilizing shared information to enhance learner retention and course quality. An efficient approach is to use computational models to process and summarize participants' feedback and suggestions within these forums, enabling ongoing evaluation of course-related elements.

Examinations of user-generated content within MOOC discussion forums reveal a multifaceted engagement, where participants share their course experiences as well as provide valuable opinions and suggestions for course enhancement [10]. While the practice of suggestion mining has traditionally been explored within Twitter data and reviews for commercial purposes [11], the fundamental objective remains constant, which includes extracting and utilizing participant insights. This process not only helps brand owners in refining product iterations but also empowers consumers to make more informed purchase decisions.

Furthermore, the principles of the mentioned task can be seamlessly applied to the realm of learning analytics. It serves both lecturers and course designers to improve course offerings and provides actionable insights for learners and policymakers, enhancing decision-making regarding course participation and promotion. This paper proposes a BERT-based hybrid multi-output deep learning model named CBiLSTM tailored for the extraction of urgent student opinions and suggestions from MOOC discussion forums. The model's primary goal is to identify and classify opinions expressed by students as either urgent or not urgent. To the best of our knowledge, in the education scope, especially within the context of MOOCs, this paper is the first to present and implement a machine learning-based model explicitly created for the extraction and classification of student opinions and suggestions from MOOC discussion forums. The primary findings of the paper are:

1. We introduce a BERT-based hybrid multi-output deep learning model to extract urgent student opinions and suggestions from MOOC discussion forums.
2. We present a method to identify and prioritize student feedback within MOOC forums, particularly focusing on opinions and suggestions.
3. The model creates the potential to enhance the decision-making process for instructors, course

designers, learners, and policymakers regarding course participation, promotion, and improvement.

4. The paper bridges the gap between suggestion mining in commercial contexts and its application in the educational domain.
5. As far as we understand, this is the first known paper to implement a deep learning model specifically for extracting urgent student suggestions within the MOOC context.

The subsequent sections outline the systematic progression of our research. In Section 2, we present a concise overview of related work in the field. Section 3 delves into the adapted method. In Section 4, we explain the data, experimental setup, and comparison methods used. The results of our analysis are reported in Section 5, with the conclusion provided in Section 6.

Related Work

This section presents a summary of prior studies relevant to the field. Given the extensive body of work on MOOCs and discussion forums over the past three years, our review focuses on empirical studies most pertinent to our research problem.

A fundamental component of MOOC learning support is the communication platform provided by discussion forums, which facilitates interaction between teachers, learners, and peers [8]. Research on user-generated content in MOOC discussion forums reveals that participants share their course experiences, voice their opinions, and provide suggestions for course enhancements [10]. Extracting student opinions and suggestions can help instructors, course designers, and policymakers enhance various aspects of the course and streamline the decision-making process. It can also help to find and extract the exact student problem that causes their dropout rates, a critical concern in the field [12]. The challenge of suggestion mining has primarily been examined in the context of reviews and Twitter data, with a predominant focus on commercial applications [11].

Ramanand et al. [11] addressed two challenges in opinion and intention mining: identifying 'wishes' for product improvements and making purchases. The proposed approaches that use English-language patterns are the first attempts at solving these problems. The wish detection method is most effective for texts with explicit wishes, like customer surveys, and moderately effective for electronic product reviews, but less so for banking service reviews. The approaches are effective in specific contexts but require improved datasets. Negi et al. [13] defined suggestion mining as identifying text that directly proposes or recommends an action or entity. They introduced the use of forum posts for suggestion mining, and showed that deep neural network algorithms outperformed SVM and rule association methods for both in-domain and cross-domain evaluations. Alotaibi et al. [14] extracted suggestions from opinionated text, utilizing the XGBoost classifier and word-embedding techniques. Their methodology achieved over 80% accuracy when evaluated on hotel reviews and Microsoft Windows App Studio discussion data. The study emphasized the importance of suggestion-related keywords and affirmed XGBoost's effectiveness in suggestion extraction.

Brun & Hagege [15] extracted suggestions for improvement from user comments. The system utilizes NLP (Natural Language Processing) techniques, including a deep syntactic parser and syntactic-semantic patterns to analyze customer reviews and identify valuable suggestions. The system achieved an F1-score of 73% on a corpus of printer reviews from the 'Epinion' website. In recent study, Laskari & Sanampudi [16] proposed a novel hybrid model for fine-grained analysis of suggestions with aspect orientation for commercial purpose. They utilized two different datasets and evaluated the performance of their approach using various machine learning, neural network, and transfer learning models. The transfer learning approach outperformed others. Almatrafi & Johri [10] proposed an approach that analyzes MOOC discussion forum posts to summarize participants' opinions on different aspects of a course and recognize suggestions for improvement. This is the first study that discussed suggestion mining in an educational context. The study used sentiment analysis to detect participants' attitudes and rule-based techniques to identify suggestions. The results show that the approach effectively identifies aspect-based sentiments and recommendations towards course design elements.

The studies [11], [13]-[16] have employed NLP techniques and machine learning models to extract valuable suggestions for improvement from user comments. While these studies have made significant contributions in their respective domains, there remains a notable gap in the research within the education domain, specifically in the context of MOOCs. Although Almatrafi & Johri's [10] work represents a significant step forward in the education domain, there is still room for further exploration. Notably, advanced machine learning techniques such as deep learning models and transfer learning, which have shown promise in other domains, have yet to be fully utilized in this context. Furthermore, while the study [17] has successfully identified urgent questions, there has been no research specifically designed to mine and analyze student opinions based on their urgency. This leaves a gap in understanding the broader spectrum of student feedback and sentiment in online learning environments. To overcome these gaps, we present a novel approach for opinion and suggestion mining in MOOC discussion forums. The primary advancement is the development of a BERT-based hybrid CBiLSTM multi-output deep learning model developed to identify and classify urgent student opinions and suggestions. Table 1 summarizes the findings of previous research results.

Table 1: Overview of past study results

Study	Data	Approach	Results
[11]	Electronic and banking reviews.	Rule-based approach	This paper introduced methods to identify 'wishes' for product improvements and purchase intentions. The approaches are effective in specific contexts but require improved datasets.
[13]	1. Travel advice2. 2. Suggestion forum. 3. Tweets with hashtags: suggestion, advice, recommendation, warning.	1. Rule-based 2. SVM 3. DL (CNN & LSTM)	The study focused on suggestion detection & extracting different types of suggestions from opinionated text. It introduced new suggestion-rich datasets and compared various approaches for suggestion detection, highlighting the potential of deep learning models.
[14]	1. Hotel reviews. 2. Microsoft Windows App Studio discussion data.	Word Embedding + XGBoost	This study extracted suggestions, tips, and advice from social media data to enhance business decision-making and customer benefits. Using word embedding and XGBoost classifiers, it achieved 80% accuracy, highlighting the significance of suggested keywords and phrases.
[15]	Printer review corpus	Rule-based Approach	The study introduced an automated approach to detect expressions suggesting product improvements within customer reviews. The system achieved F1-score of 73% on a corpus of printer reviews from the website 'Epinion'.
[16]	1. Travel Reviews 2. MS Windows Phone	1. Rule-based 2. SVM & NB 3. CNN & LSTM 4. BERT	This paper presented a hybrid approach for aspect-oriented suggestion mining from opinion reviews, featuring two key phases: aspect term extraction and suggestion classification. The transfer learning approach outperformed others.
[10]	Stanford MOOC Posts	Rule-based	This study improves MOOCs by analyzing forum feedback for course enhancement through sentiment analysis and suggestion mining. They applied a rule-based approach for suggestion extraction and achieved a 31% F1-score.

Method

This study presents a novel BERT-based CBiLSTM multi-output classification model to mine student urgent opinions in MOOC discussion forums. The model integrates CNN and Bi-LSTM layers with BERT to carry out a multi-output classification task within the context of MOOC discussion forums. First, it distinguishes between opinionated and non-opinionated student posts, allowing for the identification of those expressing subjective viewpoints. Subsequently, the model delves deeper by categorizing these opinionated posts into either urgent or non-urgent parts, which enables the differentiation of opinions that demand immediate attention or action from those that are less time-sensitive.

A. Preprocessing

To maintain the integrity of the text and ensure meaningful analysis, several preprocessing steps are undertaken. We transformed common linguistic abbreviations such as "'re," "n't," and "'s," into their corresponding full-word forms ('are', 'not', 'is'), which can improve text readability. To remove potential noise or unrelated content, we removed URLs (Uniform Resource Locators). Furthermore, symbols like question marks, exclamation marks and ampersands are replaced with their corresponding textual equivalents. Similarly, to standardize word forms, contractions like "won't" and "can't" are substituted with expanded versions, such as "will not" and "can not", facilitating uniformity in text representation.

Specific characters, including slashes and dollar, signs are eliminated, which ensures that the text remains free from interference with its semantic meaning. Stop words are intentionally retained as their existence had a positive impact on the classification result [18]. In order to facilitate lemmatization, the 'en_core_web_ls' model from Spacy is employed to reduce word variations to their base/dictionary forms. Finally, the 'course_display_name' metadata attribute is incorporated with student posts, as it notably enhances the results [19].

B. Balancing the Dataset

The Stanford MOOCPosts dataset exhibits a substantial imbalance. In the classification of student opinions as urgent or not urgent, we face a scarcity of urgent records compared to the abundance of non-urgent ones. This imbalance can hinder model training, potentially leading to biased results.

Addressing the issue of data imbalance in text classification tasks is a common challenge within the realm of natural language processing [20]. Often, certain classes or groups in a dataset are significantly underrepresented, which may negatively impact the performance of machine learning models, particularly for minority classes. Synthetic data generation and oversampling methods, such as SMOTE and AdaSyn [21], [22], have proven effective in mitigating statistical imbalances.

However, when applied to textual data, they encounter challenges related to overfitting and noise. Although Generative Adversarial Networks (GANs) like CycleGAN [23] have shown promise in generating synthetic data for numerical and image datasets, their applicability to textual data, with its inherent complexities involving grammatical structure, context, and semantic information, necessitates careful evaluation.

We use BERT (Bidirectional Encoder Representations from Transformers), which leverages the language understanding to balance the dataset. The process involves tokenization and [MASK] token insertion to create augmented samples. These samples are then processed through BERT for replacements. Table 2 presents the original data and its augmented variations. Our augmentation strategy focuses on enhancing the 'urgent' class samples which effectively mitigates data imbalance concerns and leads to significant improvements in model performance. However, as we work towards striking a balance within the urgency classification, we noticed a potential effect on the balance between 'opinion' and 'not opinion' labels shown in Fig. 1, as both aspects are intrinsically connected. Consequently, our dataset is, to some extent, balanced, leading to significant improvement in model performance.

Table 2: Primary posts with their respective augmented variations

No	Initial Posts	Enhanced Posts	Opinion	Urgent
1	how long would it take to grade the peer assessment question education education one one five number how to learn math	how long would it take me to have grade over the program peer assessment question good education more education one four one five number how to learn math	No	No
2	i think these statement show the student know how to complete a certain skill but they do not necessarily show any understanding of the content education education one one five number how to learn math	i actually think these statement would show the student we know well how to complete a certain skill but that they do not quite necessarily show any understanding in of it the content education education one point one five number how to fully learn math	Yes	No
3	i find the professor go too fast while he speak he do not stop to explain to give you time to understand I have to stop and go backwards constantly especially when he talk about percentage he be like read very fast do not know how to give a lecture maybe I be spoil by another professor of another course I find edx do not answer our post there be not staff available humanity science economy one summer two zero one four	i find the professor go far too far fast while he speak he do not stop to explain to give you time to understand i have to t stop and go even backwards constantly especially when he talk about percentage he be like read very fast do not know how to and give a lecture too maybe i be a spoil by another professor of another course i find edx do not answer about our post there to be not staff available humanity science economy one summer two zero zero one four	Yes	Yes

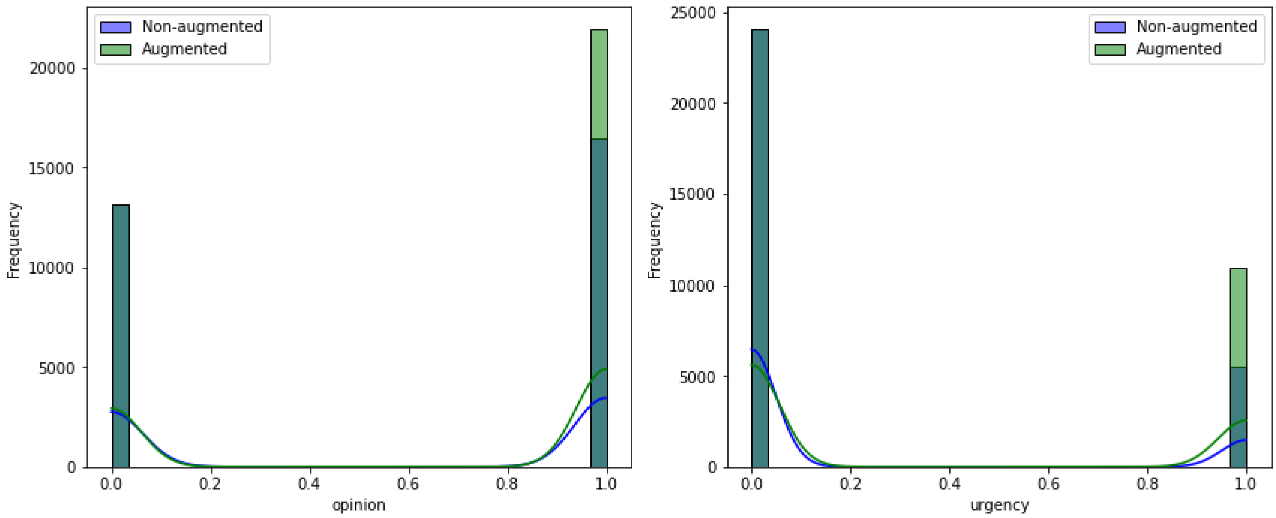

Fig. 1: Distribution of opinion and urgency samples following the augmentation process.

C. Embedding Layer

Word Embedding is used to transform words or tokens in a text into dense numerical vectors within a high-dimensional space. In the realm of NLP, conventional methodologies like Word2Vec and Glove were widely adopted prior to the advent of transformer-based models like BERT. BERT, as introduced by Devlin et al. [24], is a transformer-based model that represents contextual embeddings. It is pre-trained on a vast corpus using next sentence prediction and masked language modeling tasks. This pre-training enables BERT to understand the semantic meaning and syntactic structure of words in various contexts. The word embeddings produced by BERT not only capture individual word meanings but also their contextual relevance within sentences or documents.

Key to BERT's approach are special tokens, "[CLS]" and "[SEP]," which play an important role in handling variable-length sequences of text. The "[CLS]" token functions as a classification token, generating a fixed-size vector that represents the entire input sequence. For text classification tasks, BERT uses the last hidden state of the "[CLS]" token to represent the whole sequence. Meanwhile, the "[SEP]" token serves as a segment separator, particularly valuable for sentence-level classification tasks [24]. The flexibility of fine-tuning strategies allows researchers to explore various approaches, from training the complete architecture (all layers, including pre-trained and task-specific, are updated during fine-tuning) to selectively updating specific layers (only a subset of BERT layers is updated, while others are frozen), or maintaining BERT as a feature extractor (all layers, including pre-trained and task-specific, are kept frozen).

In this study, we utilize the pre-trained BERT model, the same one used for tokenization.

During the fine-tuning phase, the BERT model will be kept frozen, and only the CBiLSTM component will be trained to learn from BERT's representations. It's worth highlighting that the tokenization process utilizes the pre-trained BERT model, specifically the 'bert-base-uncased tokenizer' from the transformer's library. This tokenizer is designed to be case-insensitive and optimized for the specific requirements of processing textual data.

D. CBiLSTM Multi-output Hybrid Deep Learning Model

The choice of combining BERT with CNN and BiLSTM is built on the complementary strengths of these components. BERT offers contextual embeddings that capture the semantic richness of student posts, while CNN efficiently extracts local features such as opinion-indicative phrases. BiLSTM further enhances the model's capability to learn bidirectional dependencies, which is essential to understand nuanced and sequenced expressions in forum discussions. We opted for this configuration over deeper architectures such as RoBERTa, T5, and GPT variants to maintain a balance between performance and computational efficiency. This ensures the model remains feasible for deployment in real-world educational settings. This hybrid design also outperformed simpler models such as BiLSTM-only or CNN-only in our comparative evaluations, as detailed in the results section.

In text classification and sentiment analysis, CNN and Recurrent Neural Networks (RNN) are widely used. CNN excels at local feature extraction and understanding spatial relationships [25], while RNN is proficient in capturing sequential dependencies and global features [26]. Traditional RNNs face challenges like gradient problems, particularly with extensive data sequences. To overcome this, the LSTM was introduced. LSTM uses memory cells to overcome gradient issues and capture long-term relationships [26]. Conventional LSTM models involve unidirectional data flow, which effectively models past dependencies but may not capture future context, which is important in text classification.

LSTM utilizes a gating mechanism involving the forget gate (fg_t), input gate (ig_t), and output gate (og_t) to regulate information flow within its cells [27]. These gates are constructed using a sigmoid neural network layer combined with pointwise multiplication operations. At time step t, the current cell state of the LSTM is denoted as c_t. The forget gate, responsible for determining which information to discard from the previous state, is calculated as:

$$fg_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where w_f and b_f are the weight matrix and bias of the forget gate, x_t is the input at time t, and h_{t-1} is the hidden state from the previous time step. Next, the input gate identifies which part of the current input x_t should be added to the current cell state c_t. This process involves both sigmoid and tanh layers. The sigmoid layer determines what information to update in the current cell state, and is defined as:

$$ig_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

The candidate cell state is generated through a tanh layer:

$$c'_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

where W_i, W_c and b_i, b_c are the weights and biases for the input and candidate state layers, respectively.

Then, LSTM updates the previous cell state c_{t-1} with the new cell state c_t by performing element-wise multiplication of the forget gate values with the previous cell state c_{t-1}. This retains essential information from the current input while discarding irrelevant or outdated information from the previous cell state is calculated as:

$$c_t = fg_t \cdot c_{t-1} + c'_t \cdot ig_t \quad (4)$$

To generate the desired output, the LSTM leverages its output gate, is calculated as:

$$Og_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = og_t \cdot \tanh(c_t) \quad (6)$$

The final hidden state, h_t, is then directed to a densely connected layer for further processing.

BiLSTM extends LSTM by using two layers that process data in both forward and backward directions, as illustrated in Fig. 2 [28]. The bidirectional approach enhances its ability to capture bidirectional data dependencies, making it highly effective in various text classification tasks.

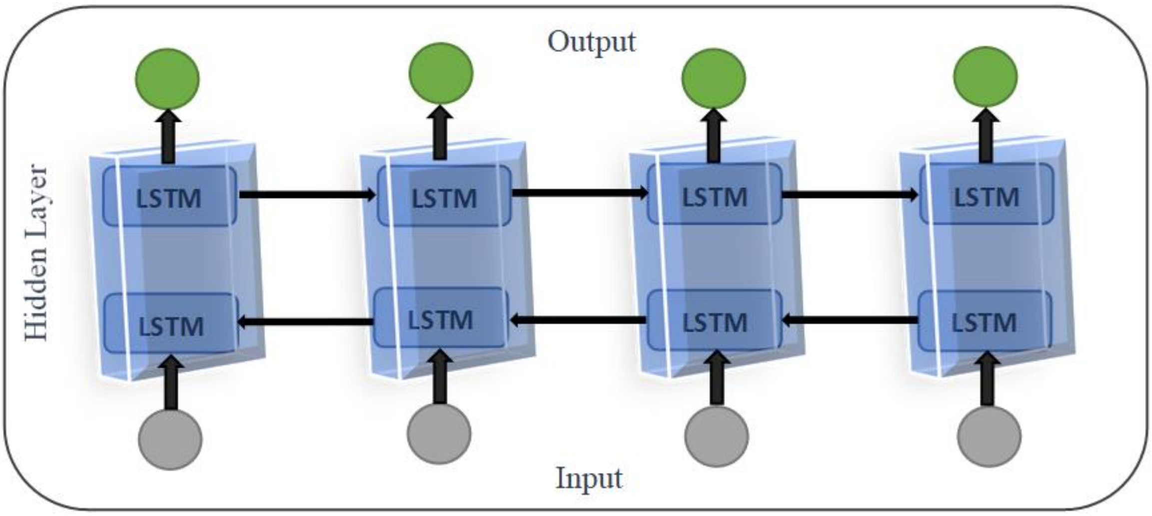

Fig. 2: Depiction of BiLSTM model.

We present a multi-output hybrid deep learning model to extract and classify urgent student opinions, which leverages a pre-trained BERT model as the embedding layer. The model's architecture involves a one-dimensional convolutional layer with 128 filters and a kernel size of 3, followed by a max-pooling layer with a pool size of 3 to capture local features. To mitigate overfitting, we applied dropout regularization with a rate

Opinion classification relies on two bidirectional LSTM layers, featuring 256 and 128 units respectively, followed by another dropout layer and a flattening operation. The dense output layer contains a single neuron with a sigmoid activation function to handle binary opinion classification.

To filter for opinions, we introduce a 'Filter_opinions' layer by performing element-wise multiplication of the output from the second LSTM layer with the opinion classification result.

For urgency classification, we implement another set of bidirectional LSTM layers, mirroring the opinion classification architecture. We set the BERT layers of the model to be non-trainable, which ensures that the pre-trained embeddings do not get updated and remain fixed during training.

We use Adam optimizer with binary cross-entropy loss function and apply some helpful techniques like model checkpoint, early stopping, and adjusting the learning rate during model training. It's necessary to note that we run the model for 10 epochs with a batch size of 200. The general overview of the model architecture is presented in Fig. 3.

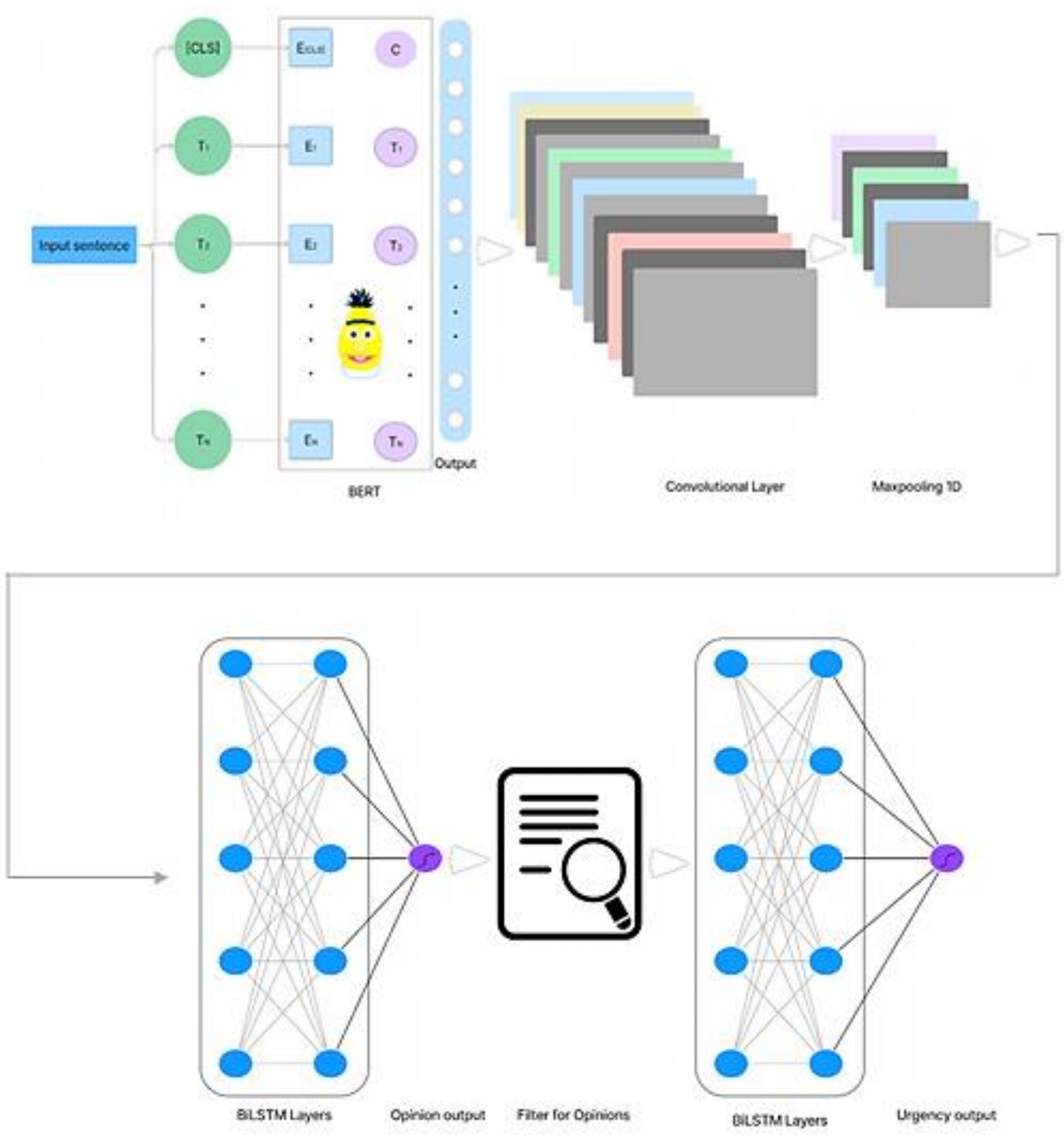

Fig. 3: Proposed model architecture.

Experiments

This section offers an overview of the data, details of the experimental setup, and the evaluation metrics.

E. Dataset and Experimental Setup

In this paper, the experiments are conducted using the Stanford MOOC Posts dataset, a benchmark corpus introduced by Agrawal et al. [29]. The corpus comprises 29,604 anonymized learner forum posts from 11 public online courses at Stanford University. These posts are categorized into three distinct domains: Humanities/Sciences, Medicine, and Education. Each post is manually labeled across various dimensions, including assessing whether the post is student opinion or non-opinion, and the overall urgency of the posts is ranked on a scale from 1 to 7. To create a binary classification task for urgent post-identification, the class labels are adjusted. Posts with urgency scores of 3 or higher are categorized as 'urgent', while those with scores less than 3 are labeled as 'not urgent'. We again adjust the urgency label for non-opinion posts by changing the value from 'urgent' to 'not urgent'. Modifying the urgency label for non-opinion posts within the Stanford MOOC Posts dataset is an important adjustment aimed at enhancing the precision and contextual relevance of the classification process. In this dataset, the urgency label originally indicates the importance of the student post, irrespective of whether it contains an opinion.

However, in the context of our multi-output model, which performs a dual classification task to identify both opinions and their urgency, it is vital to ensure that urgency classification is closely aligned with the presence of an opinion. Therefore, we made the necessary refinement by reassigning the urgency label to 'not urgent' for those student posts categorized as 'not opinion.'

This adjustment serves a twofold purpose.

First, it ensures that urgency classification is meaningful only in the context of opinionated student comments, which inherently have a subjective and potentially actionable nature.

Second, it minimizes any potential confusion or misclassification of non-opinion posts as urgent, thereby contributing to the overall accuracy and applicability of the model's outcomes.

This binary classification scheme guarantees that approximately 7% of posts are classified as urgent opinions, which can then be used by course designers, policymakers, and instructors to enhance course quality. The schema not only saves time but also empowers them to leverage students' opinions and suggestions for current and upcoming courses.

The model was implemented using TensorFlow and Keras libraries on Python 3.10, and trained on Google Colab's free GPU environment. Training was conducted for 10 epochs, taking approximately 3 hours in total. The session utilized Colab's Tesla T4 GPU, which provides 16 GB of VRAM and supports large-scale model training. Before feeding the data into the model, we conducted a series of preprocessing steps to ensure data integrity and uniformity.

For tokenization and word embedding, we leverage the pre-trained 'bert-based-uncased' model from the transformer's library, introducing specialized tokens to format input sequences correctly. Considering the model's input constraint of 512 tokens, we truncate sequences to the defined maximum length, ensuring our data aligns with the BERT model's requirements. As detailed in the CBILSTM Multi-output Hybrid Deep Learning Model subsection, we carefully select the optimizer, the number of layers and hidden units for BiLSTM components in the presented model. These decisions are made after thorough experimentation and optimization to achieve optimal performance.

F. Evaluation Metrics

Despite employing BERT-based data augmentation to address the imbalance problem, some level of imbalance still exists in the dataset. Therefore, to effectively evaluate the model performance, we use Learning Curve (LC) assessment, precision (PR), recall (RC), F1-score, weighted F1-score, and Precision-Recall Curve (PRC) analysis [30].

LC is a diagnostic tool, frequently used in machine learning, especially for models with incremental learning like deep learning [31]. We evaluate model performance by analyzing training and validation data, yielding two important curves: the Training Learning Curve, revealing how well the model learns from training data, and the Validation Learning Curve, assessing its knowledge generalization.

Precision measures how accurately the model identifies positive outcomes, while recall evaluates how well it detects relevant positive results. The F1 score combines these two aspects into a single performance measure, balancing precision and recall. Equations (7)-(9) provide the mathematical formulas for these metrics:

$$PR = TP / (TP + FP) \tag{7}$$

$$RC = TP / (TP + TN) \tag{8}$$

$$\text{F1-score} = (2 \cdot PR \cdot RC) / (PR + RC) \tag{9}$$

True Positives (TP) refer to the samples that were correctly classified as positive, True Negatives (TN) are the samples correctly identified as negative, while False Positives (FP) are the instances that were incorrectly identified as positive. To address class imbalances, we employ the F1-weighted score, which computes the weighted average of the F1 scores for each class.

PRC, a vital tool in binary classification [30], captures the balance between true positive rate and positive predictive value.

Its significance shines when handling imbalanced datasets, where one class dominates. In such cases, the Precision-Recall Plot excels in evaluating binary classifiers.

To summarize performance and contrast classifiers, we utilize the Area Under the Curve (AUC) metric. It measures the model's proficiency. The precision-recall curve's baseline, $y = P/(P + N)$, with P for positives and N for negatives, sets the standard for a no-skill classifier that can't differentiate among classes.

Result and Discussion

This section provides the performance analysis of our proposed hybrid CBiLSTM multi-output deep learning model. The model serves a dual purpose: it initially classifies student comments into 'opinion' or 'not opinion,' and subsequently further categorizes opinionated comments into 'urgent' or 'not urgent'.

Given the distinctive nature of our research focus, it is essential to acknowledge that there is only one prior study [10] in this domain, which primarily employed a rule-based approach for the extraction of student opinions and suggestions. Other studies, while insightful in their own right, operate within divergent domains and are based on different datasets, making a direct one-to-one comparison impractical. To maintain methodological clarity and uphold the integrity of our comparative analysis, the results of these previous papers are shown in Table 3.

Prior studies, as seen in Table 3, primarily highlighted their approaches' success in classifying suggestions from reviews and comments, they provided scores only for the suggestion class. The references [13]-[16], conducted outside the realm of education, especially from the MOOC context, employed rule-based, machine learning, deep learning, and transfer learning methodologies. These studies achieved F1 scores ranging from 72.7% to 89.6%, reflecting the distinctive nature of their objectives. The only study [10], operating within the same domain, applied a rule-based approach to extract student suggestions from MOOC discussion forums and achieved a 31% F1-score.

In contrast, as presented in Table 4, our proposed model demonstrates better performance, achieving precision, recall, and F1-scores of 91.1%, 86.4%, and 89.8%, respectively, for the opinion/suggestion class. Additionally, it maintains strong classification results for the not opinion class, which underscores the model's efficiency.

Additionally, we contrast the proposed model with base models, namely LSTM, BiLSTM, GRU, BiGRU, and CNN + BiGRU. For all base models as well as our proposed model, we used a BERT pre-trained model as an embedding layer, which can help train the model quickly with less epochs and minimal loss comparing to other word embedding techniques [30].

Less epochs with minimal loss mean lower cost and higher efficiency. The following subsection demonstrates the model performance using different evaluation metrics.

Table 3: Summary of prior papers

Model	Dataset	Suggestion		
		P (%)	R (%)	F1 (%)
LSTM [13]	Suggestion Forum	73.8	71.6	72.7
XGBoost [14]	MSWASR	81	83	84
Rule-based [15]	Review corpus ‘Epinion’	77	70	73
BERT-large [16]	MS Windows Phone	85.3	84.6	85.7
BERT-base [16]	Travel	90.2	89.6	89.6
Rule-based [10]	Stanford MOOCPosts	22	50	31

A. Learning Curve Analysis

As depicted in Figs. 4 to 9, the learning curves and validation curves are plotted for the different used models. We can see that the proposed CBiLSTM model achieved early stopping with minimal loss in few training iterations, resulting in improved efficiency and cost effectiveness.

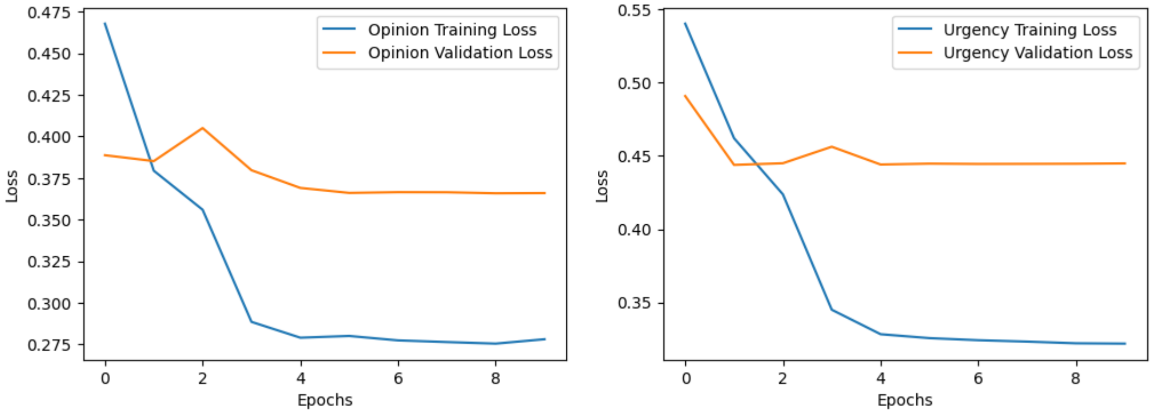

Fig. 4: Learning & validation curves with optimal loss and relevant Epochs for GRU model: Opinion (loss 0.365, Epoch 8), Urgency (loss 0.443, Epoch 5).

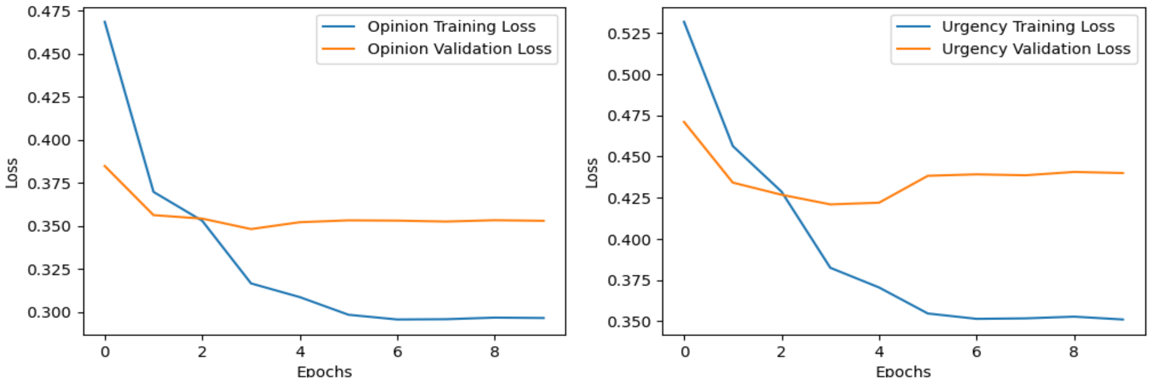

Fig. 5: Learning & validation curves with optimal loss and relevant Epochs for LSTM model: Opinion (loss 0.348, Epoch 3), Urgency (loss 0.42, Epoch 4).

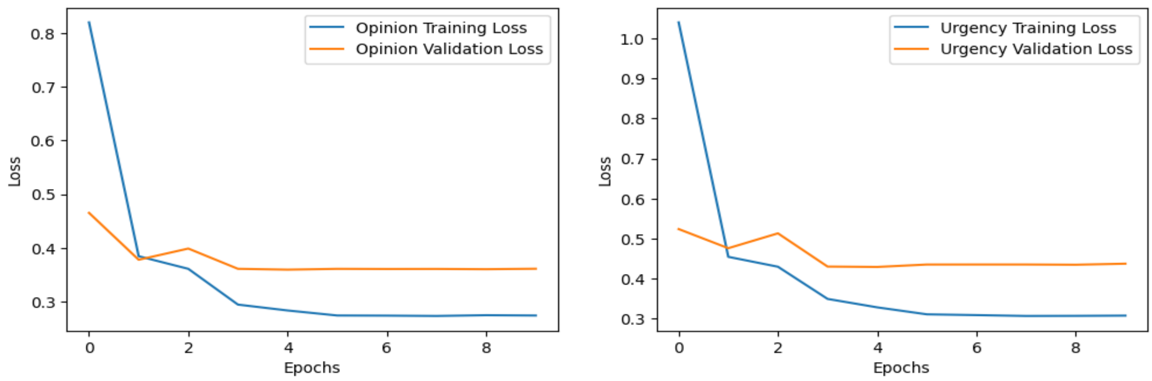

Fig. 6: Learning & validation curves with optimal loss and relevant Epochs for BiGRU model: Opinion (loss 0.339, Epoch 5), Urgency (loss 0.429, Epoch 4).

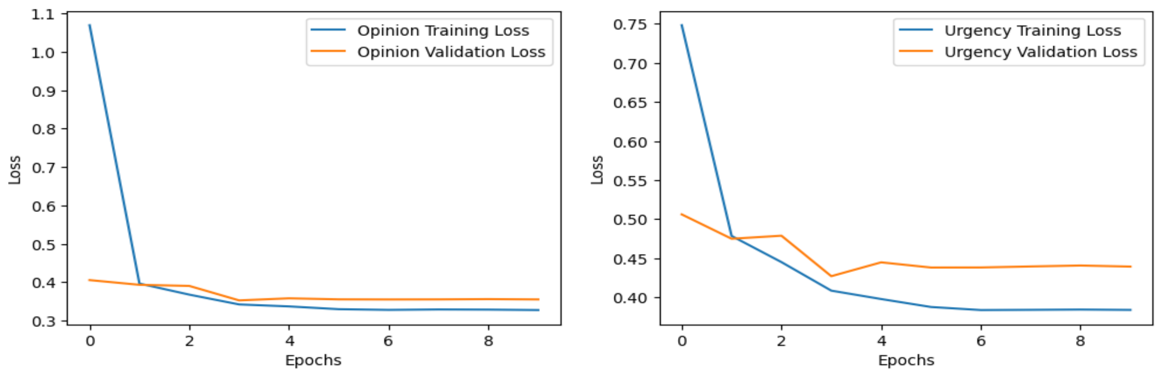

Fig. 7: Learning & validation curves with optimal loss and relevant Epochs for BiLSTM model: Opinion (loss 0.35, Epoch 4), Urgency (loss 0.426, Epoch 3).

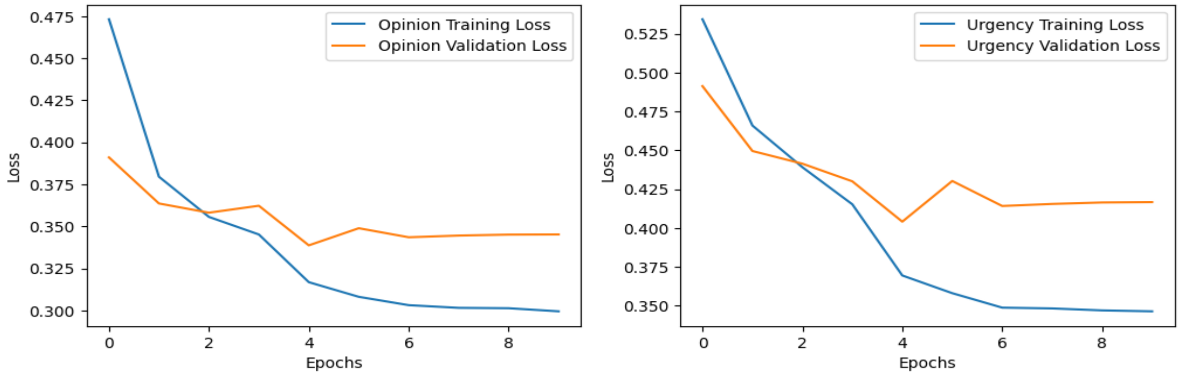

Fig. 8: Learning & validation curves with optimal loss and relevant Epochs for CBiGRU model: Opinion (loss 0.338, Epoch 4), Urgency (loss 0.404, Epoch 4).

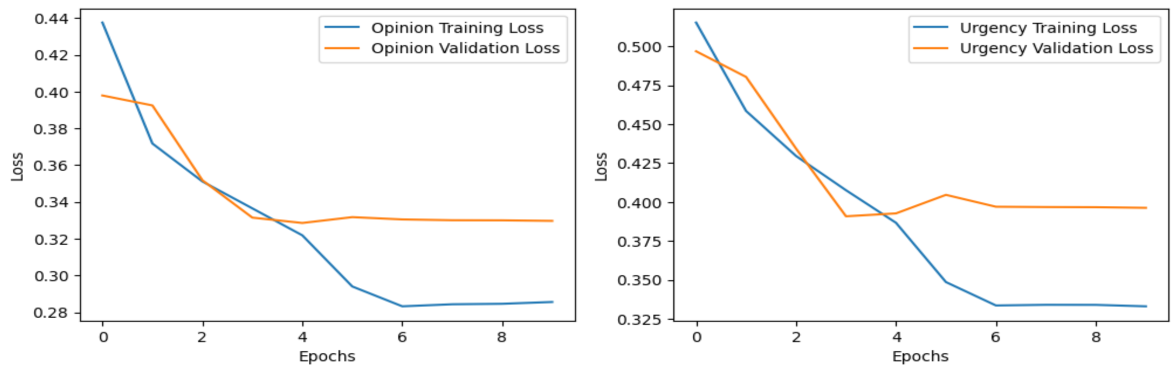

Fig. 9: Learning & validation curves with optimal loss and relevant Epochs for CBiLSTM model: Opinion (loss 0.322, Epoch 4), Urgency (loss 0.384, Epoch 3).

B. Precision-Recall Curve Analysis

In Figs. 10 and 11, we present the Precision-Recall curves that offer insight into the balance among the true positive rate and positive predictive value for opinion and urgency classification tasks. The AUC is employed as a quantitative metric to show the model performance.

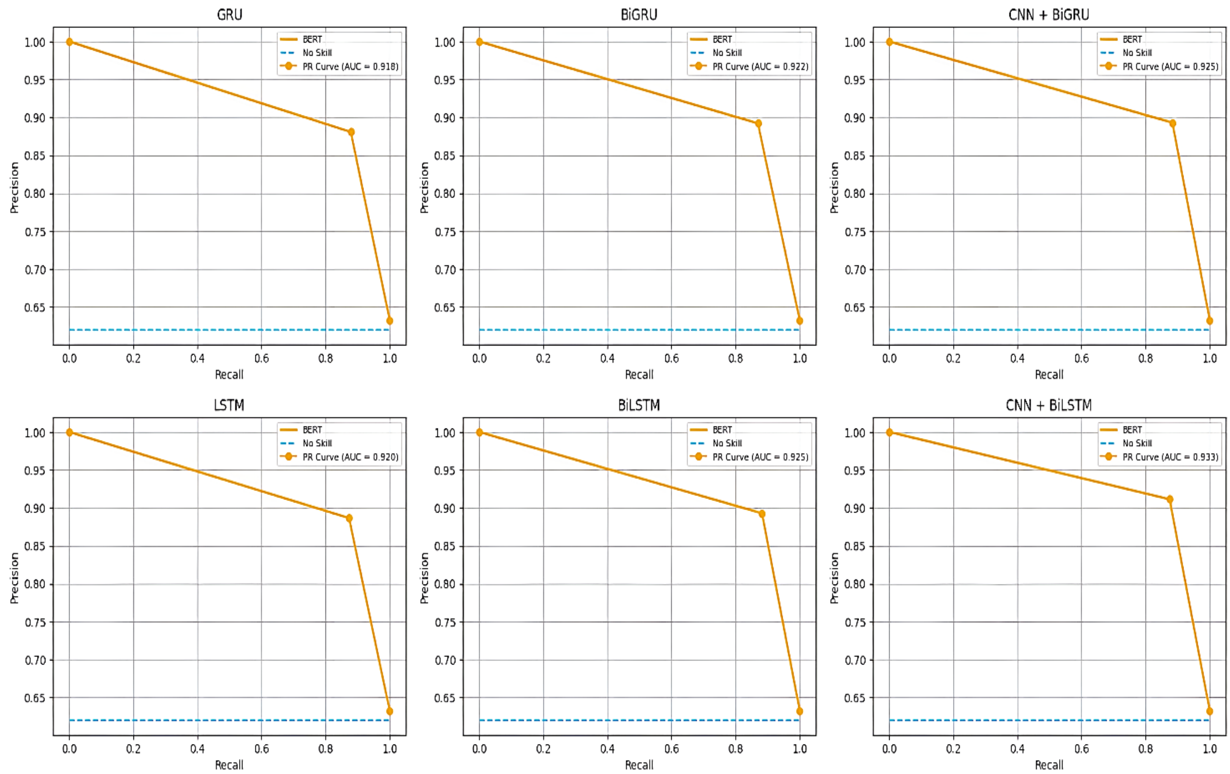

Fig. 10: Precision-Recall curves for opinion classification using GRU, LSTM, BiGRU, BiLSTM, CNN + BiGRU, and the proposed CBiLSTM models.

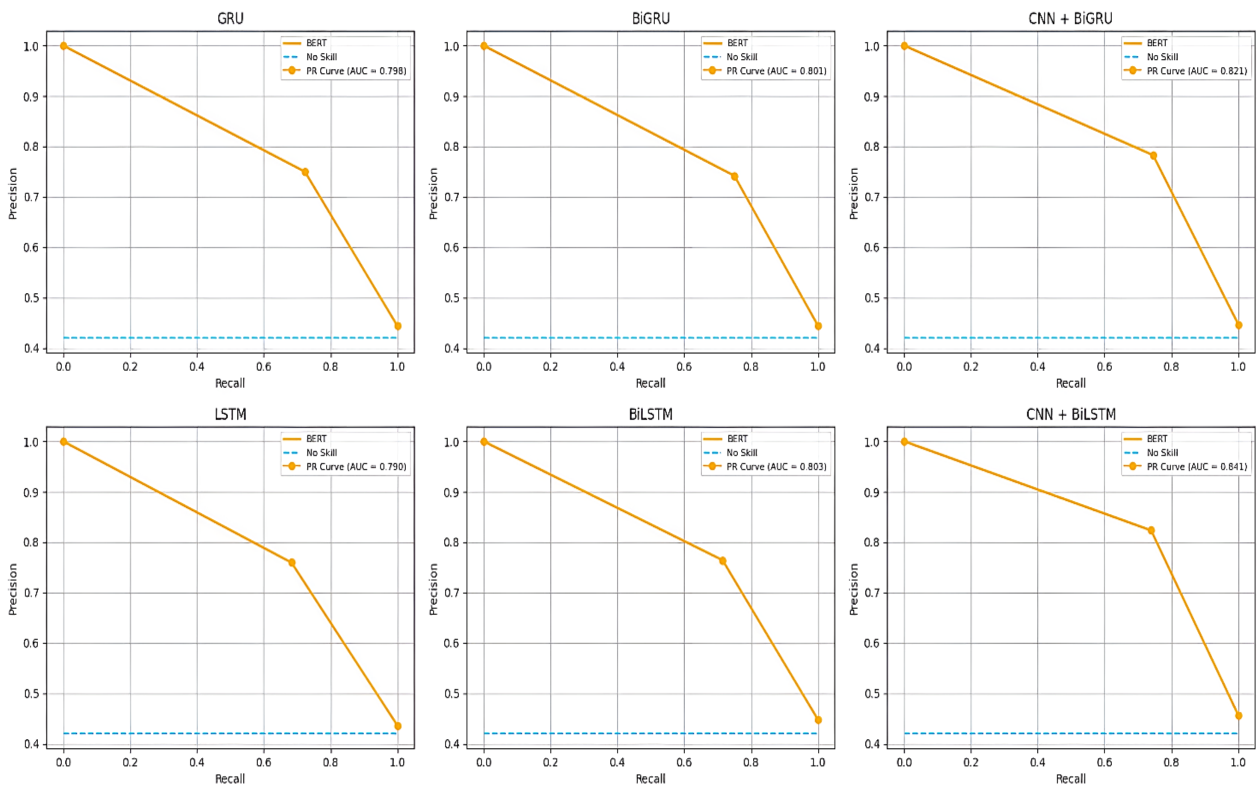

Fig. 11: Precision-Recall curves for urgency classification using GRU, LSTM, BiGRU, BiLSTM, CNN + BiGRU, and the proposed CBiLSTM models.

Furthermore, we employ a baseline for the precision-recall curves to show the balance between classes in both classification tasks. Notably, across all baseline models, CBiLSTM achieved the highest AUC values for both opinion and urgency classifications, which shows the model's effectiveness.

C. F1-Weighted Analysis

The presented CBiLSTM model outperformed the baseline models in both opinion and urgency classification tasks. As depicted in Tables 4 and 5, the CBiLSTM model has achieved better results, boasting 87.3% F1-weighted score for opinion classification and 81.1% F1-weighted score for the urgency classification task.

This represents a notable advancement of 1.3% and 1.8% over the best-performing baseline model on opinion and urgency classification tasks, respectively. Moreover, our model achieved 89.8% F1-score for opinion class and 78.4% F1-score for urgent class, a boost of 1% and 2.1% compared to the best-performing baseline model.

Table 4: Opinion classification task experimental results

Model	**Opinion/Suggestion**			**Not Opinion**			**Weight F1 (%)**
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	
LSTM	88.7	87.3	88	78.7	80.8	79.8	84.9
GRU	88.1	87.9	88	79.2	79.6	79.4	84.8
BiLSTM	89.3	88.2	88.7	80.2	81.8	81	85.9
BiGRU	89.2	86.9	88.1	78.5	81.9	80.2	85.2
CNN + BiGRU	89.3	**88.4**	88.8	**80.3**	81.8	81.1	86
CNN + BiLSTM	**91.1**	86.4	**89.8**	80.1	**84.2**	**82.5**	**87.3**

Table 5: Urgency classification task experimental results

Model	**Urgent**			**Not Urgent**			**Weight F1 (%)**
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	
LSTM	75.9	68.3	71.9	77.3	83.3	80.2	76.6
GRU	75	72.3	73.6	78.6	80.8	79.7	77
BiLSTM	76.4	71.5	73.9	78	82.1	80	77.2
BiGRU	74.2	75	74.6	79.8	79.1	79.5	77.3
CNN + BiGRU	78.3	74.5	76.3	80.2	83.3	81.7	79.3
CNN + BiLSTM	**82.5**	**74.2**	**78.4**	**80.2**	**86.7**	**83.5**	**81.1**

D. Analysis of Word Embedding

To investigate the effect of different word embeddings, we evaluated the CBiLSTM model with four BERT pre-trained models: 'bert-based-cased', 'bert-based-uncased', 'bert-large-cased', and 'bert-large-uncased'. Our analysis, as displayed in Table 6, shows that the 'bert-based-uncased' model outperformed the other three models in both opinion and urgency classification tasks. This result is attributed to the uncased model's robustness against case-related variations in text, leading to more consistent embeddings for words with varying capitalization.

E. Impact of Dataset Balance on Model Performance (Ablation Study)

Assessing a model's performance on both balanced and imbalanced data is important for understanding the model robustness and reliability in real-world scenarios, where class distributions can vary widely. As shown in Table 7, the CBiLSTM model performed well on both balanced and imbalanced data, demonstrating its robustness. This comparison serves as an ablation study that quantifies the effect of the BERT-based masked token augmentation technique used to balance the dataset. The results indicate that balancing the dataset leads to better overall performance and suggest that a balanced dataset can provide a more conducive training environment for the model to learn effective patterns without skewed bias toward more prevalent classes.

F. Comparative Analysis of CNN and Maxpooling Layers

The number of hidden layers in a neural network, which defines its depth, plays an important role in the model's ability to capture complex data patterns, thereby affecting both accuracy and training efficiency [32]. While increasing the number of hidden layers can enhance accuracy by capturing intricate patterns, it may also reach a point where further depth results in reducing returns or strengthen overfitting [33].

We conducted a set of experiments to examine how modifying key parameters in the convolutional and max-pooling layers, such as the number of layers and their respective settings, affects CNN architecture and model performance.

This investigation aimed to understand the correlation between these adjustments and the overall model performance. The goal was to determine the optimal configuration for our task. While adding more hidden layers can increase the model's capacity, it also introduces the potential for overfitting if not carefully controlled. The results of these adjustments on model performance are detailed in Table 8.

Table 6: Performance Comparison using different BERT embeddings

Model	**Opinion/Suggestion Classification**							**Urgency Classification**						
	Opinion			Non-Opinion			Weight F1 (%)	Urgent			Non-Urgent			Weight F1 (%)
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)		P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	
BERT-base-cased	90.4	85.9	88.1	79.8	84.6	82.1	85.3	81.1	74.5	77.6	78.5	87.1	82.5	79.9
BERT-base-uncased	91.1	86.4	89.8	80.1	84.2	82.5	**87.3**	82.5	74.2	78.4	80.2	86.7	83.5	**81.1**
BERT-large-cased	90.1	85.4	87.6	79.2	83.5	81.3	84.7	81	72.2	76.3	79.3	84.6	81.8	78.9
BERT-large-uncased	90.3	85.8	88	79.2	83.9	81.4	85	81.1	73.6	77.1	79.5	85.3	82.3	79.5

Table 7: Performance Comparison using balanced and imbalanced datasets

Model	**Opinion/Suggestion Classification**							**Urgency Classification**						
	Opinion			Non-Opinion			Weight F1 (%)	Urgent			Non-Urgent			Weight F1 (%)
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)		P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	
Imbalanced	90.3	85.9	88	80.8	83.4	82.1	85.2	83.1	73.3	77.8	80	84.9	82.3	79.9
Balanced	91.1	86.4	89.8	80.1	84.2	82.5	**87.3**	82.5	74.2	78.4	80.2	86.7	83.5	**81.1**

Table 8: Performance Comparison using different CNN parameters

CNN Kernel	**Number of CNN filters**	**Pool Size (Maxpooling)**	**Opinion/Suggestion Classification**							**Urgency Classification**						
			Non-Opinion			Opinion			Weight F1 (%)							Weight F1 (%)
			P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)		P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	
CNN – 2	64	3	89.5	84	87.8	78.1	82.7	80.8	85.4	81.1	72.1	76.8	78.1	84.6	81.4	79.2
CNN – 3	64	3	90.2	84.4	88.4	78.6	84	81.7	86.1	81.4	72.6	77	78.8	85.1	82	79.6
CNN – 4	64	3	90.1	84.8	88.6	79.3	83.3	81.7	86.2	81.8	73.1	77.5	79.7	85.8	82.8	803
CNN – 5	64	3	90	84.5	88.3	79.1	83.1	81.5	86	81.2	72.8	77	79.6	85.4	82.5	79.9
CNN – 2	128	3	90.6	85.3	89	79.5	83.4	81.9	86.5	82	73.5	77.8	80.4	86.1	83.3	80.7
CNN – 3	128	3	91.1	86.4	89.8	80.1	84.2	82.5	**87.3**	82.5	74.2	78.4	80.2	86.7	83.7	**81.1**
CNN – 4	128	3	90.8	86.2	89.6	79.3	84.3	82.2	87	82.1	73.7	77.9	79.8	86.2	83	80.6
CNN – 5	128	3	90.3	86	89.2	79.7	83.6	82.1	86.8	81.8	73.1	77.5	79.3	86.1	82.7	80.2
CNN – 2	256	3	90.1	85.8	89	78.9	82.7	81.2	86.3	81.1	72.9	77	78.8	85.8	82.3	79.8
CNN – 3	256	3	90.3	85.5	89	80	83.3	82.1	86.6	81.3	73.2	77.3	79.1	85.4	82.3	79.9
CNN – 4	256	3	90	85.6	88.9	79.2	83.1	81.6	86.3	79.5	72.6	76.1	77.8	86.1	81.9	79.1
CNN – 5	256	3	89.7	85.1	88.5	78.3	83.2	81.1	85.9	79.9	72.7	76.4	77.5	86.2	81.8	79.2
CNN – 3 – 4	64 – 64	6 – 3	88.8	84.7	87.9	77.8	83.2	80.9	85.5	78.7	72.1	75.5	78.3	85.5	81.9	78.8
CNN – 3 – 5	64 – 64	6 – 3	87.3	85.9	87.7	78.2	82.9	80.9	85.4	79.2	72.3	75.8	78.1	85.3	81.7	78.9
CNN – 3 – 6	64 – 64	6 – 3	88.1	85.4	87.9	78	83.3	81	85.5	78.4	73.5	76.1	78.7	85.1	81.9	79.1
CNN – 3 – 7	64 – 64	6 – 3	87.9	86.7	88.4	78.9	82.6	81.2	85.9	78.7	72.8	75.7	77.4	85	81.2	78.5
CNN – 3 – 4	128 – 128	6 – 3	88	86.3	88.3	79.4	82.7	81.5	86	79.3	73.1	76.3	77.2	85.8	81.4	79
CNN – 3 – 5	128 – 128	6 – 3	89.8	86.7	89.4	79.8	83.6	82.1	86.8	79.4	73.3	76.5	77.8	86.1	81.9	79.3
CNN – 3 – 6	128 – 128	6 – 3	89.2	85.4	88.4	78.8	83.1	81.3	86	78.1	72.8	75.6	78.2	86.2	82.2	79
CNN – 3 – 7	128 – 128	6 – 3	88.9	84.1	87.6	79.3	82.2	81.2	85.5	78.4	72.4	75.5	78.3	85.1	81.7	78.7
CNN – 3 – 4	256 – 256	6 – 3	89.3	83.8	87.6	79.2	83.7	81.8	85.8	79.2	73.9	76.7	77	84.9	80.9	79
CNN – 5 – 5	256 – 256	6 – 3	89.1	85.6	88.5	79.8	83.8	82.2	86.4	79.3	73.8	76.7	78.3	85.2	81.8	79.4
CNN – 3 – 6	256 – 256	6 – 3	88.7	84.2	87.5	78.5	82.3	80.8	85.3	78.4	72.1	75.4	77.1	84.4	80.7	78.2
CNN – 3 – 7	256 – 256	6 – 3	88.2	83.7	87	77.9	82	80.3	84.8	77.9	72.3	75.2	77.6	84.1	80.9	78.2

Conclusion

In online education domain, particularly in the context of MOOCs, the task of capturing and comprehending student opinions and suggestions is very important.

Students share their course experiences, express their perspectives, and provide suggestions to enhance the course. Extracting student opinions and suggestions can help instructors, course designers, and policymakers enhance various aspects of the course and streamline the decision-making process.

In this paper, we introduced a state-of-the-art solution for extracting and classifying student opinions within MOOC discussion forums, with a specific focus on identifying the urgency of these opinions. We developed a BERT-based hybrid multi-output deep learning model, named CBiLSTM.

It performs two classification tasks: first it determines if a student's post contains an opinion and, if so, further classifies it based on its urgency. This research bridges the gap between conventional suggestion mining in commercial contexts and its application in the education domain and paves the way for effective decision-making processes.

The performance analysis of the proposed CBiLSTM model underscores its effectiveness. Through extensive evaluations, we have demonstrated that our model surpasses baseline models, achieving F1-weighted scores of 87.3% for opinion classification and 81.1% for urgency classification.

Additionally, it maintains high F1 scores of 89.8% and 78.4% for opinion and urgent classes, respectively. The Precision-Recall curves and AUC metrics further support the model's strength, highlighting its effectiveness in balancing precision and recall, especially in the context of imbalanced datasets.

While this study uses the Stanford MOOCPosts dataset, a widely accepted benchmark in educational sentiment analysis and suggestion mining [30], the dataset's diversity enables broad coverage of student expression patterns.

However, to further strengthen model generalizability, future work may explore applying the proposed model to additional datasets from other MOOC platforms such as Coursera or edX. Future research may explore fine-grained analysis of student opinions, allowing for detailed categorization into specific aspects like course content, assessments, or instructor performance. This precision enhances course improvement strategies. Additionally, incorporating sentiment analysis can provide insights into the emotional tone behind opinions, which enriches the context for decision-making in the online education domain.

Author Contributions

Mujtaba Sultani: Conception, Data Curation, Methodological Approach, Validation, Formal Analysis, Investigation, Drafting the Main Manuscript.

Negin Daneshpour: Conception, Methodological Approach, Manuscript Review & Editing.

Acknowledgment

The authors would like to thank Shahid Rajaee Teacher Training University (SRTTU) for their support and providing the opportunity to publish this research work.

Conflict of Interest

The authors state that there are no conflicts of interest related to this publication.

Funding

This research received no external funding.

Abbreviations

CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
Bi-LSTM	Bidirectional Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers
CBiLSTM	Convolutional Neural Network +
	Bidirectional Long Short-Term Memory
MOOCs	Massive Open Online Courses

References

[1] S. Chaturvedi, D. Goldwasser, H. Daumé III, "Predicting instructor's intervention in mooc forums," in Proc. 52nd Annual Meeting of the Association for Computational Linguistics, 1: Long Papers: 1501-1511, 2014.

[2] N. A. Paufler, E. F. Sloat, "Using standards to evaluate accountability policy in context: School administrator and teacher perceptions of a teacher evaluation system," Stud. Educ. Eval., 64, 100806, 2020.

[3] J. McDonald, A. C. M. Moskal, A. Goodchild, S. Stein, S. Terry, "Advancing text-analysis to tap into the student voice: A proof-of-concept study," Assess.Evalu. Higher Educ., 45, 154–164, 2020.

[4] P. Topali, A. Ortega-Arranz, A. Martínez-Monés, S. L. Villagrá-Sobrino, "Houston, we have a problem: Revealing MOOC practitioners' experiences regarding feedback provision to learners facing difficulties," Comput. Appl. Eng. Educ., 29(4): 769-785, 2020.

[5] R. F. Kizilcec, C. Piech, E. Schneider, "Deconstructing disengagement: analyzing learner subpopulations in massive

open online courses," in Proc. the 3rd International Conference on Learning Analytics and Knowledge: 170-179, 2013.

[6] W. Pena, A. Melgar, "Ontology-based information extraction from Spanish Forum," in Proc. Computational Collective Intelligence: 7th International Conference, ICCCI 2015, Part I: 351-360, 2015. Springer International Publishing.

[7] K. Stephens-Martinez, M. A. Hearst, A. Fox, "Monitoring moocs: which information sources do instructors value?" in Proc. the first ACM Conference on Learning@ Scale Conference, 79-88, 2014.

[8] C. Zhang, H. Chen, C. W. Phang, "Role of instructors' forum interactions with students in promoting MOOC continuance," J. Global Inf. Manage., 26(3): 105-120, 2018.

[9] Y. Feng, D. Chen, Z. Zhao, H. Chen, P. Xi, "The impact of students and TAs' participation on students' academic performance in MOOC," in Proc. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining: 1149-1154, 2015.

[10] O. Almatrafi, A. Johri, "Improving MOOCs using information from discussion forums: An opinion summarization and suggestion mining approach," IEEE Access, 10: 15565-15573, 2022.

[11] J. Ramanand, K. Bhavsar, N. Pedanekar, "Wishful thinking-finding suggestions and'buy'wishes from product reviews" in Proc. the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text: 54-61, 2010.

[12] K. Talebi, Z. Torabi, N. Daneshpour, "Ensemble models based on CNN and LSTM for dropout prediction in MOOC," Expert Syst. Appl., 235, 121187, 2024.

[13] S. Negi, K. Asooja, S. Mehrotra, P. Buitelaar, "A study of suggestions in opinionated texts and their automatic detection," in Proc. the Fifth Joint Conference on Lexical and Computational Semantics: 170-178, 2016.

[14] Y. Alotaibi, M. N. Malik, H. H. Khan, A. Batool, A. Alsufyani, S. Alghamdi, "Suggestion mining from opinionated text of big social media data," Comput. Mater. Continua, 68(3): 3323- 3338, 2021.

[15] C. Brun, C. Hagege, "Suggestion mining: Detecting suggestions for improvement in users' comments," Res. Comput. Sci., 70(1): 171-181, 2013.

[16] N. K. Laskari, S. K. Sanampudi, "Aspect-oriented suggestion mining from opinion reviews," J. Theor. Appl. Inf. Technol., 101(12): 4869-4883, 2023.

[17] M. Sultani, N. Daneshpour, "Extracting urgent questions from mooc discussions: A BERT-based multi-output classification approach," Arabian J. Sci. Eng., 1-22, 2024.

[18] A. Agrawal, A. Paepcke, "The stanford MOOC Posts dataset," Accessed: Dec, 15, 2020.

[19] S. X. Guo, X. Sun, S. X. Wang, Y. Gao, J. Feng, "Attention-based character-word hybrid neural networks with semantic and structural information for identifying of urgent posts in MOOC discussion forums," IEEE Access, 7: 120522-120532, 2019.

[20] S. Shaikh, S. M. Daudpota, A. S. Imran, Z. Kastrati, "Towards improved classification accuracy on highly imbalanced text dataset using deep neural language models," Appl. Sci., 11(2): 869, 2021.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., 16: 321-357, 2002.

[22] H. He, Y. Bai, E. A. Garcia, S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in Proc. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence: 1322-1328, 2008.

[23] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, A. Courville, "Augmented cyclegan: Learning many-to-many mappings from unpaired data," in Proc. International Conference on Machine Learning: 195-204, 2018.

[24] J. Devlin, M. W. Chang, K. Lee, "Google, KT, Language, AI: BERT: pre-training of deep bidirectional transformers for language understanding," in Proc. the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1: 4171-4186, 2019.

[25] Y. Luan, S. Lin, "Research on text classification based on CNN and LSTM," in Proc. 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA) : 352-355, 2019.

[26] J. Du, C. M. Vong, C. P. Chen, "Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification," IEEE Trans. Cybern., 51(3): 1586-1597, 2021.

[27] C. Olah, "Understanding lstm networks, August 2015,".

[28] G. Liu, J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," Neurocomputing, 337: 325-338, 2019.

[29] A. Agrawal, J. Venkatraman, S. Leonard, A. Paepcke, "YouEDU: addressing confusion in MOOC discussion forums by recommending instructional video clips," in Proc. International Conference on Educational Data Mining (EDM), 2015.

[30] N. A. Khodeir, "Bi-GRU urgent classification for MOOC discussion forums based on BERT," IEEE Access, 9: 58243-58255, 2021.

[31] M. J. Anzanello, F. S. Fogliatto, "Learning curve models and applications: Literature review and research directions," Int. J. Ind. Ergon., 41(5): 573-583, 2011.

[32] K. B. Haryono, H. C. Purnomo, R. Ferdinand, H. Lucky, D. Suhartono, "Investigating the influence of layers towards speed and accuracy of neural networks," in Proc. 2022 International Conference on Informatics Electrical and Electronics (ICIEE): 1-6, 2022.

[33] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," Nature, 521: 436-444, 2015.

Biographies

Mujtaba Sultani received his B.S. degree at Kandahar University, Afghanistan in 2019, and his M.S. degree at Shahid Rajaee Teacher Training University (SRTTU), Tehran in 2024. He is an Assistant Professor at Kabul Polytechnic University (KPU). His research interest primarily includes data mining, information and knowledge management, machine learning, natural language processing, sentiment analysis, text processing and language comprehension.

- Email: Mujtaba.cs01@kpu.edu.af
- ORCID: 0009-0009-4267-703X
- Web of Science Research ID: NA
- Scopus Author ID: NA
- Homepage: NA

Negin Daneshpour is an Associate Professor in the Faculty of Computer Engineering at Shahid Rajaee Teacher Training University Tehran, Iran. She received her B.Sc. degree in Computer Engineering-Hardware from the Department of Electronics and Computer Engineering at Shahid Beheshti University, Tehran, Iran, in 1999 and her M.Sc. and Ph.D. degrees in Computer Engineering-Software from Amirkabir University of Technology, Tehran, Iran, in 2002 and 2010, respectively. Her research interests include data mining, data preprocessing, information management and natural language processing.

- Email: ndaneshpour@sru.ac.ir
- ORCID: 0000-0003-3951-4060
- Web of Science Research ID: NA
- Scopus Author ID: NA
- Homepage: https://www.sru.ac.ir/daneshpour/