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

Research paper

A Novel Hybrid Genetic Algorithm to Predict Students' Academic Performance

Y. Rohani, Z. Torabi*, S. Kianian

Faculty of Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran.

Article Info

Article History:
Received 15 September 2019
Reviewed 16 November 2019
Revised 23 January 2020
Accepted 23 May 2020

Keywords:
Classification
Educational data mining
Simulated annealing algorithm
Genetic algorithm
Educational performance prediction

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

Abstract

Background: Prediction of students' academic performance is essential for systems emphasizing students' greater success. The results can largely lead to increase in the quality of the educating and learning. Through the application of data mining, useful and innovative patterns can be extracted from the educational data.

Methods: In this paper, a new metaheuristic algorithm, combination of simulated annealing and genetic algorithms, is proposed for predicting students' academic performance in educational data mining. Although metaheuristic algorithms are one of the best options for discovering the hidden relationships between data in data science, they do not separately perform well in accurate prediction of students' academic performance. Therefore, the proposed method integrates the advantages of both genetic and simulated annealing algorithms. The genetic algorithm is applied to explore new solutions, while simulated annealing is used to increase the exploitation power. By using this combination, the proposed algorithm has been able to predict the students' academic performance with high accuracy.

Results: The efficiency of the proposed algorithm is evaluated on five different educational data sets, including two data sets of students of Shahid Rajaee University of Tehran and three online educational data sets. Our experimental results show $1.09\% - 24.39\%$ and $0.29\% - 6.57\%$ accuracy improvement of the proposed algorithm in comparison to the four similar metaheuristic and five popular classification methods respectively.

Introduction

Educational data mining (EDM) refers to the application of data mining methods on data sets, obtained from educational centers or web sites, to extract useful and novel information from data. Early prediction of the students' academic performance at the end of a training course is one of the challenges in educational centers. Today the coronavirus has forced many educational centers to e-learning. It seems they continue online education even after the pandemic ends however some studies [1]-[5] demonstrate higher dropout rates in web-based courses than traditional education ones which increase the importance of EDM. Usage of EDM can prevent financial and psychological consequences which are caused by failure of students, therefore various algorithms and methods have been used for this purpose. Metaheuristic algorithms, which are widely used to solve optimization problems, also have been used by researchers to analyze data and discover hidden patterns in a large data set [6]-[11].

To solve optimization problems effectively, several metaheuristic methods have been proposed such as: particle swarm optimization algorithm [12], artificial bee colony algorithm [13], differential evolution algorithm

[14], firefly algorithm [15] and earthworm optimization algorithm [16]. In these population-based optimization algorithms, the two concepts of "Exploration" and "Exploitation" have received much attention. Exploration refers to the ability to search various unknown areas in the space of a solution in order to find the probable solutions, while exploitation refers to the ability of an algorithm to improve existent solutions, to increase their quality [17].

Genetic algorithm (GA) as a population-based metaheuristic algorithm has both exploration and exploitation strategy and has been widely used in solving optimization problems [8], [18], [19]. Crossover and mutation operators focus more on exploring the problem space, while selection operator focuses on exploiting existing solutions. Thus, in GA algorithm the power of exploration is high, but the power of exploitation is not. In this paper, we proposed a new GA algorithm using simulated annealing (SA) approach to achieve the optimal combination of exploration and exploitation. SA is a single-solution based algorithm that starts with a solution and tries to enhance it. In other words, SA has only exploitation strategy and could strengthen the power of the genetic algorithm. By using this combination, the proposed algorithm has been able to produce high-quality solutions. Therefore, the algorithm can predict students' academic performance more accurately in comparison with other well-known classification methods.

Based on the above explanations, the contributions of this paper could be summarized as follow:

- Proposing an efficient algorithm for classification problems in educational data sets, which is based on combination of GA and SA.
- Proposing a new mutation operator for GA.
- Implementing the proposed algorithm on different educational data sets which leads to the highest prediction accuracy in comparison with the best classification methods.

The rest of this paper is organized as follows. Some recent works on educational data mining and combining metaheuristic algorithms are explained briefly in related work section. In genetic algorithm and simulated annealing algorithm sections, the concept and definitions of the GA and SA algorithms are described respectively. The proposed method is discussed thoroughly in proposed method section. The implementation details and comparison of the proposed work and other common methods are provided in results and discussion section. Finally, conclusion of this paper is presented.

Related Work

In EDM, statistics, machine learning, and data mining techniques are utilized to analyze data collected during teaching and learning [20] in order to find appropriate solutions to educational research. Such solutions are used by both instructors and students to improve teaching and learning. Early prediction of students' learning outcomes is one of the main concerns in EDM which can be categorized in two classes [21]:

- Predicting students' score in a specific course or score point average (GPA).
- Predicting students' academic performance (pass/fail) or dropout.

Unlike some works such as prediction of students' engagement [22] or prediction of slow learners [23], most of the studies about prediction models fall into this category. Several methods have been recently utilized in the prediction of students' grade which is a continuous value [24], [25]. In addition, predicting students' academic performance (pass/fail) or dropout, has been receiving significant attention. In this case, the main goal is to construct a learning model that predicts whether a student will pass/fail or dropout/complete a course.

Students' academic performance or grade depends on several factors such as background characteristics, previous scores, teaching and learning approaches, relationships between student-student, student-teacher, and student-content [25], [26]. Researches in EDM propose prediction models through some of these factors. For example, [27] with consideration of background characteristics and teaching features, presented a study to investigate whether the performance of teachers can predict students' academic performance. Table 1 summarizes recent researches proposed to predict students' outcome. Nevertheless, some works from each category are briefly described here. For the prediction of final grades, the authors of [28] applied popular algorithms such as Ordinary Least Squares (OLS), Support Vector Machine (SVM), Classification and Regression Tree (CART), k-Nearest Neighbor (kNN), Random Forest (RF) and AdaBoost R2 where SVM reaches the best result. In similar work [29], Kostopoulos proposed a multi-scheme semi-supervised regression approach (MSSRA) using three different k-NN algorithms regressors. The prediction model is based on features such as background characteristics, academic performance and interactions within the learning platform where the results showed the superiority of the MSSRA in comparison with other regression methods.

To achieve accurate prediction of students' dropout, Mubarak et al. [30] extracted significant features from students' weekly interaction with course content. They presented two models based on Logistic Regression (LR) and Input-Output Hidden Markov Model (IOHMM), which have better prediction accuracy in comparison with baseline of machine learning models. Moreover, with offering instructors' intervention methods, they reduced the rate of dropout. In similar study [31], Burgos et al. conducted a study over the scores of 100 students for several distance learning courses where Logistic Regression models were used for students' dropout prediction. With the usage of this result, they designed a tutoring action plan reducing the dropout rate by 14%. With the aim of discovering patterns that motivate students to drop out, Sarrah et al. [25] developed the Bayesian Profile Regression (BPR) method to identify students who are more likely to drop out. Due to the performance, motivation and resilience of students, this technique draws the profile of students at high risk of academic failure.

In order to predict students' academic performance, Chui et al. [32] proposed a reduced training vector-based support vector machine (RTV-SVM) algorithm. In their research three classes are defined, namely, pass, marginal, and fail.

The authors showed that the RTV-SVM has reached accuracy of 91.2%. Moreover, in large database, the RTV-SVM can be adopted to reduce the training time. In another study [33], the authors addressed high students' failure rates in introductory programming courses. Therefore, several educational data mining techniques were used for prediction of students' academic failure on two data sets including personal and educational information about 262 students from distance education and 161 students from on-campus. They also analyzed the impact of preprocessing and fine-tuning of input parameters to increase the prediction accuracy where the results showed that SVM reaches the best performance. All these studies, manipulate data sets of students' learning behavior, activities, and interactions stored in files and databases. One of the main concerns of such works is acquiring the highest prediction accuracy. Nevertheless, they applied different methods to obtain notable results.

Table 1: Research on prediction of students' learning outcomes

Ref.	Methods	Aim	Type
[28]	OLS, SVM, CART, kNN, RF , AdaBoost R2	Prediction of students' grade	Regression course grades
[43]	GA, Quadratic Bayesian Classifier, kNN, Parzen-window, Multi-Layer Perceptron, Decision Tree.	Prediction of students' grade	Regression course grades
[29]	MSSRA	Prediction of students' grade	Regression course grades
[44]	J48, REPTree	Prediction of students' grade	Regression course grades
[32]	RTV-SVM	Prediction of students' academic performance	Multi-class classification
[25]	BPR	Prediction of students' academic performance	Binary classification
[45]	Gradient Boosting Machine	Prediction of students' academic performance	Binary classification
[33]	Naive Bayes, Decision Tree, SVM, Neural Network	Prediction of students' academic performance	Binary classification
[46]	Decision Tree, Naive Bayes, Neural Network, SVM,kNN	Prediction of students' academic performance	Binary classification
[47]	Regression, Decision Tree	Prediction of students' academic performance	Binary classification
[48]	Ensemble model of Decision Tree, Gradient Boost algorithm and Naive Bayes	Prediction of students' academic performance	Binary classification
[49]	Naive Bayes, J48, RF, Naive Bayes Multiple Nominal, K-star and IBk	Prediction of students' academic performance	Multi-class classification
[31]	LR	Prediction of students' dropout	Binary classification
[30]	LR, IOHMM	Prediction of students' dropout	Binary classification
[50]	JRip, OneR, PART and Ridor	Prediction of students' dropout	Binary classification
[51]	Improved Decision Tree algorithm based on ID3	Prediction of students' dropout	Binary classification
[52]	Multilayer Perception, Naive Bayes, SMO, J48, REPTree	Prediction of slow learners	Binary classification
[53]	Decision Tree, J48, Naive Bayes, CART, JRIP Decision Rules, Gradient Boosting Trees,	Prediction of students' engagement	Binary classification

In this paper, we focus on two popular metaheuristic algorithms, namely SA and GA to obtain an accurate binary classifier in EDM to predict the students' academic performance. In recent years, the combination of metaheuristic algorithms has been used by numerous researchers in the field of optimization [6]. Moreover, hybrid metaheuristic algorithms have shown superior performance in solving many practical or academic problems [7]. In the following, several works on application of hybrid metaheuristic algorithms in different problems are described.

SA is used in numerous hybrid metaheuristic algorithms. In [9], Martin and Otto introduced a hybrid algorithm between the Markov chain and SA, in which the Markov chain is allocated just to detect local optimizations. With combination of two Tabu search algorithms and SA, Lenin et al. [10] proposed a new way to solve the reaction power problem.

In [11], similar combination proposed for symmetric traveling salesman problem. Wang et al. [34] proposed a new hybrid SA for scheduling in dual-resource cellular manufacturing system. In [35], to achieve an automatic diabetic retinopathy screening system, a new hybrid algorithm was proposed by using SA and ensemble bagging classifier. For feature selection, different hybrid metaheuristic algorithms were developed such as combination of local search operations and GA [8], combination of whale algorithm and SA [36]. Several combinations of SA and GA have been proposed in the literature for optimization of signal timing, navigation and routing in the supply chain, thermal structure problem, and flow shop scheduling [37]-[42].

Genetic Algorithm

Genetic algorithm is a metaheuristic algorithm inspired by the principle of the natural selection and natural genetics [18]. GA represents the solutions in the form of chromosomes and the fitness of the chromosomes is evaluated by fitness function which is created according to the objective function of the optimization problem. A collection of chromosomes is called a population where initially a random population is created. Individual solutions are selected to be parents for generating a new population through their fitness values, where fitter solutions are typically more likely to be selected. A pair of parent solutions creates new children by crossover and mutation operators where new solutions typically share many features of their parents. Mutation operator is used to avoid getting stuck in the optimal local trap. The process of selection, crossover, and mutation operators improves the population and continues until a new population of appropriate size is generated. The best solution in the last population is returned as the best approximation of the global optimum. Compared to the traditional metaheuristic methods, GA is well converged due to the adaptation of the biological evolution model [54]. Algorithm 1 shows the pseudo code of GA.

Algorithm 1: Genetic Algorithm

```
begin
  Generate initial population
  repeat
     Evaluate the individual solutions
     Select pairs of best-ranking individuals
     Apply crossover operator
     Apply mutation operator
  until terminating condition is not met
end
```

Simulated Annealing Algorithm

SA proposed by Kirkpatrick et al. [55] is a metaheuristic method inspired by the annealing procedure used in metallurgy, suitable in solving complex optimization problems. SA starts with a randomly generated solution in high temperature. In each iteration, a neighbor solution is generated according to a predefined neighborhood structure and evaluated using a fitness function. If the new solution is better than the original one, it is accepted, otherwise it can still be accepted with probability $e^{-\frac{\theta}{k_B T}}$, where θ is the difference between the fitness of the current solution and the generated neighbor, k_B is Boltzmann's constant and T is the current temperature [36]. By accepting worse solutions, SA can avoid being trapped on local optimum. The parameter T is gradually decreased by a cooling function as SA proceeds until the termination condition is met [56]. Algorithm 2 shows the pseudo code of SA algorithm.

Algorithm 2: Simulated Annealing Algorithm

```
begin
  Initialize solution x, temperature T
  repeat
     repeat
        Create neighborhood solution y
        if θ ≤ 0 then accept y
        otherwise accept y by the probability e^(-θ/(k_B T))
     until inner-loop terminating condition is not met
     Decreasing T gradually
  until outer-loop terminating condition is not met
end
```

Proposed Method

In this paper, the G-SA algorithm, which is a new hybrid of GA and SA algorithms, is presented to predict students' academic performance. The framework of the G-SA algorithm is shown in Fig. 1. As the figure shows, an initial population of states is generated first. After generating the initial population, one randomly chosen state, namely current state, is utilized to crossover with the best state of the population (i.e. Gbest) where offspring1 and offspring2 are produced. On the other hand, an offspring3 is created by the mutation operation. The best of these three offsprings is compared to the current state. If the best offspring is better, it would replace the current state. Otherwise it would replace with probability $e^{-\frac{\theta}{k_B T}}$. This process continues until the new population would be completed with n new states where n is the size of population. After that, based on the current temperature it is decided to generate a new population or stop generating new population. The best state of final population is added to the rule set and all the states from the training set that can be evaluated by this rule are removed from the training set.

The algorithm runs again with the remaining states of the training set to obtain another rule. This continues until the size of the training set reaches a certain threshold (10% of the training set). In the end, a set of rules is obtained that is used for prediction. Algorithm 3 shows how the G-SA works.

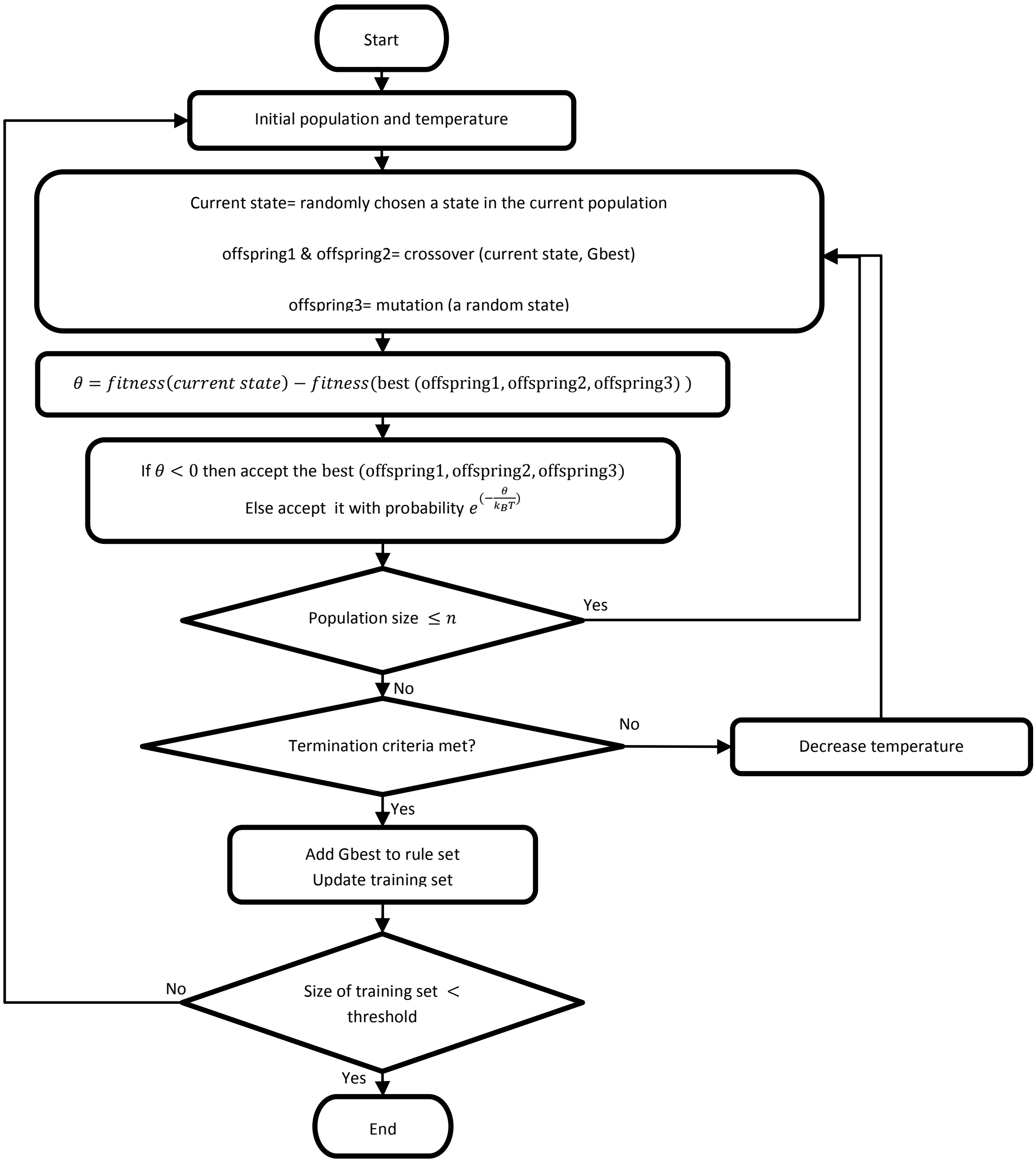

Fig. 1: Framework of the G-SA algorithm.

Algorithm 3: G-SA Algorithm

```
begin
 repeat
   Initialize first population and T
   Find the best state obtained so far, which has highest fitness value (i.e.Gbest)
   repeat
     repeat
       (Crossover)
         Select random state u as current state
         for each feature index j in u
           Select randomly bit RN between [0, 1]
           if RN > 0.6 then
             Offspring1_j = Gbest_j
             Offspring2_j = u_j
           else
             Offspring1_j = u_j
             Offspring2_j = Gbest_j
       (Mutation)
         Select random state v
         Offspring3 = v
         for each feature index j in v
           Select randomly bit RN between [0, 1]
           if RN <= 0.1 then
             Update Offspring3_j value using Equation 1
       new state = best (offspring1, offspring2, offspring3)
       θ = fitness(u) − fitness(new state)
       if θ ≤ 0 then Accept new state
       otherwise Accept it by the probability e^(−θ/(k_B T))
       (Update Gbest)
         Find the best state obtained so far, which has highest fitness value
     until population size reaches n
     Decrease T gradually
   until outer-loop terminating condition is not met
   Add Gbest as a rule to rule set
   Update training set
 until the size of the training set reaches threshold
end
```

A. Problem formulation and initial population

The aim of the G-SA algorithm is to predict the students' academic performance (pass/fail) using some background characteristics and previous scores.

Let $X = [x_1, x_2, \dots, x_n]$ be pattern of the initial population where x_i is a chromosome that represents a candidate solution which is modeled as an array of genes. In educational data sets, genes of a chromosome are dependent on a given data set and may contain background characteristics, previous scores, teaching and learning approaches, relationships between student-student, student-teacher, and student-content. Last gene in each chromosome indicates academic performance. Operators of GA in the proposed method are applied to these chromosomes. Here, since the proposed algorithm is a combination of SA and GA, these chromosomes are called states.

For more clarification, Fig. 2 shows an example of a state in our problem modeling for Math dataset explained in Section 6-A. In this figure, the first 10 features are background characteristics of student such as age, sex, parents' jobs and educations. Next 3 features are scores of three exams and last feature represents the success or failure of student and must be predicted.

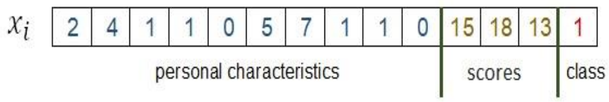

Fig. 2: An example of a state.

B. Crossover and Mutation

One of the important points that can directly affect the accuracy and convergence speed of the algorithm is the neighborhood selection strategy. In basic SA, some operators such as insertion, reversion and swap are used to produce a neighborhood state, which sometimes does not result in a global optimal solution. Crossover and mutation operators in GA can be good options for generating new neighborhood states. The quality of the new states generated by crossover and mutation operators depends on the fitness of the parents. In this study, the best state in the population named Gbest is used in crossover operator.

Gbest and a randomly chosen state (i.e. current state) are combined under crossover operation to generate offspring1 and offspring2. The G-SA uses the crossover operation [57] implemented by generating a vector of random numbers in the range [0, 1] having the same length as the parents. In each feature of a current state, if the value of the random vector is below 0.6, offspring1 takes this feature from Gbest, otherwise from a current state. If the value of the random vector is above 0.6, offspring2 takes this feature from a current state, otherwise from Gbest.

One-point or two-point uniform crossover operators are usually used in GA where parents are randomly selected based on their fitness from a large search space. The crossover operator of [57] by deciding for each feature independently led to accurate classifier [57]. In addition, one of the parents in the crossover operator of [57] is Gbest, increasing the fitness of the produced offsprings and convergence speed.

After performing the crossover and producing two offsprings, offspring3 is generated by the mutation operator. In mutation operator, another random state is selected and a random vector is generated in the range [0, 1] having the same length as the random state. If the value of each feature of random vector is less than or equal to 0.1 then the corresponding feature of random state would change according to (1), where i is a number of feature, $\emptyset_i$ is a random number from a normal distribution with mean 0 and standard deviation $sigma$. For each feature, $sigma$ is obtained according to (2), where $Varmax_i$ and $Varmin_i$ are the upper and lower bounds of the *i*-th feature respectively.

$$v_i = x_i + \emptyset_i * sigma_i \tag{1}$$

$$sigma_i = 0.1 * (Varmax_i - Varmin_i) \tag{2}$$

Fig. 3, shows an example of how a mutation operator works. As can be seen in this example, only the fourth feature of the random vector is less than 0.1. Therefore, by changing it in random state according to (1), a new offspring is produced that differs from the random state only in one feature.

In the conventional mutation operator, a new offspring is generated by changing only one feature of parent state which leads to the offspring different from the parent in only one feature. To avoid getting stuck in local optimum, the proposed mutation operator in this study, can change more than one features increasing the mutation power. Therefore, the proposed mutation operator by creating offspring with more distance from the parent state, prevents the algorithm trapping into local optimum.

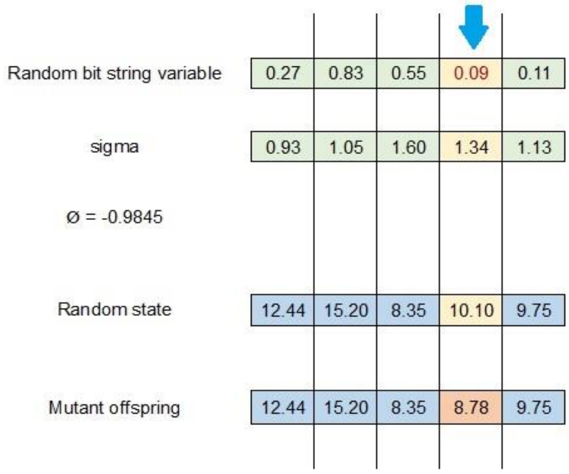

Fig. 3: Mutation operator in G-SA algorithm.

C. *Fitness function*

In this paper, (3), is used to evaluate the states in population. Each state in the training set is considered as a rule for prediction. All the states in the training set are compared to this rule, and the values of the two TP and TN criteria are specified.

$$f(x_i) = \frac{TP + TN}{size(Training\ Set)} \tag{3}$$

The TP criterion contains the number of samples that are correctly predicted as positive and TN contains the number of samples that are correctly predicted as negative.

Results and Discussion

To indicate the superiority of the G-SA algorithm, this section presents and analyzes the results of the proposed algorithm and other state-of-the-art classification methods. The G-SA algorithm has been compared to five well-known classification algorithms in data mining, including Decision Tree J48, Naive Bayes (NB), Multilayer Perceptron (MP), LR, and SVM. Furthermore, it has been compared to the four following metaheuristic methods.

- Basic ABC algorithm [58].
- GBC algorithm [57].
- Basic SA algorithm [55].
- SA-GA algorithm [41].

In the following, datasets which are used for evaluation are described and then parameters of algorithms are estimated. Finally, the G-SA algorithm results are compared with other classification methods.

A. *Data Sets*

In order to evaluate the proposed algorithm for prediction of student performance, five different educational data sets including S5f, E-circ, Math, Port, and Deeds have been used.

The detailed descriptions of these five data sets are presented in Table 2 with respect to the number of classes, samples, and features.

S5f and E-circ data sets have been obtained from the database of computer engineering students at Shahid Rajaee University of Tehran during the years of 2011-14. S5f data set includes the students' average grades in the first five semesters.

The purpose is to predict the performance of students in the fifth semester using the average grades of the first four semesters. E-circ data set is compiled to predict students' performance in Electric Circuits course which is one of the most challenging courses using the scores of Discrete Structures, Physics, Mathematics, Differential Equations and Logic Circuits.

The other three data sets were compiled by researchers in the field of educational data mining to predict the educational academic status of students and can be accessed online on the UCI[1] site. Math and Port data sets related to the students of Portuguese school, whose data were gained in two courses of Mathematics and Portuguese Language respectively and have 33 identical features [59].

Since the large number of features increases the execution time of the algorithm greatly, first feature selection operation was applied to these two data sets. For this purpose, with the Information Gain method in WEKA, data mining tool, these data sets were analyzed and among the 33 features, top 10 features were selected for each of the Math and Port data sets.

Deeds data set is related to Logic Circuits course at the University of Genoa, Italy, and includes scores of 16 tests out of 100 samples, with the final test score being considered as a class [60]. In Deeds data set, among its features, six characteristics were selected to predict the class.

Table 2: Statistics of educational data sets

Data sets	Number of features	number of samples	Class Fail	Class Pass
S5f	5	128	10	118
E-circ	8	255	105	150
Math	11	395	130	265
Port	10	649	100	549
Deeds	6	114	63	51

[1] https://archive.ics.uci.edu/ml/index.php

The features of S5f, E-circ and Deeds data sets are some exam scores in different courses. In the Math and Port data sets, in addition to the three test scores, they also include some personal characteristics of students. Table 3 shows the features of the Math and Port data sets. Some of these features are used in the math and some in the port data set.

Table 3: Features of the math and Port database

Attribute	Description (Domain)
age	student's age (numeric: from 15 to 22)
school	student's school (binary: Gabriel Pereira or Mousinho da Silveira)
Pstatus	parent's cohabitation status (binary: living together or apart)
Medu	mother's education (numeric: from 0 to 4)
famsize	family size (binary: < 3 or > 3)
famrel	quality of family relationships (numeric: from 1 : very bad to 5 : excellent)
reason	reason to choose this school (nominal: close to home, school reputation, course preference or other)
studytime	weekly study time (numeric: 1 : < 2 hours, 2 : 2 to 5 hours, 3 : 5 to 10 hours or 4 : > 10 hours)
failures	number of past class failures (numeric: n if 1 : n < 3, else 4)
schoolsup	extra educational school support (binary: yes or no)
famsup	family educational support (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
freetime	free time after school (numeric: from 1 : very low to 5 : very high)
goout	going out with friends (numeric: from 1 : very low to 5 : very high)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (Class – Fail or Pass)

B. *Parameter settings for the algorithms*

In this paper, MATLAB has been used for implementation of the algorithms with 10-fold cross validation method for all of the algorithms. To attain a fair comparison, we ran each algorithm with different parameters several times to find appropriate initialization values for the best results.

In ABC-based algorithms (i.e. ABC and GBC) the best results in five data sets are obtained by setting the maximum number of iterations to 400 and the colony size to 40 and both the number of food source and the rate of food source abandonment to 20.

In the implementation of SA-based algorithms (i.e. SA and SA-GA) the iteration of the main and sub loop set to 1000 and 10 respectively. Also, the initial population, temperature and the rate of temperature reduction, set to 20, 10 and 0.95 respectively. Due to the random nature of the initial population selection, each of the algorithms runs 50 times, and the average of accuracy was considered as the accuracy of the algorithm.

Population size of the G-SA algorithm set to 1000. Accuracy of the proposed method with different values for temperature reduction $(Alpha)$ and initial temperature (T_0) on the educational data sets are shown in Table 4. As can be seen, in the four data sets, the G-SA algorithm with $T_0 = 10$ and $Alpha = 0.95$ has achieved the highest accuracy, so it is considered as the initial settings for G-SA algorithm.

Table 4: The classification accuracy performance of the G-SA algorithm with different T_0 and $Alpha$

parameters	S5f	E-circ	Port	Math	Deeds
T_0=10 ; *Alpha*=0.99	95.62	79.43	94.13	**93.60**	93.62
Alpha=0.95	**96.30**	**80.29**	**94.84**	93.46	**94.49**
Alpha=0.9	95.18	79.22	94.35	93.32	93.21
Alpha=0.8	95.11	78.17	93.75	92.77	92.31
Alpha=0.75	94.43	78.38	93.63	92.41	93.15
T_0=20 ; *Alpha*=0.99	94.26	78.35	93.56	93.29	92.99
Alpha=0.95	95.33	79.91	93.80	93.14	93.82
Alpha=0.9	93.85	77.69	92.62	93.10	91.53
Alpha=0.8	94.27	77.33	93.32	91.33	90.23
Alpha=0.75	92.51	76.79	92.17	92.11	89.22

C. Experimental results of comparisons with other classification methods

The evaluation criterion for comparing the G-SA and other algorithms are accuracy and F-measure. Accuracy is computed by dividing the set of correct predictions by the sum of all predictions according to the (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

The variables used in the Equation 4 are as follow:

TP: The number of true positive samples that are correctly predicted as positive.

TN: The number of true negative samples that are correctly predicted as negative.

FP: The number of true negative samples that are incorrectly predicted as positive.

FN: The number of true positive samples that are incorrectly predicted as negative.

In F-measure criterion, the incorrect prediction rate along with the correct prediction rate, are used to evaluate the efficiency of the algorithms. Equation (5) shows the F-measure criterion, in which the two criteria Precision and Recall are obtained according to (6) and (7), respectively.

$$F-measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Table 5 shows the accuracy of the G-SA algorithm along with other relevant metaheuristic algorithms.

The numbers marked in blue are the highest accuracy in each data set related to the G-SA method.

The G-SA algorithm has improved accuracy compared to the basic ABC algorithm in the data set of Math and Deeds by 19.01% and 24.39%, respectively.

The minimum improvement in accuracy by the proposed method compared to the basic ABC algorithm has occurred in the data set of the Port which is 4.69%.

Maximum and minimum accuracy improvements compared to the GBC are 19.69% in the data set of E-circ and 5.15% in the data set of Port.

In addition, the maximum accuracy improvement of the G-SA algorithm compared to the SA and SA-GA methods are related to the data set of Deeds, which are 6.61% and 3.15%, respectively and the minimum accuracy improvement over the SA and SA-GA methods are 1.62% and 1.09% in the data set of Port.

Table 5: Accuracy performance of the proposed method, ABC, GBC, SA, and SA-GA

Data Sets	ABC	GBC	SA	SA-GA	G-SA
S5f	91.33	86.05	93.81	94.22	**96.30**
E-circ	61.28	60.60	77.13	78.48	**80.29**
Port	90.15	89.69	93.22	93.75	**94.84**
Math	76.75	77.34	89.00	92.10	**93.52**
Deeds	70.10	74.39	87.88	91.34	**94.49**

The accuracy of the G-SA algorithm, along with other popular classification methods, can be seen in Table 6. The proposed algorithm has the highest accuracy which marked in blue in each data set. The G-SA algorithm in different data sets, compared to the J48, has improved the accuracy from 1.12% to 6.57%. The highest and lowest accuracy improvements of the G-SA algorithm in comparison with NB are 5.68% and 0.64% in the Math and Deeds data sets respectively. The maximum accuracy improvement compared to SVM and MP are 4.66% and 5.68% in Math data set. The minimum accuracy improvement of the G-SA in Table 6 is 0.29% and related to LR in the data set of E-circ.

Table 6: Accuracy performance of the G-SA and five well-known classification methods

Data Sets	J48	NB	LR	SVM	MP	G-SA
S5f	94.53	93.75	94.53	92.97	93.75	**96.30**
E-circ	73.72	75.68	80.00	79.60	75.29	**80.29**
Port	93.62	91.83	92.91	91.52	90.91	**94.84**
Math	92.40	87.84	91.49	88.86	87.84	**93.52**
Deeds	88.59	93.85	92.98	93.85	93.85	**94.49**

The results of the proposed method with F-measure criterion in comparison with metaheuristic algorithms and conventional classification methods are shown in Tables 7 and 8, respectively. As can be seen, the results of the F-measure criterion also show better performance in comparison with other methods.

For better representation, Fig. 4, and Fig. 5, show the accuracy and F-measure results of the G-SA algorithm respectively, along with relevant metaheuristic algorithms and conventional classification methods. The results show that combination of two metaheuristic algorithms in G-SA leads to more accurate results.

Table 7: F-measure criterion of the proposed method, ABC, GBC, SA, and SA-GA

Data Sets	ABC	GBC	SA	SA-GA	G-SA
S5f	0.92	0.88	0.93	0.95	**0.96**
E-circ	0.63	0.61	0.77	0.78	**0.81**
Port	0.90	0.90	0.93	0.94	**0.97**
Math	0.78	0.79	0.90	0.92	**0.94**
Deeds	0.71	0.74	0.88	0.93	**0.95**

It can be noted that in most cases, the accuracy of each of the metaheuristic algorithms (i.e. ABC, GBC, SA, and SA-GA) is lower than the conventional classification methods, however, in G-SA algorithm after combining SA and GA properly, accuracy improved significantly.

Table 8: F-measure criterion of the proposed method and five well-known methods in classification

Data Sets	J48	NB	LR	SVM	MP	G-SA
S5f	0.94	0.93	0.94	0.90	0.93	**0.96**
E-circ	0.73	0.76	0.80	0.79	0.75	**0.81**
Port	0.93	0.92	0.92	0.90	0.91	**0.97**
Math	0.92	0.88	0.91	0.88	0.87	**0.94**
Deeds	0.89	0.94	0.93	0.94	0.94	**0.95**

Experimental results confirm the good balance between the power of exploration and exploitation of the proposed algorithm.

The crossover operator in the proposed algorithm utilized from the best global state generates high fitness states.

Such crossover operator sometimes may cause to trap the algorithm in local optimum.

The proposed method by introducing the new mutation operator prevents trapping in local optimum.

In the new mutation operator, instead of changing only one feature, more than one features may be changed, which increases the mutation power.

Due to the methaheuristic strategy and random nature of the initial population selection from the training set, increasing the volume of data not only would not limit the algorithm, but also can probably increase the accuracy as can be seen in the experimental results.

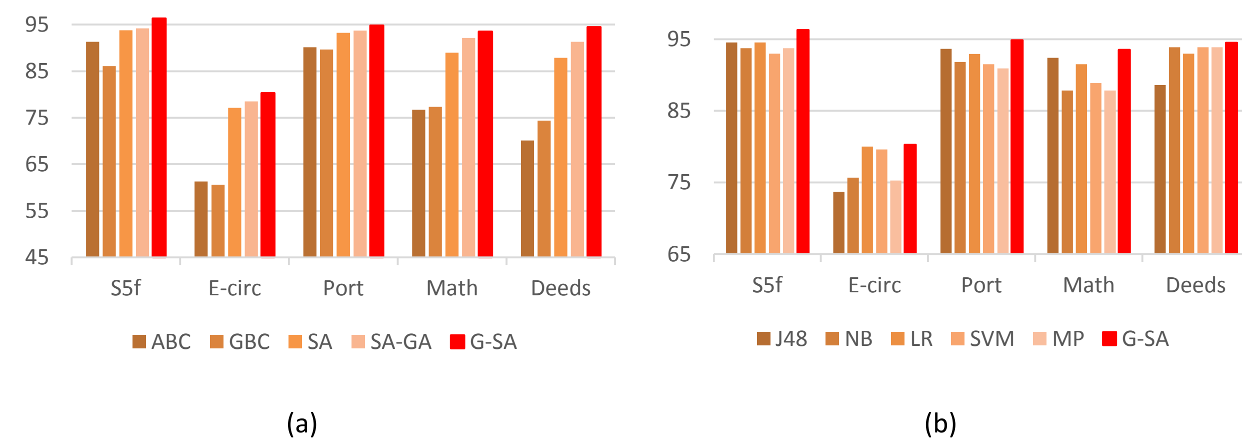

Fig. 4: Diagram of the accuracy performance of the proposed method and nine other classifiers.

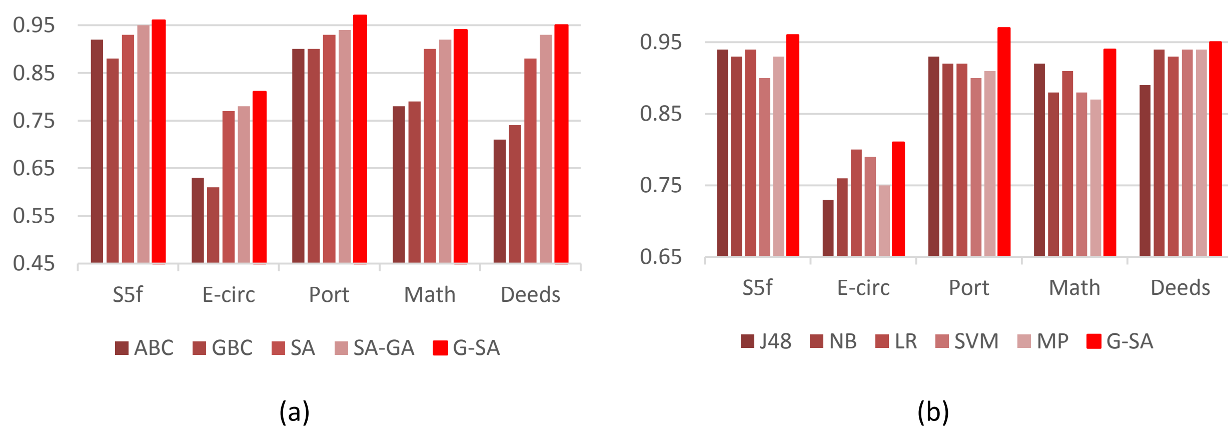

Fig. 5: Diagram of the F-measure criterion of the proposed method and nine other classifiers.

Conclusion

In this paper, a new method called G-SA, is presented for prediction of students' academic performance during educational courses, which can be used to prevent possible failures of students. The G-SA algorithm uses the advantages of both simulated annealing and genetic algorithms by combining them. By relying on the best global solution in the proposed algorithm, crossover and mutation operators produce stronger neighbors and ultimately lead to better solutions. The combination of the two algorithms also balances the power of exploration and exploitation in the proposed algorithm, which has not only helped to speed convergence, but has also been able to get rid of local optimum. Experimental results from the implementation of the G-SA algorithm show that the proposed algorithm improves accuracy performance from 1.09% to 24.39% compared to other metaheuristic comparison methods and from 0.29% to 6.57% compared to well-known conventional classification methods.

Author Contributions

This paper is the result of Y. Rohani's MSc thesis supervised by Z. Torabi, and S. Kianian.

Conflict of Interest

The authors declare that there is no conflict of interests 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

GA	Genetic Algorithm
SA	Simulated Annealing
G-SA	Genetic Simulated Annealing Algorithm
Gbest	Global best states
J48	Decision Tree J48
NB	Naive Bayes
LR	Logistic Regression
SVM	Support Vector Machine
MP	Multilayer Perceptron
n	Size of Population
k_B	Boltzmann's Constant
T	Current Temperature
S5f	Student performance data set in the fifth semester
E-circ	Student performance data set in Electric Circuits course
Math	Student performance data set in Mathematics course
Port	Student performance data set in Portuguese language course
Deeds	Student performance data set in Digital Electronics Education and Design Suite

References

[1] E. Black, K. Dawson, J. Priem, "Data for free: Using LMS activity logs to measure community in online courses," The Internet and Higher Education, 11(2): 65-70, 2008.

[2] C. Chen, Y. Chen, C. Liu, "Learning performance assessment approach using web-based learning portfolios for e-learning systems," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37(6): 1349-1359, 2007.

[3] J. Hung, K. Zhang, "Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching," MERLOT Journal of Online Learning and Teaching, 4(4): 426-437, 2008.

[4] M. Roblyer, L. Davis, S. Mills, J. Marshall, L. Pape, "Toward practical procedures for predicting and promoting success in virtual school students," The Amer. Jrnl. of Distance Education, 22(2): 90-109, 2008.

[5] A. Juan, T. Daradoumis, J. Faulin, F. Xhafa, "Developing an information system for monitoring student's activity in online collaborative learning," in Proc. International Conference on Complex, Intelligent and Software Intensive Systems: 270-275, 2008.

[6] E. Talbi, Metaheuristics: from design to implementation. John Wiley & Sons; 2009.

[7] E. Talbi, "A taxonomy of hybrid metaheuristics," Journal of heuristics, 8(5): 541-564, 2002.

[8] I. Oh, J. Lee, B. Moon, "Hybrid genetic algorithms for feature selection," IEEE Transactions on pattern analysis and machine intelligence, 26(11): 1424-37, 2004.

[9] O. Martin, S. Otto, "Combining simulated annealing with local search heuristics. Annals of Operations Research," 63(1): 57-75, 1996.

[10] K. Lenin, B. Reddy, M. Suryakalavathi, "Hybrid Tabu search-simulated annealing method to solve optimal reactive power problem," International Journal of Electrical Power & Energy Systems, 82: 87-91, 2016.

[11] Y. Lin, Z. Bian, X. Liu, "Developing a dynamic neighborhood structure for an adaptive hybrid simulated annealing–tabu search algorithm to solve the symmetrical traveling salesman problem," Applied Soft Computing, 49: 937-52, 2016.

[12] J. Kennedy, R. Eberhart, "Particle swarm optimization," in Proc. of ICNN'95-International Conference on Neural Networks, 4: 1942-1948, 1995.

[13] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical report-tr06, Erciyes university, engineering faculty, computer engineering department; 2005.

[14] Z. Meng, J. Pan, "HARD-DE: Hierarchical archive-based mutation strategy with depth information of evolution for the enhancement of differential evolution on numerical optimization, IEEE Access, 7: 12832-54, 2019.

[15] H. Wang, W. Wang, H. Sun, S. Rahnamayan, "Firefly algorithm with random attraction," International Journal of Bio-Inspired Computation, 8(1): 33-41, 2016.

[16] G. Wang, S. Deb, L. Coelho, "Earthworm optimisation algorithm: a bio-inspired metaheuristic algorithm for global optimisation problems," International Journal of Bio-Inspired Computation, 12(1): 1-22, 2018.

[17] A. Singh, K. Deep, "Exploration–exploitation balance in Artificial Bee Colony algorithm: a critical analysis," Soft Computing, 23(19): 9525-9536, 2019.

[18] H. Braun, "On solving travelling salesman problems by genetic algorithms," in Proc. International Conference on Parallel Problem Solving from Nature: 129-133, 1990.

[19] Y. Deng, Y. Liu, D. Zhou, "An improved genetic algorithm with initial population strategy for symmetric TSP," Mathematical Problems in Engineering, 2015: 1-7, 2015.

[20] C. Romero, S. Ventura, "Data mining in education," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1): 12-27, 2013.

[21] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, O. Ragos, "Implementing AutoML in educational data mining for prediction tasks," Applied Sciences, 10(1): 90, 2020.

[22] T. Soffer, A. Cohen, "Students' engagement characteristics predict success and completion of online courses," Journal of Computer Assisted Learning, 35(3): 378-89, 2019.

[23] A. Rajeswari, C. Deisy, "Fuzzy logic based associative classifier for slow learners prediction," Journal of Intelligent & Fuzzy Systems, 36(3): 2691-704, 2019.

[24] R. Morsomme, E. Smirnov, "Conformal Prediction for Students' Grades in a Course Recommender System," In Conformal and Probabilistic Prediction and Applications, 2019: 196-213, 2019.

[25] A. Sarra, L. Fontanella , S. Di Zio, "Identifying students at risk of academic failure within the educational data mining framework," Social Indicators Research, 146(1-2): 41-60, 2019.

[26] A. Agudo-Peregrina, S. Iglesias-Pradas, M. Conde-González, A. Hernández-García, "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and

online learning," Computers in human behavior, 31: 542-50, 2014.

[27] C. Herodotou, B. Rienties, A. Boroowa, Z. Zdrahal, M. Hlosta, "A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective," Educational Technology Research and Development, 67(5): 1273-306, 2019.

[28] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," International Educational Data Mining Society, 2015: 392- 395, 2015.

[29] G. Kostopoulos, S. Kotsiantis, N. Fazakis, G. Koutsonikos, C. Pierrakeas, "A Semi-Supervised Regression Algorithm for Grade Prediction of Students in Distance Learning Courses," International Journal on Artificial Intelligence Tools, 28(04): 1940001, 2019.

[30] A. Mubarak, H. Cao, W. Zhang, "Prediction of students' early dropout based on their interaction logs in online learning environment," Interactive Learning Environments, 20: 1-20, 2020.

[31] C. Burgos, M. Campanario, D. de la Peña, J. Lara, D. Lizcano, M. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," Computers & Electrical Engineering, 66: 541-56, 2018.

[32] K. Chui, D. Fung, M. Lytras, M. Lam, "Predicting at-risk university students in a virtual learning environment via a machine learning algorithm," Computers in Human Behavior, 107: 105584, 2020.

[33] E. Costa, B. Fonseca, M. Santana, E. de Araújo, J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," Computers in Human Behavior, 73: 247-56, 2017.

[34] J. Wang, C. Liu, K. Li, "A hybrid simulated annealing for scheduling in dual-resource cellular manufacturing system considering worker movement," Automatika, 60(2): 172-80, 2019.

[35] S. Sreng, N. Maneerat, K. Hamamoto, R. Panjaphongse, "Automated diabetic retinopathy screening system using hybrid simulated annealing and ensemble bagging classifier," Applied Sciences, 8(7): 1198, 2018.

[36] M. Mafarja, S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," Neurocomputing, 260: 302-12, 2017.

[37] P. Vasant P, "Hybrid simulated annealing and genetic algorithms for industrial production management problems," International Journal of Computational Methods, 7(02): 279-97, 2010.

[38] Z. Li, P. Schonfeld, "Hybrid simulated annealing and genetic algorithm for optimizing arterial signal timings under oversaturated traffic conditions," Journal of advanced transportation, 49(1): 153-70, 2015.

[39] Y. Li, H. Guo, L. Wang, J. Fu, "A hybrid genetic-simulated annealing algorithm for the location-inventory-routing problem considering returns under E-supply chain environment," The Scientific World Journal, 2013: 1-11, 2013.

[40] L. Junghans, N. Darde, "Hybrid single objective genetic algorithm coupled with the simulated annealing optimization method for building optimization," Energy and Buildings, 86: 651-62, 2015.

[41] H. Wei, S. Li, H. Jiang, J. Hu, J. Hu, "Hybrid genetic simulated annealing algorithm for improved flow shop scheduling with make span criterion," Applied Sciences, 8(12): 2621, 2018.

[42] F. Erchiqui, "Application of genetic and simulated annealing algorithms for optimization of infrared heating stage in thermoforming process," Applied Thermal Engineering, 128: 1263-72, 2018.

[43] B. Minaei-Bidgoli, W. Punch, "Using genetic algorithms for data mining optimization in an educational web-based system," in Proc. Genetic and evolutionary computation conference 2003: 2252-2263, 2003.

[44] S. Natek, M. Zwilling, "Student data mining solution–knowledge management system related to higher education institutions," Expert systems with applications, 41(14): 6400-6407, 2014.

[45] I. papadogiannis, V. Poulopoulos, M. Wallace, "A Critical Review of Data Mining for Education: What has been done, what has been learnt and what remains to be seen," International Journal of Educational Research Review, 5(4):353-372, 2020.

[46] A. RİMİ, A. IBRAHİM, O. BAYAT, "Developing Classifier for the Prediction of Students' Performance Using Data Mining Classification Techniques," AURUM Mühendislik Sistemleri ve Mimarlık Dergisi, 4(1):73-91, 2020.

[47] H. Hassan, R. Mohamad, R. Ali, Y. Talib, H. Hsbollah, "Factors Affecting Students' Academic Performance in Higher Education: Evidence from Accountancy Degree Program," International Business Education Journal, 13(1): 1-6, 2020.

[48] P. Kamal, S. Ahuja, "An ensemble-based model for prediction of academic performance of students in undergrad professional course," Journal of Engineering, Design and Technology, 2019.

[49] B. Kapur, N. Ahluwalia, R. Sathyaraj, "Comparative study on marks prediction using data mining and classification algorithms," International Journal of Advanced Research in Computer Science, 8(3): 1-5, 2017.

[50] C. Siebra, R. Santos, N. Lino, "A Self-Adjusting Approach for Temporal Dropout Prediction of E-Learning Students," International Journal of Distance Education Technologies (IJDET), 18(2):19-33, 2020.

[51] S. Sivakumar, S. Venkataraman, R. Selvaraj, "Predictive modeling of student dropout indicators in educational data mining using improved decision tree," Indian Journal of Science and Technology, 9(4): 1-5, 2016.

[52] P. Kaur, M. Singh, G. Josan, "Classification and prediction-based data mining algorithms to predict slow learners in education sector," Procedia Computer Science, 57: 500-8, 2015.

[53] M. Hussain, W. Zhu, W. Zhang, S. Abidi, "Student engagement predictions in an e-learning system and their impact on student course assessment scores," Computational intelligence and neuroscience, 2018..

[54] J. McCall, "Genetic algorithms for modelling and optimisation. Journal of computational and Applied Mathematics," 184(1): 205-22, 2005.

[55] S. Kirkpatrick, C. Gelatt, M. Vecchi, "Optimization by simulated annealing," science, 220(4598): 671-680, 1983.

[56] T. Wu, C. Chang, S. Chung, "A simulated annealing algorithm for manufacturing cell formation problems," Expert Systems with Applications, 34(3): 1609-1617, 2008.

[57] H. Alshamlan, G. Badr, Y. Alohali, "Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer

classification," Computational biology and chemistry, 56: 49-60, 2015.

[58] D. Karaboga, B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," Applied soft computing, 8(1): 687-97, 2008.

[59] P. Cortez, A. Silva. Using data mining to predict secondary school student performance, in Proc. 5th Annual Future Business Technology Conference: 5-12, 2008.

[60] M. Vahdat, L. Oneto, D. Anguita, M. Funk, M. Rauterberg, "learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator," in Proc. Design for teaching and learning in a networked world 2015: 352-366, 2015.

Biographies

Yaser Rohani received his BSc in 2002 from the Islamic Azad University of Mashhad, Iran and his MSc in 2020 from Shahid Rajaee Teacher Training University, Tehran, Iran, all in computer engineering. He is currently working as a computer teacher at the Ministry of Education, in Technical Conservatory. His research interests are data mining, optimization and classification algorithms.

Zeinab Torabi received her Ph.D. degree in Computer System Architecture from Shahid Beheshti University, Tehran, Iran, in 2016. She is currently an Assistant Professor of Computer Engineering in Department of Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran. Her research interests include computer arithmetic, residue number system, design and analysis of algorithms.

Sahar kianian received her B.Sc. degree in computer Engineering (2007) from razi University, also M.Sc. and Ph.D. degrees in computer Engineering from Isfahan University (2010 and 2016, respectively). She is an assistant professor of computer engineering at Shahid Rajaee University. Her research interests are the application of algorithms, machine learning and data science to complex networks, focuses on protein interactions, connections of neurons and relationships among people. Applications include disease prediction, drug discovery, event detection and tracking, recommendation system, web mining and social influence mining.

How to cite this paper:

Y. Rohani, Z. Torabi, S. Kianian, "A Novel Hybrid Genetic Algorithm to Predict Students' Academic Performance," Journal of Electrical and Computer Engineering Innovations, 8(2): 219-232, 2020.

DOI: 10.22061/JECEI.2020.7230.373

URL: http://jecei.sru.ac.ir/article_1459.html