# Metaheruistic optimization based machine-learning approach for analysis of academic students' strategies

Lu Wang*
*School of Education, University of South California, LA, USA*

**Abstract**. In this technology world, education is also becoming one of the basic necessities of human life like food, shelter, and clothes. Even in day-to-day daily activities, the world is moving toward an automated process using technology developments. Some of the technology developments in day-to-day life activities are smartphone, internet activities, and home and office appliances. To cope with these advanced technologies, the persons must have basic educational qualification to understand and operate those appliances easily. Apart from this, the education helps the person to develop their personal growth in both knowledge and wealth. With the development of technologies, different Artificial Intelligence techniques have been applied on the datasets to analyze these factors and enhance the teaching method. But the current techniques were applied to one or two data models that analyze either their educational performance or demographic variable. But these models were not sufficient for analyzing all the factors that affects the education. To overcome this, a single optimized machine-learning approach is proposed in this paper to analyze the factors that affect the education. This analysis helps the faculty to enhance their teaching methodology and understand the student's mentality toward education. The proposed Hybrid Cuckoo search-particle swarm optimization was implemented on three datasets to determine the factors that affect the education. These optimal factors are determined by identifying their relations to the final results of an individual person. All these optimal factors are combined and grades are grouped to analyze the proposed optimization process performance using regression neural network. The proposed optimization-based neural network was tested on three data models and its performance analysis showed that the proposed model can achieve higher accuracy of 99% that affects the individual education. This shows that the proposed model can help the faculty to enhance their attention to the students individually.

Keywords: Education, demographic factors, optimization, hybrid, cuckoo search optimization, particle swarm, regression neural network

## 1. Introduction

The world is utilizing technological advancements to move toward an automated procedure. People need to acquire a foundational education in order to deal with these new technologies. A person's personal advancement in wealth and knowledge is facilitated by education. By introducing new ideas and effectively utilizing the resources at hand, it also aids in the development of nations. The current models used for estimating the academic performance were discussed below.

In [1] and [2], the decision tree models were used for estimating academic performance. Along with decision tree, the Naïve Bayes algorithm, and neural network are also used for estimating the student performance [3]. The deep learning algorithm is also used for estimating the dropout possibility of the students along with the traditional machine-learning algorithms [4]. In [5], the deep learning algorithms were implemented for estimating the

*Corresponding author. Lu Wang. E-mail: lwang417@usc.edu.

student performance. The importance of preprocessing step in estimating the student performance was analyzed in [6]. In [7, 8] also, the preprocessing importance was analyzed using the correlation map before classification. In [9], the machine-learning algorithms like decision tree and multilayer perceptron network model were analyzed.

In [10] and [11], also the ML model performance for estimating the student performance was evaluated. A complete review of machine-learning models for estimating the student performance was presented in [12]. In [13], the stochastic gradient descent optimization algorithm was used to tune the network performance of multilayer perceptron in student performance prediction. In [14] also proposed a complete automate approach of student performance prediction using machine-learning technique. In [15] also analyzed the different types of ML models in estimating the student performance.

### 1.1. Related works

A increasing body of research has focused on incorporating machine learning methods into meta-heuristics to address combinatorial optimisation issues. This integration intends to enhance the performance of the me-ta heuristics in terms of the quality of the solutions, the convergence rate, and the robustness of the search [21].

According to [20], forecasting student retention offers information on possibilities for deliberate student counselling. The review offers a research viewpoint on predicting student retention using machine learning through a number of significant discoveries, including the identification of the variables employed in prior studies and the prediction approaches Educational data mining is a group of data mining applications in the field of education, according to [22]. The analysis of student and teacher data is a focus of these programmes. The analysis could be put to use for classification or forecasting. Examined are machine learning techniques including Nave Bayes, ID3, C4.5, and SVM. [23] suggests a new machine learning-based model to forecast undergraduate students' final test marks using the results of their midterm exam grades as the source data. An educational process and data mining plus machine learning (EPDM + ML) model was introduced in [24] and used to contextually analyse the performances of the teachers and provide recommendations based on information obtained from students' evaluations of instruction (SET). Higher education institutions have

begun integrating technology into their conventional teaching methods in [25] in an effort to improve learning and teaching. In this study, five machine learning algorithms were used to analyse two datasets in order to predict and categorise student performance.

### 1.2. Shortcomings of the current techniques

Because the traditional machine-learning algorithms like SVM, ANN, and K-means techniques were applied to the educational dataset to observe the pattern in the set, and it helps to estimate the possible future outcomes. Even though the existing techniques can estimate the student performance but its performance was lagging in the following process:

The attribute analysis that affects the student performance was performed through the unsupervised learning or correlation analysis. Some of the models utilized thresholds for predicting the performance.

### 1.3. Motivation of study

Demographic elements including a parent's educational level, environment, and one's own desire in education are among the many aspects that influence an individual's education. On the datasets, a variety of artificial intelligence algorithms have been used to analyze these variables and improve the teaching strategy. These models were not enough for analysing every element that has an impact on education. In this paper, a single, optimized machine-learning strategy is suggested.

## 2. Materials and methods

The survey analysis showed that the machine-learning algorithm is not only efficient in image classification or pattern recognition but it also plays a vital role in data mining applications like academic performance analysis to identify the relation between the inputs and outputs. Because the traditional machine-learning algorithms like SVM, ANN, and K-means techniques were applied to the educational dataset to observe the pattern in the set, and it also helps to estimate the possible future outcomes.

As the aforementioned both processes are depending on the user criteria, its performance will not be stable in predicting the results. Hence, in this, an optimized machine-learning approach is proposed for finding the factors that affect the student's performance. This helps the faculty to counsel their

students and improve their performance. The proposed method was tested on the three types of datasets as follows:

- Student exam performance dataset
- Student academic performance dataset
- Portuguese student academic performance dataset.

The above three datasets have different input attributes and a common output attribute called student performance with different notations. These input attributes were processed by the proposed hybrid cuckoo-search and particle swarm optimization to find the optimal attribute that affects the student performance using regression neural network (RNN). Then, these factors are combined and tested using the RNN model for final evaluation. The working of the proposed model is explained briefly below.

## 3. Proposed work

In this, the hybrid cuckoo search and particle swarm optimization were used to find the optimal factors that affect the student performance and was determined by solving the objective function. Here, the objective function is to minimize the regression neural network (RNN) classifier error rate. This process was performed on three types of datasets. These datasets were downloaded from the Kaggle website. Before performing the aforementioned process, the datasets undergone the following process:

- Data collection
- Preprocessing
- Assigning values to categorical variables
- Normalization
- Data Exploration

Then, the preprocessed data will be subjected to the optimal attribute selection using hybrid CSO-PSO approach. The optimization process finds the factors that affect the student performance using the following objective function.

$$HCSO - PSO_{obj\ function} = \min\left(\text{RNN Error Rate}\right) \tag{1}$$

From this, the optimal factors from each model are collected, and the final performance was grouped as low, medium, and high. Then, the transformed dataset was trained using regression neural network, and its

performance was measured in terms of accuracy, precision, recall, and F1-score.

### 3.2. Dataset

In this paper, the three types of student performance data were collected from the kaggle websites [16–18]. The three datasets describe the following information:

- In the first dataset [16], the student performance was analyzed in terms of student and parent behavior.
- In the second dataset [17], the student exam performance was analyzed in terms of students reading pattern and parent educational qualification.
- In the third dataset [18], the student performance was analyzed in terms of student and parent nature.

The data distribution in this three datasets is presented in Table 1.

The above information in the dataset were either in categorical or numerical formats. But for an efficient data mining process, the data should be in a single format. In particular, the numeric data format helps to analyze the information in faster and efficiently. Hence, in this, the datasets were subjected to a preprocessing step

### 3.3. Preprocessing

For an efficient and faster analysis, the datasets were subjected to a data transformation process.

In data transformation process, the binary variables like true or false, yes or no, or any two class variables like gender, college, or address nature are transformed to 1 and 0

$$V_{[binary]} \xrightarrow{transformation} V_{[0,1]}. \tag{2}$$

In categorical variables like educational stage, student nationality, and other categories are transformed into 1 to n categories based on number of categorical options.

$$V_{[categorical\ 1\ to\ n]} \xrightarrow{transformation} V_{[1\ to\ n]} \tag{3}$$

Once the datasets are preprocessed using the above Equations 2 and 3, the remaining data also combined together in each set to form a complete numerical dataset. Then, the transformed data was subjected to Z-score normalization to normalize the values

Table 1
Data distribution in three datasets

| Variables | Dataset 1 | Dataset 2 | Dataset 4 |
|---|---|---|---|
| Total number of records | 480 | 1001 | 650 |
| Input attributes | 15 | 7 | 32 |
| Output attributes | Educational stage:<br>Low<br>Medium<br>High | Test score<br>0 to 100 | Scores G3<br>0 to 20 |
| Features analyzed | Demographic<br>Academic details<br>Student behavior<br>Parent response | Demographic Information<br>Student academic information<br>Parent | Demographic<br>Academic information<br>Parent information<br>Student behavior |
| Demographic | Gender<br>Nationality<br>Place of birth | Gender<br>Place<br>Lunch | Gender<br>Age<br>School<br>Address<br>Reason for school selection |
| Academic Information | Student information<br>Grade<br>Subject<br>Section id<br>Semester<br>Absence<br>Subject materials:<br>Revised<br>Announcements viewed<br>Discussion groups participated | Test preparation<br>Read score<br>Write score | Nursery<br>Higher<br>Grade<br>Study time<br>Absence<br>Health status<br>Educational support from<br>school and parents |
| Student behavior | Raised hand | | Internet<br>Relationship<br>Alcohol consumption during<br>workout and weekend |
| Parent | Parent response to the groups<br>Parent satisfaction | Parent education | Parent education status<br>Parent occupation<br>Parent cohabited<br>Family time |

between 0 and 1 in all categories. The z-score normalization calculates the new values for all the variables using Equation 4.

$$V_{new} \xrightarrow{z-score} \frac{V - mean(V)}{std(V)} \quad (4)$$

Then, the normalized values are subjected to the data exploration process for the first stage of data reduction.

### 3.4. Data exploration

In this stage, the data distribution individually was analyzed which least impact on the target variable is called output variables in each set. The sample data exploration of the set in terms of gender and school is shown in Figs. 1 and 2.

In the above Fig. 1, the demographic variables like gender, subjects, guardian of the student were presented in figure format for dataset 1 with gender as base variable.

In Fig. 2, the demographic variables like gender, nationality, and lunch activities of the student were presented in figure format for dataset with gender as the base variable.

In Fig. 3, the demographic variables like gender, locality, guardian of the student were presented in figure format for a dataset with gender as base variable.

In the data exploration part, the individual variable analysis was only carried out which gives an idea of the data distribution in the set. Based on this, the general variables like gender, nationality, and place of birth are removed from the sets. Then, the remaining variables were subjected to attribute selection using the proposed hybrid CSO-PSO.

### 3.5. HCSO-PSO-based attribute selection

In this, the preprocessed variable from the set was subjected to optimization process to select the optimal attributes that are responsible for the student performance was determined by the proposed
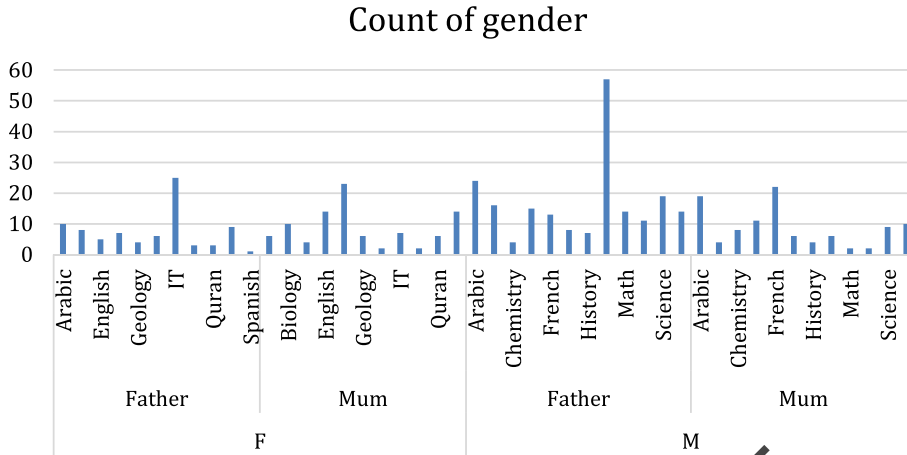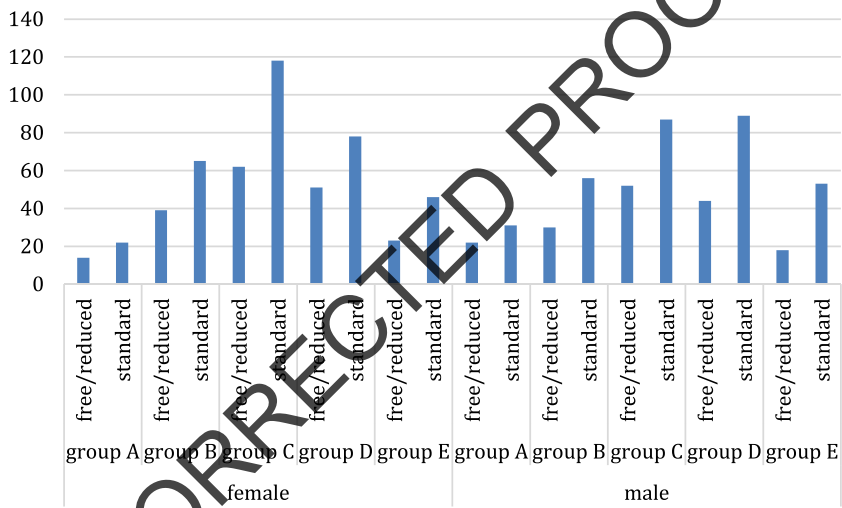
## Count of gender
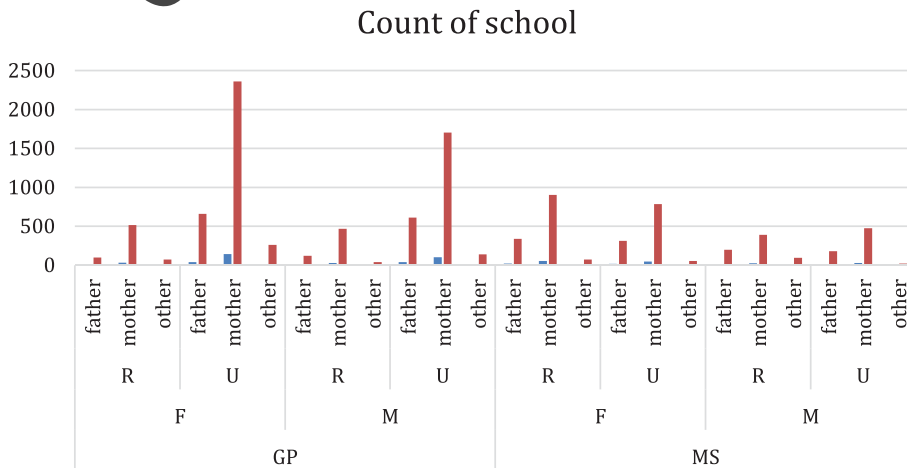


Fig. 1. Dataset 1 exploration.



Fig. 2. Dataset 2 exploration.

## Count of school



Fig. 3. Dataset 3 exploration.
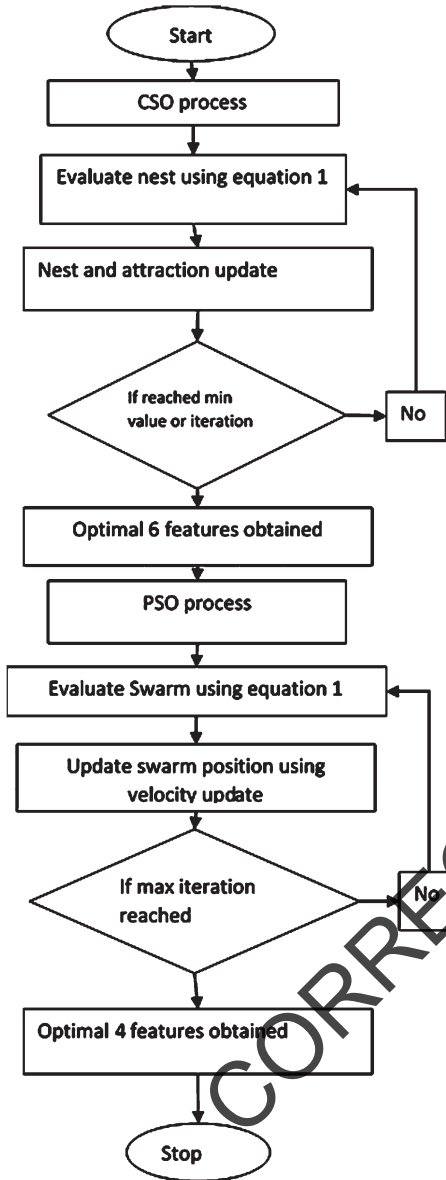
Fig. 4. HCSO-PSO processes.

Table 2
Initialization values for CSO

| Parameter | Value |
| --- | --- |
| Total no of nests | 20 |
| Maximum iterations | 50 |
| Attraction of nest | 0.25 |
| Lower bound | 1 for all sets |
| Upper bound | 12 for set 1 |
| | 6 for set 2 |
| | 29 for set 3 |
| No of output | 6 |

Table 3
Initialization values for PSO

| Parameter | Value |
| --- | --- |
| Number of swarms | 50 |
| Maximum iteration | 50 |
| Inertia weight | 0.8 |
| C1 | 1.5 |
| C2 | 2.0 |
| Lower bound | 1 |
| Upper bound | 6 |
| Optimal attributes | 4 |

HCSO-PSO process. Both processes utilize the same objective function mentioned in Equation 2 was used. The boundary variables were only modified in the first stage of process using cuckoo search optimization. While in the second stage, the boundary variables are common to find the four optimal variables. The proposed method process is represented in the flow chart in Fig. 4.

### 3.5.1. First stage attribute selection using CSO

In this stage, the optimal six attributes from each set are determined using Table 2 Cuckoo search initialization parameters. Using Equation 1 and the steps mentioned in the [19] was used to find the six best attributes for evaluate the student performance.

Apart from dataset 2, in datasets 1 and 3, the optimal attributes were selected from the set. In set 2, only six attributes are present; hence, all the six attributes were selected for the second stage process. In set 1, the student behavior and education were selected. In set 3, the student behavior like absences, study time, romantic, internet, and education g1, g2 was selected for set 3. These reduced sets were subjected to the second stage of processing using particle swarm optimization.

### 3.5.2. Second stage attributes selection using PSO

In this stage, the optimal four attributes that affect the student performance were determined using Table 3 Particle swarm initialization parameters. Using Equation 1 and the swarm position update equations were used to find the four optimal attributes in all the sets.

The PSO begins its process by minimizing Equation 1 objective function. After the first iteration, the swarm update its position using Equation 5.

$$Swarm_{newPosition}$$
$$= Swarm_{currentPosition} + Swarm_{newvelocity} \quad (5)$$

Before position update, the swarm update its velocity using Equation 6 to find the optimal attributes that help to enhance the academic performance of students

$$velocity_{new}$$
$$= Rand\,(0,\,1)*(LBS - Swarm_{currentPosition} \quad (6)$$
$$+(GBS - Swarm_{currentPosition}$$

where LBS is the local best solution and GBS is the global best solution. With these equations, the swarm determines the optimal four attributes from each set. In set 1, the student performance like discussion, view, raising hand, and response was selected as optimal attributes. In set 2, apart from lunch and place of birth, all other attributes are selected. In set 3, the grade performance, study time and romantic were selected.

### 3.6. Regression neural network

The reduce attributes from the hybrid CSO-PSO algorithm, a new dataset was framed with the reduced attribute set for analysis. Then, the new set is divided into training and testing using hold-out approach with a 0.3% value. This helps to split the dataset into 70% as training and 30% as testing.

The optimal four attributes will act as the first layer or input to the regression neural networks. In the corresponding test score, categorical variables were given as the outputs to the network.

The kernel function was multiplied with the inputs to determine the neuron weights using Equations 7 and 8.

$$kernel\,(x,\,x_n) = e^{-k_d \big/ 2\sigma^2} \quad (7)$$

$$k_d = (x - x_n)'\,(x - x_n) \quad (8)$$

In this layer, the kernel function in the step 2 is the denominator and the corresponding output is the numerator.

The inputs and outputs are related by dividing the numerator and denominator parts.

$$Y = \frac{\sum_{k=1}^{N} yk * Kernel(x,\,x_n)}{\sum_{k=1}^{N} Kernel(x,\,x_n)} \quad (9)$$

Then, the transformed dataset was trained using a regression neural network, and its performance was measured in terms of accuracy, precision, recall, and F1-score.
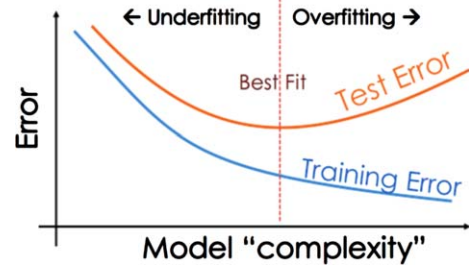


Fig. 5. Model types based on error.

## 4. Evaluation metrics

In this, the proposed HCSO-PSO performance in attribute selection was evaluated using regression neural network training and testing model evaluation metrics. Because the RNN was modeled using the selected attributes to estimate the student performance. In this, the following metrics were used for evaluation, and it plays a major role in the evaluation process.

### 4.2. Accuracy

Equation 10 is used for calculating the accuracy. This value evaluates the model in terms of the correct estimation of student performance.

$$Accuracy = \frac{TP + TN + FP + FN}{Overall\ Records} \quad (10)$$

The accuracy indicates the model types like underfitting, overfitting, and correct model. This accuracy will be calculated for both the training and testing to determine the model types.

#### 4.2.1. Under fitting model
When a model has lower accuracies in both the sets, then that model is an underfitting model as shown in Fig. 6.

#### 4.2.2. Overfitting model
When a model has higher accuracy in training and lower accuracy in testing means, the model is highly correlated to the training set, and it will act as an overfitting model as shown in Fig. 6.

#### 4.2.3. Accurate model
When the model accuracy is the same and higher for both training and testing means, it is a correct model for future estimation.

From Fig. 5, the following points were observed.

Table 4
Model types based on error

| Model | Error nature |
| --- | --- |
| Underfitting | Training and testing error is high |
| Overfitting | Testing error > training error |
| Correct model | Training and testing error is optimal and somewhat equal |

### 4.3. Precision

Equation 11 is used for calculating the model precision. This value is used to identify the model performance in estimating the values accurately.

$$Precision = \frac{TP}{TP + FP} \qquad (11)$$

### 4.4. Recall

Equation 12 is used for calculating the model precision. This value is used to identify the model performance in estimating the total number of times the model can find the values accurately.

$$Recall = \frac{TP}{TP + FN} \qquad (12)$$

### 4.5. F1-score

F1 score is calculated from precision and recall and is given in Equation 13.

$$F1 - score = 2 * \frac{Precision \cdot Recall}{Precision + Recall} \qquad (13)$$

Using Equations 10 to 13, the proposed HCSO-PSO-based regression neural network model is used in estimating the student performance was evaluated and the corresponding results are given below.

## 5. Results and discussion

In this, the proposed method HCSO-PSO was analyzed on the three datasets to find the optimal attributes for estimating the student academic performance. Then, the optimal attributes were trained using the regression neural network and tested. Then, the network performance was evaluated using accuracy, precision, recall and F1-score.

### 5.1. Dataset

In this, the sample screenshots of the three datasets are shown in Figs. 6–8.

### 5.2. Preprocessing and exploration

In the preprocessing step, the data were transformed and normalized the values 0 and 1 as shown in the sample Fig. 9 for set 2.

### 5.3. Data exploration

The results of data exploration are presented in Figs. 1–3 in the proposed work section.

### 5.4. Attribute selection using HCSO-PSO

The optimization process performance will be evaluated using the convergence curve. The convergence curve is the graph drawn between the iterations and fitness function. This shows the optimization performance speed in finding the optimal attribute selection.

| gender | NationalITy | PlaceofBirth | StageID | GradeID | SectionID | Topic | Semester | Relation | raisedhan | VisITedRe | Announce | Discussio | ParentAns | Parentsch | StudentA | Class |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 15 | 16 | 2 | 20 | Yes | Good | Under-7 | M |
| M | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 20 | 20 | 3 | 25 | Yes | Good | Under-7 | M |
| M | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 10 | 7 | 0 | 30 | No | Bad | Above-7 | L |
| M | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 30 | 25 | 5 | 35 | No | Bad | Above-7 | L |
| M | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 40 | 50 | 12 | 50 | No | Bad | Above-7 | M |
| F | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 42 | 30 | 13 | 70 | Yes | Bad | Above-7 | M |
| M | KW | KuwaIT | MiddleSchool | G-07 | A | Math | F | Father | 35 | 12 | 0 | 17 | No | Bad | Above-7 | L |
| M | KW | KuwaIT | MiddleSchool | G-07 | A | Math | F | Father | 50 | 10 | 15 | 22 | Yes | Good | Under-7 | M |
| F | KW | KuwaIT | MiddleSchool | G-07 | A | Math | F | Father | 12 | 21 | 16 | 50 | Yes | Good | Under-7 | M |
| F | KW | KuwaIT | MiddleSchool | G-07 | B | IT | F | Father | 70 | 80 | 25 | 70 | Yes | Good | Under-7 | M |
| M | KW | KuwaIT | MiddleSchool | G-07 | A | Math | F | Father | 50 | 88 | 30 | 80 | Yes | Good | Under-7 | H |
| M | KW | KuwaIT | MiddleSchool | G-07 | B | Math | F | Father | 19 | 6 | 19 | 12 | Yes | Good | Under-7 | M |
| M | KW | KuwaIT | lowerlevel | G-04 | A | IT | F | Father | 5 | 1 | 0 | 11 | No | Bad | Above-7 | L |
| M | lebanon | lebanon | MiddleSchool | G-08 | A | Math | F | Father | 20 | 14 | 12 | 19 | No | Bad | Above-7 | L |
| F | KW | KuwaIT | MiddleSchool | G-08 | A | Math | F | Mum | 62 | 70 | 44 | 60 | No | Bad | Above-7 | H |
| F | KW | KuwaIT | MiddleSchool | G-06 | A | IT | F | Father | 30 | 40 | 22 | 66 | Yes | Good | Under-7 | M |
| M | KW | KuwaIT | MiddleSchool | G-07 | B | IT | F | Father | 36 | 30 | 20 | 80 | No | Bad | Above-7 | M |
| M | KW | KuwaIT | MiddleSchool | G-07 | A | Math | F | Father | 55 | 13 | 35 | 90 | No | Bad | Above-7 | M |
| F | KW | KuwaIT | MiddleSchool | G-07 | A | IT | F | Mum | 69 | 15 | 36 | 96 | Yes | Good | Under-7 | M |
| M | KW | KuwaIT | MiddleSchool | G-07 | B | IT | F | Mum | 70 | 50 | 40 | 99 | Yes | Good | Under-7 | H |

Fig. 6. Dataset 1 sample.

| gender | race/ethnicity | parental level of education | lunch | test preparation co | math scor | reading sc | writing sco |
|--------|----------------|----------------------------|-------|---------------------|-----------|------------|-------------|
| female | group B | bachelor's degree | standard | none | 72 | 72 | 74 |
| female | group C | some college | standard | completed | 69 | 90 | 88 |
| female | group B | master's degree | standard | none | 90 | 95 | 93 |
| male | group A | associate's degree | free/reduced | none | 47 | 57 | 44 |
| male | group C | some college | standard | none | 76 | 78 | 75 |
| female | group B | associate's degree | standard | none | 71 | 83 | 78 |
| female | group B | some college | standard | completed | 88 | 95 | 92 |
| male | group B | some college | free/reduced | none | 40 | 43 | 39 |
| male | group D | high school | free/reduced | completed | 64 | 64 | 67 |
| female | group B | high school | free/reduced | none | 38 | 60 | 50 |
| male | group C | associate's degree | standard | none | 58 | 54 | 52 |
| male | group D | associate's degree | standard | none | 40 | 52 | 43 |
| female | group B | high school | standard | none | 65 | 81 | 73 |
| male | group A | some college | standard | completed | 78 | 72 | 70 |
| female | group A | master's degree | standard | none | 50 | 53 | 58 |
| female | group C | some high school | standard | none | 69 | 75 | 78 |
| male | group C | high school | standard | none | 88 | 89 | 86 |
| female | group B | some high school | free/reduced | none | 18 | 32 | 28 |

Fig. 7. Dataset 2 sample.

| school | sex | age | address | famsize | Pstatus | Medu | Fedu | Mjob | Fjob | reason | guardian | traveltime | studytime | failures | schoolsup | famsup | paid | activities | nursery | higher |
|--------|-----|-----|---------|---------|---------|------|------|------|------|--------|----------|------------|-----------|----------|-----------|--------|------|------------|---------|--------|
| GP | F | 18 | U | GT3 | A | 4 | 4 | at_home | teacher | course | mother | 2 | 2 | 0 | yes | no | no | no | yes | yes |
| GP | F | 17 | U | GT3 | T | 1 | 1 | at_home | other | course | father | 1 | 2 | 0 | no | yes | no | no | no | yes |
| GP | F | 15 | U | LE3 | T | 1 | 1 | at_home | other | other | mother | 1 | 2 | 0 | yes | no | no | no | yes | yes |
| GP | F | 15 | U | GT3 | T | 4 | 2 | health | services | home | mother | 1 | 3 | 0 | no | yes | no | yes | yes | yes |
| GP | F | 16 | U | GT3 | T | 3 | 3 | other | other | home | father | 1 | 2 | 0 | no | yes | no | no | yes | yes |
| GP | M | 16 | U | LE3 | T | 4 | 3 | services | other | reputation | mother | 1 | 2 | 0 | no | yes | no | yes | yes | yes |
| GP | M | 16 | U | LE3 | T | 2 | 2 | other | other | home | mother | 1 | 2 | 0 | no | no | no | no | yes | yes |
| GP | F | 17 | U | GT3 | A | 4 | 4 | other | teacher | home | mother | 2 | 2 | 0 | yes | yes | no | no | yes | yes |
| GP | M | 15 | U | LE3 | A | 3 | 2 | services | other | home | mother | 1 | 2 | 0 | no | yes | no | no | yes | yes |
| GP | M | 15 | U | GT3 | T | 3 | 4 | other | other | home | mother | 1 | 2 | 0 | no | yes | no | yes | yes | yes |
| GP | F | 15 | U | GT3 | T | 4 | 4 | teacher | health | reputation | mother | 1 | 2 | 0 | no | yes | no | no | yes | yes |
| GP | F | 15 | U | GT3 | T | 2 | 1 | services | other | reputation | father | 3 | 3 | 0 | no | yes | no | yes | yes | yes |
| GP | M | 15 | U | LE3 | T | 4 | 4 | health | services | course | father | 1 | 1 | 0 | no | yes | no | yes | yes | yes |

Fig. 8. Dataset 3 sample.

## 5.5. First stage attribute selection using CSO

Figure 10 shows that the CSO finds the six optimal attributes after 30 iterations by maintaining the error rate as minimum below 0.2 for all the sets. Apart from dataset 2, in datasets 1 and 3, the optimal attributes were selected from the set. In set 2, only six attributes are present; hence, all the six attributes were selected for the second stage process. In set 1, the student behavior and education were selected. In set 3, the student behavior like absences, study time, romantic, internet, and education g1, g2 was selected for set 3. These reduced sets were subjected to the second stage of processing using particle swarm optimization.

Figure 11 shows that the CSO find the six optimal attributes after 30 iterations by maintaining the error rate as minimum below 0.2 for all the sets. In set 1, the student performance like discussion, view, raising hand, and response was selected as optimal attributes. In set 2, apart from lunch and place of birth, all other attributes are selected. In set 3, the

grade performance, study time, and romantic were selected. Then, the selected attributes were split as training and testing and then subjected to regression neural network modeling.

Around 50 times of the iterations were performed, and various fitness function were used as inputs. The accuracy, precision and recall were then assessed as shown in Fig. 12. As a result, the existing methods achieves 95.65% and 77.98% of accuracy, where the proposed achieves 98.9% which is 3.35% and 11.08% better than aforementioned methods. Then, the proposed modeled network performance was evaluated using evaluation metrics and the corresponding values are tabulated in Table 5.

From Table 5, it was observed that the proposed HCSO-PSO-based RNN is the best model by having higher accuracy in training and testing. This proves that the proposed student's behavioral and past performance is best for predicting their academic performance.

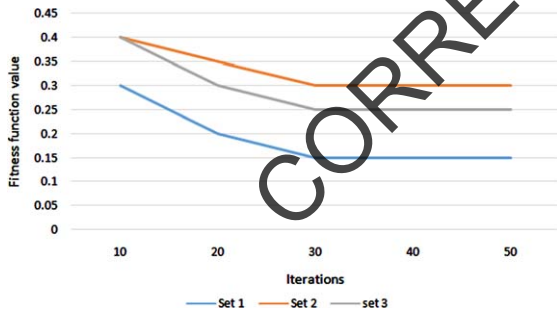| gender | race/ethnicity |
|--------|----------------|
| 0 | 2 |
| 0 | 3 |
| 0 | 2 |
| 1 | 1 |
| 1 | 3 |
| 0 | 2 |
| 0 | 2 |
| 1 | 2 |
| 1 | 4 |
| 0 | 2 |
| 1 | 3 |
| 1 | 4 |
| 0 | 2 |
| 1 | 1 |
| 0 | 1 |
| 0 | 3 |
| 1 | 3 |
| 0 | 2 |
| 1 | 3 |

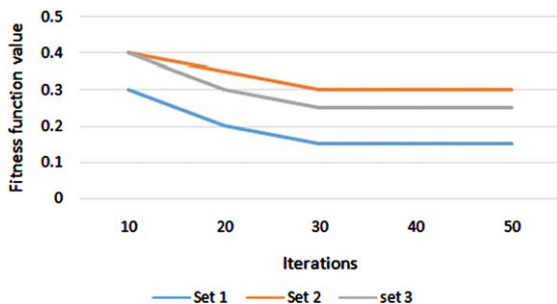Fig. 9. Preprocessed and normalized set.

Fig. 10. CSO convergence.

Fig. 11. PSO convergence.
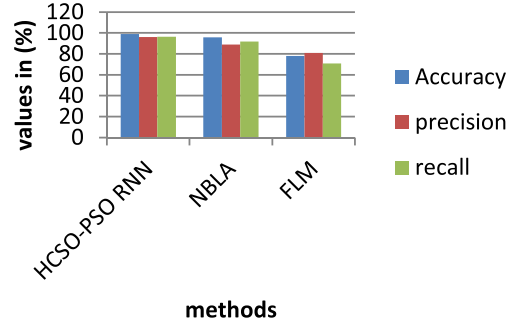
Fig. 12. Comparison for various parameters.

## 6. Discussion

This study's findings have several ramifications for academic advisors. Academic advisors could start conversations about the types of study strategies students are using, the number of hours per week they devote to studying, and the ways they use self-regulation strategies like picking study locations and dealing with distractions because the participants benefited from the study skills they learned during the intervention. advisers can develop courses or teach study techniques to one or a small group of students, regardless of their work contexts and areas of expertise (e.g., professionals at face-to-face conferences, group facilitators, academic members with significant caseloads, or online advisers for distance learners). Teaching students how to learn just addresses a portion of the issue for college-bound students. This study suggests that in order for students to learn and successfully apply the concepts in their lessons, they need time. Advisors can and should be the ones to explain the use and benefit of these study techniques and to motivate students to keep using their newly acquired strategies. They are frequently better equipped than others on campus to see issues and follow up for a few weeks or even months.

### 6.1. Implications and applications

Data mining methods are effective at locating hidden knowledge in huge datasets. When properly implemented in the educational sector, they can have significant positive effects on higher education institutions, students, education policymakers, and the prosperity of the nation. This work made a beneficial contribution to the field of educational data mining by finding the elements linked to academic probation. Policymakers, professionals, and educa-

Table 5
Performance comparison

| Algorithm | Testing (%) | Training (%) | Precision (%) | Recall (%) | Accuracy (%) |
|---|---|---|---|---|---|
| Proposed HCSO-PSO RNN | 98.7 | 98.9 | 96.2 | 96.4 | 98.9 |
| Neural-Based Learning Algorithm (NBLA) | 85.65 | 88.98 | 88.98 | 91.89 | 95.65 |
| Existing Method: Fuzzy Learning Model (FLM) | 80.9 | 79 | 80.98 | 70.78 | 77.98 |

tors can benefit from the findings by having a better understanding of the causes of the academic failure that can result in probation status. By taking into consideration certain other aspects besides high school performance, this can assist management in making better strategic decisions about the admission standards to universities. The study can also aid in determining the best strategy for pre-major and college admission requirements. The results of this study may be useful to academic advisors as well. According to substantial data, academic supervisors can, for instance, become more aware of the elements that have a negative impact on students' performance and adjust their advising approaches. Students can also comprehend and steer clear of the issues that contribute to subpar academic achievement. Additionally, after a student enrols in an institution, the proposed model can be used to forecast their academic performance there. The connected educational stakeholders can take the necessary actions in accordance with the model's predicted results. To avoid any barriers to academic growth, faculty should keep track of their advisees' academic performance, meet with them twice a semester, and develop appropriate action plans. Students who are on supervision or who are at risk of being placed on probation should get ongoing follow-up sessions from them. Students will successfully exit the probationary period thanks to this. Additionally, parents should, whenever possible, be involved in their child's academic circumstances. Parents who are well-informed could work together to create a much better learning environment for their children.

## 7. Conclusions

In this technology world, education is also becoming one of the basic necessities of human life like food, shelter, and clothes. Even in day-to-day daily activities, the world is moving toward an automated process using technology developments. Some of the technology developments in day-to-day life activ-

ities are smartphone, internet activities, and home and office appliances. Demographic elements including parent's educational level, environment, and one's own desire in education are among the many aspects that influence an individual's education. On the datasets, a variety of artificial intelligence algorithms have been used to analyze these variables and improve the teaching strategy. These models were not enough for analysing every element that has an impact on education. In this paper, a single, optimized machine-learning strategy is suggested. From the result evaluation, it was observed that the proposed HCSO-PSO-based RNN is the best model by having higher accuracy in training and testing. This proves that the proposed student behavioral and past performance is best for predicting their academic performance.

The results of this study indicate that some students may not be ready for the rigours of postsecondary education when they first enrol in higher education. The qualitative results strongly imply that research participants who were put on academic probation lacked basic academic skills, including the ability to study, communicate with their teachers, and complete required reading. They also employed scant study, self-control, and time management techniques. The negative effects of high school and family experiences, which do not adequately prepare students for success on their own in college, are also highlighted by the findings The study's subjects had completed high school with honours. However, other people lacked the drive or abilities to seek assistance when they were struggling academically. They also lacked the study, time management, and other skills required to deal with academic obstacles. Future studies should look at the intricate relationships between low performance in college students and the framework academic advisers might create to address bad study habits, either through individual interventions or Learning Skills Courses (LSCs). These initiatives may serve as a basis for developing effective, differentiated initiatives to improve learning and expand the motivating and learning techniques employed by

college.

## Funding statement

## Conflicts of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1] F.J. Kaunang and R. Rotikan, Students' academic performance prediction using data mining. *In 2018 Third International Conference on Informatics and Computing (ICIC), IEEE*, pp. 1–5, 2018. Available: http://dx.doi.org/10.1109/IAC.2018.8780547.

[2] Y.K. Saheed, T.O. Oladele, A.O. Akanni and W.M. Ibrahim, Student performance prediction based on data mining classification techniques, *Nigerian Journal of Technology* **37**(4) (2018), 1087–1091. Available: https://doi.org/10.4314/njt.374.1874.

[3] M. Alloghani, D. Al-Jumeily, A. Hussain, A.J. Aljaaf, J. Mustafina and E. Petrov, September: Application of machine learning on student data for the appraisal of academic performance. *In 2018 11th International Conference on Developments in ESystems Engineering (DeSE), IEEE*, pp. 157–162, 2018. Available: https://doi.org/10.1109/DeSE.2018.00038.

[4] M. Nagy and R. Molontay, Predicting dropout in higher education based on secondary school performance. *In 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES), IEEE*, pp. 000389–000394, 2018. Available: https://doi.org/10.1109/INES.2018.8523888.

[5] B.H. Kim, E. Vizitei and V. Ganapathi, GritNet: Student performance prediction with deep learning, *arXiv preprint arXiv:1804.07405*, 2018. Available: https://doi.org/10.48550/arXiv.1804.07405.

[6] B. Sekeroglu, K. Dimililer and K. Tuncal, Student performance prediction and classification using machine learning algorithms. *In Proceedings of the 2019 8th International Conference on Educational and Information Technology*, pp. 7–11, 2019. Available: https://doi.org/10.1145/3318396.3318419.

[7] H.R. Hasan, A.S.A. Rabby, M.T. Islam and S.A. Hossain, Machine learning algorithm for student's performance prediction. *In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1–7, 2019.

[8] A. Jain and S. Solanki, An efficient approach for multiclass student performance prediction based upon machine learning. *In 2019 International Conference on Communication and Electronics Systems (ICCES)*, IEEE, pp. 1457–1462, 2019.

[9] M. Imran, S. Latif, D. Mehmood and M.S. Shah, Student academic performance prediction using supervised learning techniques, *International Journal of Emerging Technologies in Learning* **14**(14) (2019).

[10] A.S. Hashim, W.A. Awadh and A.K. Hamoud, Student performance prediction model based on supervised machine learning algorithms. *In IOP Conference Series: Materials Science and Engineering*, IOP Publishing **928**(3) (2020), 032019.

[11] D.T. Ha, P.T.T. Loan, C.N. Giap and N.T.L. Huong, An empirical study for student academic performance prediction using machine learning techniques, *International Journal of Computer Science and Information Security (IJCSIS)* **18**(3) (2020), 21–28.

[12] J.L. Rastrollo-Guerrero, J.A. Gómez-Pulido and A. Durán-Domínguez, Analyzing and predicting students' performance by means of machine learning: A review, *Applied Sciences* **10**(3) (2020), 1042.

[13] S. Ranjeeth, T.P. Latchoumi and P.V. Paul, Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model, *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)* **14**(6) (2021), 1728–1741.

[14] H. Zeineddine, U. Braendle and A. Farah, Enhancing prediction of student success: Automated machine learning approach, *Computers & Electrical Engineering* **89** (2021), 106903.

[15] A. Khan and S.K. Ghosh, Student performance analysis and prediction in classroom learning: A review of educational data mining studies, *Education and Information Technologies* **26**(1) (2021), 205–240.

[16] I. Aljarah, Students' Academic Performance Dataset, *Kaggle*, 2016. Available: https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data.

[17] J. Seshapanpu, Students performance in exams, *Kaggle*, 2018. Available: https://www.kaggle.com/datasets/spscientist/students-performance-in-exams.

[18] D.S. Sean, Student performance data set, *Kaggle*, 2020. Available: https://www.kaggle.com/datasets/larsen0966/student-performance-data-set.

[19] X.S. Yang and S. Deb, Cuckoo search via Lévy flights. *In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), IEEE*, pp. 210–214, 2009.

[20] T. Cardona, E. Cudney, R. Hoerl and J. Snyder, Data mining and machine learning retention models in higher education, *Journal of College Student Retention: Research, Theory & Practice* **25**(1) (2023), 51–75.

[21] K. Mamaghan, Maryam, M. Mohammadi, P. Meyer, A. Mohammad and El-Ghazali Talbi, Machine learning at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art, *European Journal of Operational Research* **296**(2) (2022), 393–422.

[22] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz and K. Phasinam, Classification and prediction of student performance data using various machine learning algorithms, *Materials Today: Proceedings* **80** (2023), 3782–3785.

[23] Yağci Mustafa, Educational data mining: Prediction of students' academic performance using machine learning algorithms, *Smart Learning Environments* **9**(1) (2022), 11.

[24] Okoye Kingsley, Arturo Arrona-Palacios, Claudia Camacho-Zuñiga, Joaquín Alejandro Guerra Achem, Jose Escamilla and Samira Hosseini, Towards teaching analyt-

ics: A contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification, *Education and Information Technologies* (2022), 1–43.

[25] Chung Jae Young and Sunbok Lee, Dropout early warning systems for high school students using machine learning, *Children and Youth Services Review* **96** (2019), 346–353.