International Journal of Automation and Computing
www.ijac.net

18(6), December 2021, 981-992
DOI: 10.1007/s11633-021-1312-1

# Data Augmentation and Deep Neuro-fuzzy Network for Student Performance Prediction with MapReduce Framework

Amlan Jyoti Baruah[1] Siddhartha Baruah[2]

[1] Department of Computer Science and Engineering, Assam Kaziranga University, Jorhat 785006, India

[2] Department of Computer Application, Jorhat Engineering College, Jorhat 785007, India

**Abstract:** The main aim of an educational institute is to offer high-quality education to students. The system to achieve better quality in the educational system is to find the knowledge from educational data and to discover the attributes that manipulate the performance of students. Student performance prediction is a major issue in education and training, specifically in the educational data mining system. This research presents the student performance prediction approach with the MapReduce framework based on the proposed fractional competitive multi-verse optimization-based deep neuro-fuzzy network. The proposed fractional competitive multi-verse optimization-based deep neuro-fuzzy network is derived by integrating fractional calculus with competitive multi-verse optimization. The MapReduce framework is designed with the mapper and the reducer phase to perform the student performance prediction mechanism with the deep learning classifier. The input data is partitioned at the mapper phase to perform the data transformation process, and thereby the features are selected using the distance measure. The selected unique features are employed for the data segmentation process, and thereafter the prediction strategy is accomplished at the reducer phase by the deep neuro-fuzzy network classifier. The proposed method obtained the performance in terms of mean square error, root mean square error and mean absolute error with the values of 0.338 3, 0.581 7, and 0.391 5, respectively.

**Keywords:** Educational data mining (EDA), MapReduce framework, deep neuro-fuzzy network, student performance, data augmentation.

**Citation:** A. J. Baruah, S. Baruah. Data augmentation and deep neuro-fuzzy network for student performance prediction with MapReduce framework. *International Journal of Automation and Computing*, vol.18, no.6, pp.981–992, 2021. http://doi.org/10.1007/s11633-021-1312-1

## 1 Introduction

Educational data mining system is the bridge between computer science and education. Here, the subfields, like machine learning, computer science, and data mining are employed for performance prediction. Data mining is utilized to uncover hidden patterns in unstructured data. Accordingly, it is dedicated to find the knowledge and create relevant information[1]. Due to the advanced data mining system, it is more feasible to mine the educational information and gather the relevant information[2]. As a result, the relevant information gains more benefit to its handlers[3]. The increasing growth in the educational system distilling a large volume of information has led to the development of educational data mining (EDM)[4]. EDM is a particular field of data mining concept that ensures to discover the invisible patterns from the data to assist in making effective decisions for the administrators, teachers, and students. EDM makes the predictions that further classify the domain content knowledge, educational functionalities, applications, learner behavior, and assessment outcomes. The outcome helps the students in the learning process, whereas the tutors in the teaching area increase the educational practices and administrators in managing procedures[5]. Accordingly, student performance prediction is assumed as the important application offered by the EDM. The performance prediction strategy is based on different factors[6, 7], such that some of the parameters are elaborated for EDM[8]. The prediction of student performance gains an important factor in the education system. It is used to compute the future student′s performance after enrollment into the university, and thus it is employed to determine the best and poor score[9]. Accordingly, these results help to make the decision for admission more efficient and help to increase the quality of the academic services[10].

The administrators use the predictive results for analyzing the performance of the students in the subsequent

Research Article
Manuscript received May 11, 2021; accepted August 18, 2021
Recommended by Associate Editor Mao-Guo Gong
Colored figures are available in the online version at https://link.springer.com/journal/11633

semesters[11]. The lecturers select the appropriate learning mechanisms for the students based on the scores and determine how it would help the students increase their performance to a certain extent[12]. Such advantages desire the growth of computerized techniques for predicting the result with reliable accuracy[10]. There are different approaches for computing the student performance. Data mining is the popular method commonly employed in the educational field. Moreover, classification is a major popular method employed to predict student performance[13]. Various machine learning approaches[14] and deep learning approaches[15] include artificial neural network, naive Bayes, K-nearest neighbor, support vector machine (SVM)[16], convolutional neural network (CNN)[17], and the decision tree[18] are used in the data augmentation, mining, and classification purpose[19]. The AlexNet is a fundamental model developed for transferred learning to adapt classification[20]. The traditional researchers mainly focused on the result of previous semesters for predicting the student performance for the next semesters or current grade point average (GPA) to improve their academic performance. However, they did not evaluate the factors, like activity incentive grades and English entrance testing grades, as it degrades the student′s performance[21−23].

One of the commonly used prediction models is the artificial neural network-based prediction approach. The neural network-based performance prediction model is defined as demographic background to estimate the prediction mechanism. This approach improves the accuracy of the prediction rather than comparing it with the regression-based prediction mechanism. Accordingly, the improvement in the accuracy network-based model is based on a training algorithm[8]. In recent years, numbers of research works have been developed in the field of student performance prediction using knowledge tracing, deep learning, and cognitive diagnosis[24]. Deep knowledge tracing (DKT) gains more attractive in the area of performance prediction. Here, the deep learning model is combined with the knowledge training mechanism for predicting performance. The deep learning tracking is operated based on deep learning as well as the long and short-term memory model (LSTM) for tracking the performance of student knowledge with respect to time. In most cases, the oblivion gate in LSTM is utilized for simulating the process so that student mastery of previously learned knowledge reduces with respect to time[25]. DKT gained more success in analyzing sequential education data such that it mainly uses a recurrent neural network (RNN)[26] for tracking the student′s knowledge. Moreover, the DKT achieved higher performance in the knowledge tracing tasks, and it mainly focuses on exercise results by ignoring the features of student behavior that influence the performance of students[27].

## 1.1 Motivation

The student performance prediction mainly helps to make an efficient decision for the admission, and it helps to increase the quality of the academic services. This prediction mechanism makes the teachers find the behavior of students to improve the excellence of teaching. Recently, various approaches have been developed for computing student performance, but these methods have some challenges, which are given below:

1) The major issues that affect the performance of students are psychological status, family, and health, which require more consideration in analyzing the student learning behavior.

2) Some methods ignore the features of student behavior that influence the performance of the student.

3) The increasing growth in the educational system leads to a large volume of information, and the handling of the information is challenging in the student performance prediction systems.

These challenges in the existing student performance prediction methods are considered as a motivation and a novel method named fractional competitive multi-verse optimization-based deep neuro-fuzzy network (FCMVO-based DNFN) is developed for the student performance prediction. The MapReduce framework is modeled with the mapper and reducer phase so that the data transformation, feature selection, and the data augmentation process are accomplished at the mapper phase, whereas the student performance prediction mechanism is done at the reducer phase. The mapper phase comprises a number of mappers such that it is used to generate the data augmented result. The input data is partitioned into different data with varying dimensions, and the partitioned data is passed to the data transformation process, where the data transformation mechanism is done at the mapper phase by employing the log transformation model. After completing of the data transformation process, the feature selection mechanism is accomplished at the mapper phase by the Damerau-Levenshtein (DL) distance measure in order to choose the unique features. By considering unique selected features, the data augmentation process is done at the mapper phase by employing the Bootstrap model. Finally, the student performance prediction mechanism is performed at the reducer phase with a deep learning classifier named deep neuro-fuzzy network (DNFN) trained by the newly designed optimization algorithm named FCMVO algorithm. Accordingly, the proposed FCMVO is the incorporation of fractional calculus (FC) and competitive multi-verse optimization (CMVO) algorithm, respectively.

The major contribution of the research is explained as follows:

The MapReduce framework-based student performance prediction model is designed using the deep learning classifier named DNFN. The proposed FCMVO is employed to generate an effective prediction result with respect to fitness measures in such a way that the optimal solution is obtained with the best fitness value.

The paper is organized as follows: Section 2 explains the review of different traditional student performance prediction methods. Section 3 presents the proposed MapReduce framework-based deep learning classifier for predicting student performance. Section 4 explains the results and discussions of the proposed method, and finally, the conclusions are made in Section 5.

## 2 Literature survey

Some of the existing student performance prediction methods are reviewed in this section. Dien et al.[21] introduced the LSTM and CNN classifiers for predicting the student's performance. This method offered better prediction results by employing data transformation. However, this method failed to analyze a different group of student performance based on a different levels of marks. In [8], a cumulative dragon fly-based neural network (CDF-NN) was developed for predicting student performance with MapReduce architecture. Here, the features were captured from student information at the training phase, and the mapper is employed to generate intermediate data. However, it failed to use the deep learning model. In [28], a Bayesian knowledge tracing (BKT)-based LSTM model was developed for acquiring mastery level of individual skills using meaningful factors. It reported high potential in offering personalized instruction in the real world educational system. However, a major complexity of this model was a high rate of time complexity. In [29], a GridNet model was developed for analyzing the performance of student learning prediction. This method reported accurate predictions. However, indirect data was not incorporated with this method.

In [30], a machine learning model was modeled for classification and the prediction of student performance. Here, the raw data was effectively pre-processed before performing the classification process. This method was used to predict any category of educational data but only processed with limited datasets. In [31], an enhanced RNN (EERNN) model was developed by exploiting both the student text content and exercise records for predicting the performance of the student. Here, each state of the student was summarized into the integrated vector and traced with the RNN model. It generated better prediction results but failed to track the knowledge states of the student. In [10], a multi-adaptive neuro-fuzzy inference system (MANFIS) was developed by considering the learning strategy and multiple parameter sets to find the effective performance of the student. It used the representative set that generated better accuracy results with less computational complexity. However, this method failed to analyze the dynamic neuro-fuzzy model and perform the analysis of convergence rate. In [27], a machine learning approach was developed for analyzing the performance of students. Here, features of student behavior were integrated for increasing the prediction result. The fusion attention-based RNN model was employed for predicting the student's performance. It achieved better prediction accuracy but failed to employ the relation among knowledge ideas to gather the knowledge structure in the prediction process.

## 3 Proposed MapReduce framework-based deep learning model for student performance prediction

The student performance prediction[32−34] becomes critical demand in the institution as it supports students in generating higher performance in learning. A deep learning method is developed in this research to compute the student performance in education data mining systems based on the MapReduce framework. The proposed method involves different phases, such as data transformation, feature selection, data augmentation, and performance prediction. Accordingly, the data transformation process is done using log transformation. The unique features are selected at the feature selection phase by distance measure[32]. With selected features, the data augmentation process is carried out with the Bootstrap method[33]. Finally, the performance prediction is achieved with the DNFN classifier such that the training of the classifier is accomplished by the proposed CMVO algorithm. Fig. 1 represents a schematic view of the proposed method.

MapReduce is the simplified programming paradigm used for the distributed processing system in order to model an efficient and effective large-scale data processing framework. The advantage of using the MapReduce framework is that it effectively handles fault tolerance, low level issues, and parallelism. The MapReduce framework is composed of the mapper and reducer functions. Let us assume the dataset as $M$ with input data $A$:

$$M = \{A_1, A_2, \cdots, A_i, \cdots, A_n\}; i \in [m \times n]. \quad (1)$$

Here, $A_i$ denotes the input data with the dimension of $[m \times n]$, respectively.

### 3.1 Mapper phase

In the MapReduce framework, the mapper phase is employed to process the functions, like data transformation, feature selection, and data augmentation. The mapper function is used to take the input data and generate intermediate data pairs. The mapper phase contains a number of mappers such that each mapper is responsible for performing the process of data transformation, feature selection, and data augmentation process. The steps performed in the mapper phase are elaborately explained as belows.

#### 3.1.1 Data transformation based on log transformation

The input data $A_i$ with the dimension $[m \times n]$ is parti-

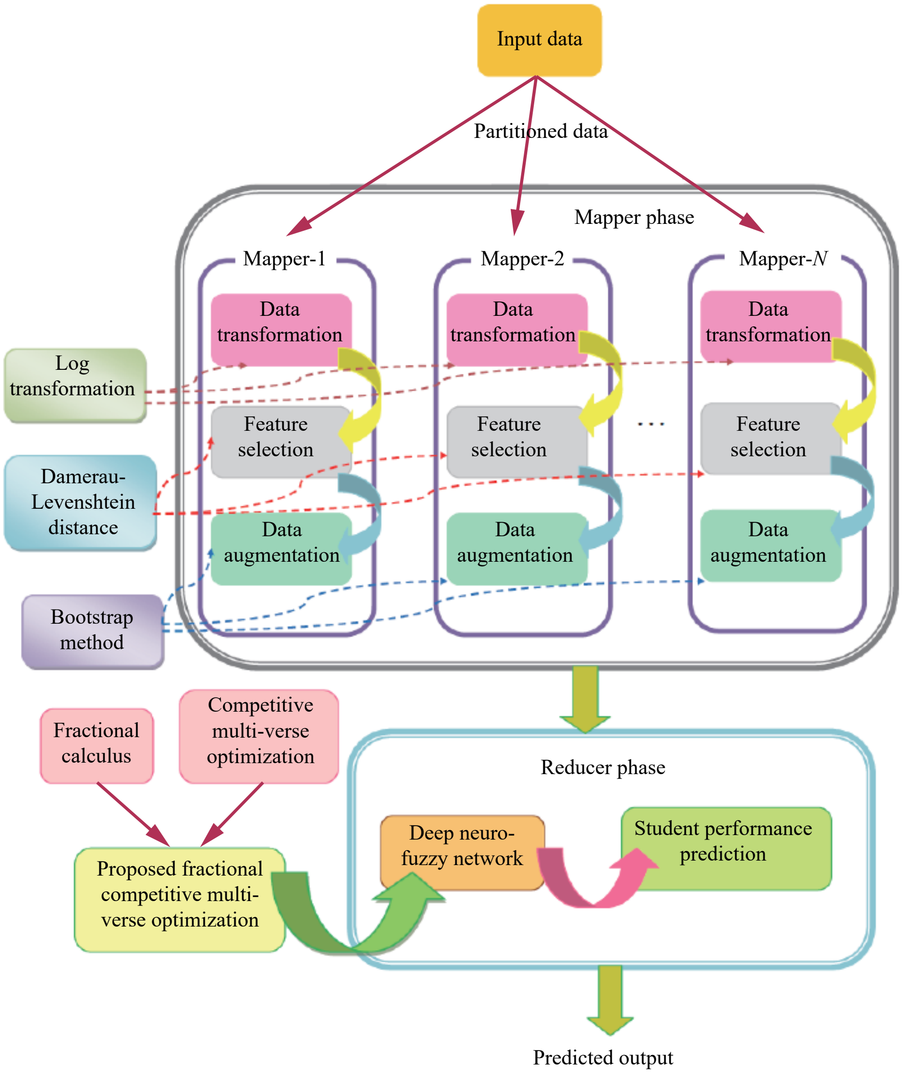


Fig. 1 Schematic view of proposed FMVO-based DNFN for student performance prediction with MapReduce framework

tioned into a number of data points by varying the dimensions, such as $[m_1 \times n]$, $[m_2 \times n]$, $[m_3 \times n]$, etc., for an easy-to-read format, which facilitates easy analysis. The partitioned data is fed to the data transformation process, where the log transformation is utilized to transform data into some other form. Data transformation is mainly employed to augment the training data. The log transformation is used to minimize the spatial variance of data.

Log transformation: It is employed to find skewed data such that this transformation is commonly used in most of the research fields. This transformation is used to transform the skewed data to the normality form. The log-transformed data follows the normal distribution. It is the process of considering the mathematical function and applying it to the input data. Here, each data is replaced with $\log(\cdot)$, which is depends on the analysis. In general, base 10 is commonly used for the transformation process. It is mainly useful to compress the data while plotting the histograms. The log of data is used to restore the symmetry of the data.

$$B = \log_{10}(A_i). \tag{2}$$

The log-transformed result is denoted as $B$, used to perform the feature selection process at the mapper phase.

#### 3.1.2 Feature selection using Damerau–Levenshtein distance

The result of the data transformation process is passed to the feature selection mechanism, where the optimal and the best features are selected by employing the DL distance. Feature selection is the mechanism of selecting the subset of relevant features for performing the student performance prediction. It maintains the physical meanings of original features and offers better interpretability and readability. Thus, it plays a key role in improving the prediction performance, preventing overfitting, and minimizing the computational cost. By removing the features through the feature selection process, the computational and the storage cost is reduced.

Damerau–Levenshtein distance: The DL distance[35] is commonly applied to the string values that show the minimum number of operations needed for one value to be converted into another value. The operations included in the DL distance are insertion, replacement, and deletion of values. Moreover, it also considers the transposition of the adjacent values. For example, the strings "ca" and "abc" have the DL distance of "2", as "ca"→"ac"→ "abc", respectively. Let us consider two sets of data as $u$ and $v$ in such a way that the DL distance of these two sets of data is expressed as

$$DL_{u,v}(e,w) = \begin{cases} \max(e,w) \quad , & \text{if } \min(e,w) \\ \min \begin{cases} DL_{u,v}(e-1,w)+1 \\ DL_{u,v}(e,w-1)+1 \\ DL_{u,v}(e-1,w-1)+1 \ \ (u_e \neq v_w) \end{cases} & \text{, otherwise} \end{cases} \tag{3}$$

where $u$ and $v$ are the two sets of data, $e$ and $w$ specify the length of each respective set of data. Each mapper selects the features with the dimension of $[m_1 \times l_1]$, $[m_2 \times l_2]$, $[m_3 \times l_3]$, etc. Finally, the unique features selected by the DL distance at the mapper phase are specified as $f$ with the dimension of $[m \times l]$, respectively.

#### 3.1.3 Data augmentation based on the Bootstrap method

After selecting the unique features with the DL distance, the data augmentation process is accomplished at the mapper phase by the Bootstrap method[36]. Data augmentation is an important method employed to increase the diversity and the amount of data by randomly augmenting the data. It is specifically used to describe the invariance in the data domain. Bootstrap is the resampling method used for estimating the statistics of the population by sampling the dataset with the replacement. This method is otherwise called the sampling with replacement. It is employed to compute the data augmentation process. This process is done by taking small features, computing the statistic, and taking the average of computed statistics. The steps involved in the Bootstrap method are explained as follows:

1) A sample feature is selected from the unique features.

2) Draw a sample from the original features with replacement, and each re-sampled sample is termed Bootstrap sample.

3) Compute the statistic for each sample.

4) Construct the sampling distribution with the Bootstrap statistics by computing the standard error of statistics and finally generate the confidence level.

Finally, the augmented result generated by the Bootstrap method is represented as $B$ with the dimension of $[p \times l]$.

### 3.2 Reducer phase

The reducer function is employed to perform the process of student performance prediction with the DNFN system. The reducer function used the data augmentation result for generating the prediction performance. The reducer takes the output of the mapper and processes each of the data for generating the result. The output obtained from the reducer is the final predicted result of student performance.

#### 3.2.1 Student performance prediction using the proposed fractional competitive multi-verse optimization-based deep neuro-fuzzy network

The prediction of student performance becomes essential in the educational system because of the failure reduction and enhancement of student success and performance. Due to the large quantity of data in the database, it is more complex to predict the performance of the student in the learning system. The process of accomplishing student performance prediction is carried out with the DNFN, which is trained by the proposed FCMVO algorithm.

**1) Structure of deep neuro-fuzzy network**

The prediction of student performance is computed by the deep learning classifier at the reducer phase by considering the input as the data augmented result. DNFN[37] is the hybrid of fuzzy logic and the deep neural network classifier. Here, the deep learning classifier named DNFN is employed for generating the prediction result to analyze the performance of the student in the educational system. In this hybrid model, the deep neural network is first employed, and later the fuzzy logic is used to compute the system objectives. The entire system is built with the following layers: input layer, hidden layer, validation and output layer. The input layer depends on the input data value and the fuzzification value, which is the process of converting the input value into a fuzzy value in the system. The number of hidden layers employed in this architecture is three, namely the rule layer, normalization layer, and defuzzification layer. The output layer is termed the defuzzification layer, which is the process of obtaining a single value from the output of the aggregated fuzzy set. The most important factors employed in the DNFN structure are consequents and premises. Here, the premise factors are related to fuzzification at the input layer, and the consequent factor is related to the defuzzification process. Each input or output factor is mapped with an entity or node in the neuro-fuzzy network at each layer. The degree of each input is allocated with the value between 0 and 1.

Input layer: Let us consider two premises as $a$ and $b$ and one consequent as $k$, they are given as

$$H_{1,q} = \alpha R_q(a) \text{ or } H_{1,q} = \alpha S_{q-2}(b)\ , \quad \forall q = 1,2,3,4. \tag{4}$$

Here, $a$ and $b$ indicate input to each $q$-th entity, $\alpha R$ and $\alpha S_{q-2}$ indicate antecedent membership functions, and $H_{1,q}$ denotes the degree of membership. Here, membership functions are specified by bell-shaped functions also called as Gaussian membership functions such that it is assigned with the maximum value of 1, and minimum value of 0, respectively.

$$\alpha R_q(a) = \frac{1}{1 + \left|\frac{a - c_q}{h_q}\right|^{2d_q}} \tag{5}$$

where $d_q$, $c_q$ and $h_q$ indicate membership functions of premise parameter such that it is optimized by the training process.

Rule base layer: This layer is utilized to describe the set of rules. Each entity of this layer multiplies linguistic variable value for satisfying degree of membership. The output of this layer is given as

$$H_{2,q} = \chi_q = \alpha R_q(a)\,\alpha S_{q-2}(b),\ \forall q = 1, 2. \tag{6}$$

Here, $\chi_q$ denotes generic network parameter weight.

Normalization layer: At this layer, each entity computes the ratio of firing strength of the $q$-th rule with summation of the firing strength of all the rules. The output of each rule is normalized through the firing strength of the rule and is given as

$$H_{3,q} = \bar{\chi}_q = \frac{\chi_q}{\chi_1 + \chi_2},\ \forall q = 1, 2. \tag{7}$$

Defuzzification layer: Individual rule consequents are calculated for specifying the overall effect on output. Accordingly, the output of this layer is represented as

$$H_{4,q} = \bar{\chi}_q g_q = \bar{\chi}_q(z_q a + K_q b + L_q),\ \forall q = 1, 2. \tag{8}$$

Here, $z$, $K$ and $L$ indicate consequent parameters set.

Output layer: The final output generated at this layer is represented as

$$J = \sum_q \bar{\chi}_q g_q = \frac{\sum_q \chi_q g_q}{\sum_q \chi_q}. \tag{9}$$

The training process of DNFN is done by the proposed optimization algorithm named FCMVO algorithm. Fig. 2 portrays the structure of DNFN.

**2) Algorithmic procedure of proposed fractional competitive multi-verse optimization algorithm**

The training procedure of DNFN is done with the proposed FCMVO algorithm, which is derived by the integration of FC[38] and CMVO[39]. CMVO is the population-based algorithm, in which the population is grouped based on bicompetitions for generating two sets, namely winners and losers. In each competition, the location of losers is modified through the observation from winners rather than considering personal and the global best position. Also, it aims to solve the optimization problems by considering the single as well as multiple objectives. Furthermore, it preserves the population density and ensures premature convergence. The FC is used to enhance the computational performance of the algorithm. Also, it preserves privacy in data mining. Due to these advantages, the integration of the FC and CMVO algorithm improves the efficiency of the proposed method. The algorithmic steps involved in the proposed FCMVO-based DNFN are explained as follows:

**i) Initialization.** Initialize a random universe based on the population size $P$ and the size of the search space.

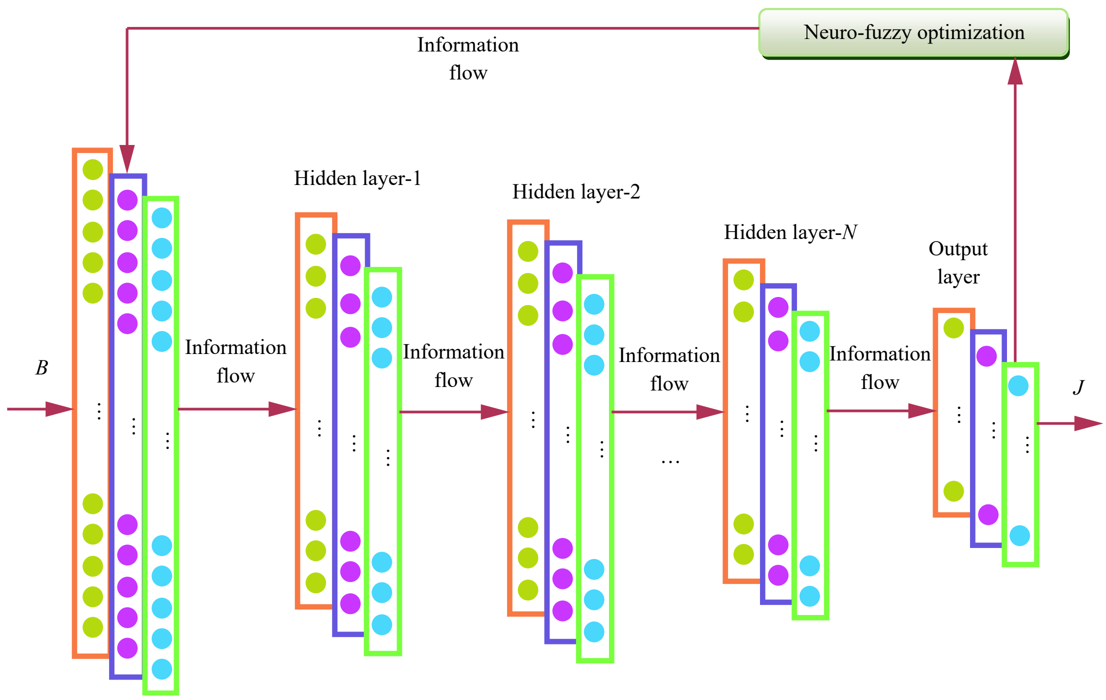


Fig. 2 Architecture of deep neuro-fuzzy network

Then specify the coefficients as $C$ and $D$, respectively.

**ii) Compute fitness measure.** The fitness measure is utilized to determine the optimal solution for the prediction problem. The fitness function used to generate the best fitness value is specified as

$$F = \frac{1}{N}\sum_{j=1}^{N}\left[O_j - J_j\right]^2 \tag{10}$$

where $F$ denotes fitness measure, $N$ indicates the total number of samples.

**iii) Compute the coefficients $D$ and $C$.** The two coefficients are computed as

$$D = 1 - \left(\frac{r^{1/t}}{r_{\max}^{1/t}}\right) \tag{11}$$

$$C = \min - r \times \left(\frac{\min - \max}{r_{\max}}\right). \tag{12}$$

Here, $r$ denotes current iteration, $r_{\max}$ indicates the maximum number of iterations, $t$ represents the accuracy of exploitation over iterations, and its value is set to 6, min specifies lower bound of $C$, and max indicates upper bound of $C$, respectively. Here, the coefficient $C$ denotes, existence probability and $D$ represents traveling distance rate.

**iv) Update solution:** The mathematical equation employed to update the position is given as

$$M_s^x(r+1) = T_1 D + T_2\left(M_y(r) - M_s(r)\right) + T_3\left(M(r) - M_s(r)\right) \tag{13}$$

$$M_s^x(r+1) = T_1 D + T_2 M_y(r) - T_2 M_s(r) + T_3 M(r) - T_3 M_s(r). \tag{14}$$

Subtracting the term $M_s(r)$ on both sides of the above equation is represented as

$$M_s^x(r+1) - M_s(r) = T_1 D + T_2 M_y(r) - T_2 M_s(r) + T_3 M(r) - T_3 M_s(r) - M_s(r). \tag{15}$$

According to the concept of FC, (15) is modified as

$$X^{\beta}\left[M_s^x(r+1)\right] = T_1 D + T_2 M_y(r) - T_2 M_s(r) + T_3 M(r) - T_3 M_s(r) - M_s(r) \tag{16}$$

$$M_s^x(r+1) - \beta M_s(r) - \frac{1}{2}\beta M_s(r-1) - \frac{1}{6}(1-\beta)M_s(r-2) - \frac{1}{24}\beta(1-\beta)(2-\beta)M_s(r-3) = T_1 D + T_2 M_y(r) - T_2 M_s(r) + T_3 M(r) - T_3 M_s(r) - M_s(r) \tag{17}$$

$$M_s^x(r+1) = \beta M_s(r) + \frac{1}{2}\beta M_s(r-1) + \frac{1}{6}(1-\beta)M_s(r-2) + \frac{1}{24}\beta(1-\beta)(2-\beta)M_s(r-3) + T_1 D + T_2 M_y(r) - T_2 M_s(r) + T_3 M(r) - T_3 M_s(r) - M_s(r) \tag{18}$$

$$M_s^x(r+1) = M_s(r)\left[\beta - T_2 - T_3 - 1\right] + \frac{1}{2}\beta M_s(r-1) + \frac{1}{6}(1-\beta)M_s(r-2) + \frac{1}{24}\beta(1-\beta)(2-\beta)M_s(r-3) + T_1 D + T_2 M_y(r) + T_3 M(r) \tag{19}$$

where $T_1$, $T_2$ and $T_3$ denote the random numbers that lie in the range of $[0,1]$, $\beta$ lies in the range of $[0,1]$, $M_y(r)$ denotes winner universes in the $r$-th round of competition, $M_s(r)$ indicates loser universes in the $r$-th round of competition, and $M_s(r-1)$ denotes loser universes in the $(r-1)$-th round.

**v) Termination:** The above steps are repeated until the best solution is obtained. Algorithm 1 portrays the pseudo-code of the proposed FCMVO-based DNFN.

**Algorithm 1.** FCMVO-based DNFN

1) **Input:** $C$, $D$, and $r_{\max}$
2) **Output:** $M_s^x(r+1)$
3) Initialize the set of population
4) for $(r < r_{\max})$ do
5) Compute the fitness measure
6) Calculate the coefficients $C$ and $D$
7) for each individual do
8) Update the solution
9) end for
10) end for
11) Return best solution

## 4 Results and discussions

This section explains the results and discussions of developed FCMVO-based DNFN with respect to the performance measures.

### 4.1 Experimental setup

The implementation of the proposed approach is carried out in Python running on a Windows 10 OS computer with 2GB RAM, and Intel i3 core processor. Table 1 shows the experimental setup of the proposed method.

### 4.2 Dataset description

The dataset utilized for experimentation contains random data that includes the performance of randomly selected students. The data attributes considered here are student grades, social features, school-related features, and demographics. Here, the student performance data is partitioned into 64 attributes such that they include cer-

Table 1 Experimental setup

| Parameter | Value |
|---|---|
| Mapper size | 5 |
| Feature size | 10 |
| Augment (over sample data) | 10 000 |
| Iteration | 10 |
| Layers | 10 |
| Population size | 10 |

tain details of the students, such as location, types of school, gender, semester marks, family details, etc. This information is used to generate the prediction of student performance. Accordingly, the dataset is separated into training and testing data, where training data uses the semester score of each student. This data is used to generate the prediction result for test data.

## 4.3 Evaluation metrics

The performance of the developed scheme is analyzed by employing the measures, such as mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE).

**MSE:** It is the measure that shows the error difference between the original and the estimated values, which is defined in (10).

**RMSE:** It is the square root of error difference among the original and the estimated value and is specified as

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(O_j - J_j)}. \qquad (20)$$

**MAE:** It shows the average absolute difference between the actual result and the predicted result.

## 4.4 Performance analysis

This section elaborates the performance analysis of the developed method by varying the number of layers and training data.

**1) Analysis based on number of layers**

Fig. 3 represents the performance analysis of the proposed scheme by varying layers at training data of 70%. Here, the number of layers varies as 5, 10, 15 and 20. Figs. 3(a)−3(c) portray the performance analysis based on MSE, RMSE and MAE. When considering the number of layers as 5, the MSE acquired by the proposed FCMVO-based DNFN with iteration 10 is 0.383 2, iteration 20 is 0.373 5, iteration 30 is 0.373 0, iteration 40 is 0.363 9, and iteration 50 is 0.343 9. Considering the layers as 5, the RMSE measured by the proposed model with iteration 10 is 0.619 0, iteration 20 is 0.611 2, iteration 30 is 0.610 8, iteration 40 is 0.603 2, and iteration 50 is 0.586 4. By considering the number of layers is increased to 20, the MAE reported by the developed approach by considering the iteration 10 is 0.359 5, iteration 20 is 0.245 2, iteration 30 is 0.228 0, iteration 40 is 0.170 8, and iteration 50 is 0.130 5. From this analysis, it is clear that the proposed method reveals the minimum MSE, RMSE and MAE when increasing the number of layers and iterations.

**2) Analysis based on training data**

Fig. 4 depicts the analysis made by the proposed model based on the training data by setting the number of layers as 10. The training data percentage is set to 60%,

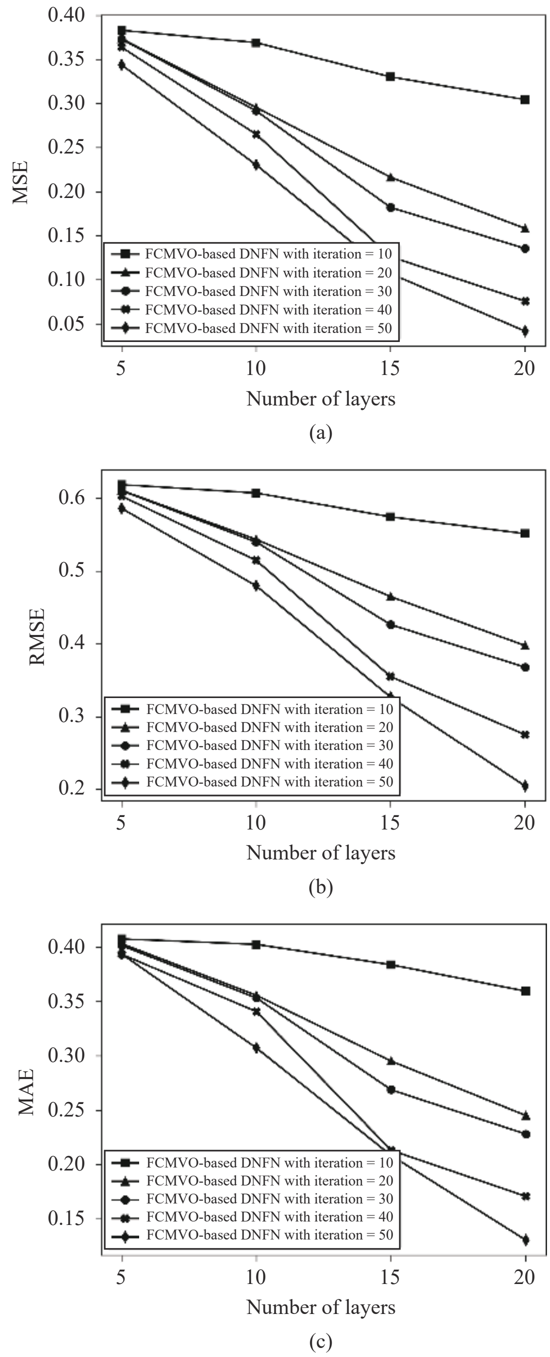


Fig. 3 Analysis based on number of layers: a) MSE, b) RMSE, c) MAE.

70%, 80% and 90%. Fig. 4(a) portrays the performance analysis made with the MSE measure. When considering the training data as 60%, the MSE measured by the proposed FCMVO-based DNFN for iterations 10, 20, 30, 40 and 50 is 0.376 1, 0.324 4, 0.301 3, 0.280 3 and 0.241 6, respectively.

The analysis computed with the RMSE measure is portrayed in Fig. 4(b). At 60%, the RMSE acquired by the proposed FCMVO-based DNFN with the iterations 10, 20, 30, 40 and 50 is 0.613 3, 0.569 6, 0.548 9, 0.529 4 and 0.491 5, respectively. Fig. 4(c) depicts the analysis with the MAE metric. When the training data is considered as 90%, the MAE computed by the proposed approach for iterations 10, 20, 30, 40 and 50 is 0.391 5, 0.336 9, 0.333 2, 0.282 5 and 0.272 2, respectively. Here, the minimum MSE, RMSE and MAE of the proposed FCMVO-based DNFN occurred at 90% of training data with iteration 50 because of the better interpretability and readability of the proposed method.

## 4.5 Comparative methods

The performance of the developed method is analyzed by considering the existing approaches, such as deep learning[21], CDF-NN[8], LSTM[28] and MANFIS-S[10].

## 4.6 Comparative analysis

This section presents the comparative analysis of the proposed FCMVO-based DNFN with respect to the training data. Fig. 5(a) portrays the analysis of MSE. When considering the training data as 60%, the MSE achieved by the proposed FCMVO-based DNFN is 0.376 1, which is 40.8%, 27.48%, 7.77% and 6.02% minimum than the MSE of the existing deep learning methods, such as CDF-NN, LSTM and MANFIS-S, respectively. The minimum MSE is considered as the best performance, and the proposed system attained the minimum MSE at 90% of training data with the value of 0.338 3.

Fig. 5(b) depicts the analysis with RMSE. By considering 80% training data, RMSE measured by the proposed FCMVO-based DNFN is 0.608 1, which is 17.27%, 15.02%, 3.77% and 2.23% minimum than the RMSE of existing methods, such as deep learning, CDF-NN, LSTM and MANFIS-S, respectively.

The analysis made with the MAE measure is represented in Fig. 5(c). Here, the minimum MAE is considered as the best performance. When considering 60% of training data, the MAE achieved by the proposed FCMVO-based DNFN is 0.409 5, which is 23.96%, 18.26%, 3.65% and 1.7% minimum than the MAE of other existing methods, such as deep learning, CDF-NN and LSTM, respectively. From this analysis, it is clear that the proposed FCMVO-based DNFN attains minimum MSE, RMSE and MAE than other existing methods, such as deep learning, CDF-NN and LSTM, due to the effectual training of the DNFN by the proposed FCMVO algorithm.

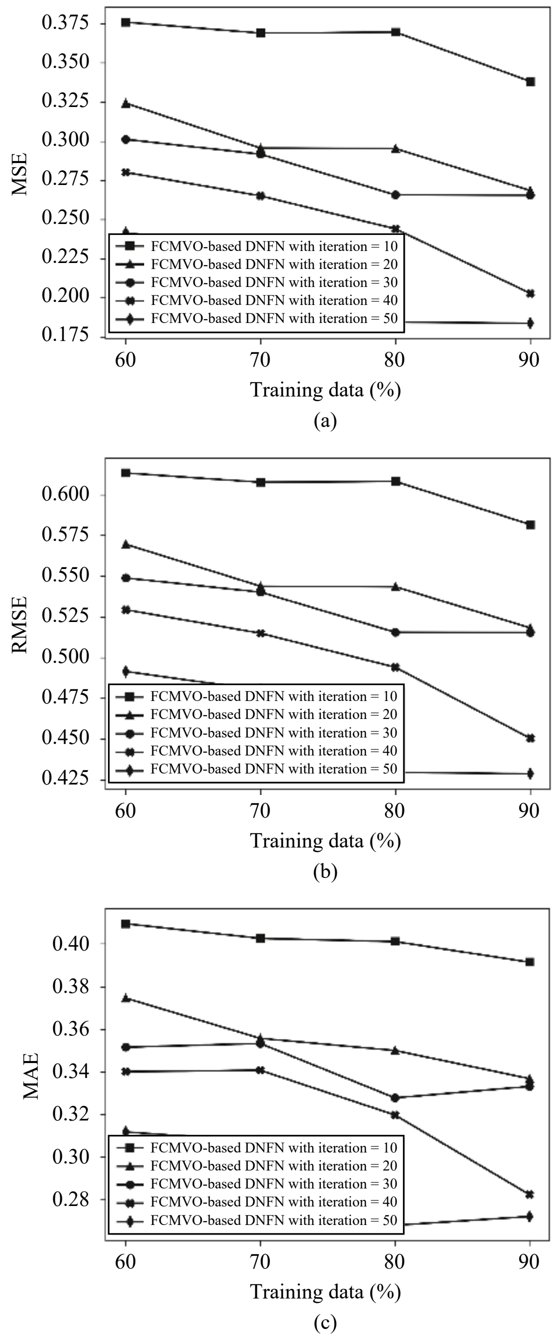


Fig. 4 Analysis based on training data: a) MSE, b) RMSE, c) MAE.

## 4.7 Comparative discussions

Table 2 portrays the comparative discussions of the proposed method and shows the values obtained at 90% of training data based on the best performance results. The MSE achieved by the proposed FCMVO-based DNFN is 0.338 3, which is 45.29%, 27.76%, 11.9% and 8.37% minimum than the MSE of existing methods, such as

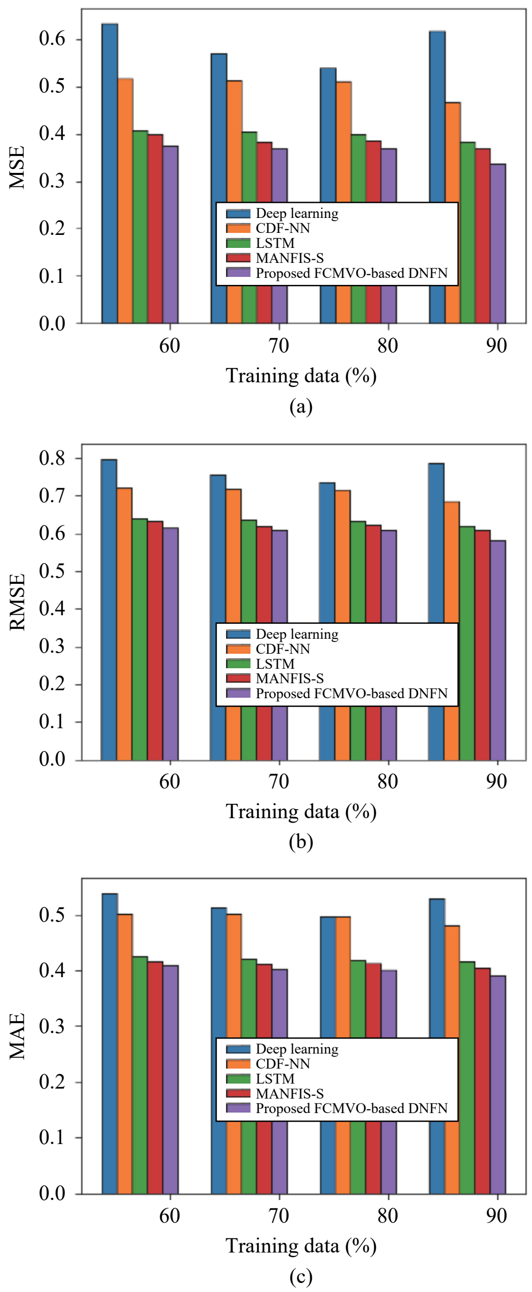


Fig. 5 Comparative analysis: a) MSE, b) RMSE, c) MAE.

deep learning, CDF-NN, LSTM and MANFIS-S, respectively. Moreover, the MAE of deep learning, CDF-NN, LSTM, MANFIS-S and proposed FCMVO-based DNFN is 0.528 0, 0.481 7, 0.415 4, 0.404 1 and 0.391 5, respectively.

The computational complexity of the proposed FCMVO-based DNFN is $O(P\,[m \times n])$, where $P$ denotes the population size and $[m \times n]$ denotes the dimension of the problem. The computational time of the proposed and the existing methods are provided in Table 3, in which the proposed FCMVO-based DNFN has a minimum computational time of 6 s.

The reasons for the best performance of the proposed

Table 2 Comparative discussions

| Metrics | Deep learning | CDF-NN | LSTM | MANFIS-S | Proposed FCMVO-based DNFN |
|---|---|---|---|---|---|
| MSE | 0.618 4 | 0.468 3 | 0.384 0 | 0.369 2 | **0.338 3** |
| RMSE | 0.786 4 | 0.684 4 | 0.619 7 | 0.607 6 | **0.581 7** |
| MAE | 0.528 0 | 0.481 7 | 0.415 4 | 0.404 1 | **0.391 5** |

Table 3 Computational time

| Methods | Deep learning | CDF-NN | LSTM | MANFIS-S | Proposed FCMVO-based DNFN |
|---|---|---|---|---|---|
| **Time (s)** | 12 | 10 | 8.5 | 7 | 6 |

FMVO-based DNFN are given below. At first, the spatial variance of the data is minimized using the log transformation. Then, the feature selection is carried out based on the DL distance, which offers better interpretability and readability. Also, it plays a key role in improving the prediction performance, preventing overfitting, and minimizing the computational cost. Then, the Bootstrap method is used for data augmentation, which increases both the diversity and the amount of data by randomly augmenting the data. Finally, the integration of the FC and CMVO improves the efficiency of the proposed method. Thus, the proposed method offers minimum MSE, RMSE and MAE than the existing methods, such as deep learning, CDF-NN, LSTM and MANFIS-S. The proposed method is used to make an effective decision for the administrators, teachers and students. Also, it is used to compute the future student performance after enrollment into the university, and thus it is employed to determine the best and poor scores. Moreover, it is used for understanding the students and their learning environment.

## 5 Conclusions

This research presents the MapReduce framework-based deep learning classifier for accomplishing the process of student performance prediction in the education data mining system. The proposed scheme comprises two functions, namely mapper and reducer functions, in such a way that the mapper function is situated at the mapper phase used to perform the process of data transformation, feature selection, and data augmentation, whereas the reducer function is located at the reducer phase used to achieve the process of student performance prediction. In the data transformation process, log transformation is used to transform the data to some other forms. The feature selection process is done with DL distance to select the unique features to increase the quality of prediction performance. Here, the Bootstrap method is employed to perform the data augmentation process, and finally, the DNFN is used to achieve the student performance prediction model. The training process of DNFN is carried out by the proposed FCMVO algorithm. The performance obtained by the proposed method is shown in terms of MSE, RMSE and MAE with the values of 0.338 3, 0.581 7 and 0.391 5, respectively. Anyhow, the performance of the proposed system is not evaluated using the publicly available dataset. This limitation will be considered in the future extension of the work. Also, the performance of the proposed system will be further enhanced by considering some other deep learning classifiers.

## References


[1] V. Ravuri, S. Vasundra. Moth-flame optimization-bat optimization: Map-reduce framework for big data clustering using the Moth-flame bat optimization and sparse Fuzzy C-means. *Big Data*, vol. 8, no. 3, pp. 203–217, 2020. DOI: 10.1089/big.2019.0125.

[2] A. N. Jadhav, N. Gomathi. DIGWO: Hybridization of dragonfly algorithm with improved grey wolf optimization algorithm for data clustering. *Multimedia Research*, vol. 2, no. 3, pp. 1–11, 2019.

[3] A. Sheshasaayee, M. N. Bee. E-learning: Mode to improve the quality of educational system. *Smart Computing and Informatics*, S. C. Satapathy, V. Bhateja, S. Das, Eds., Singapore: Springer, pp. 559−566, 2018. DOI: 10.1007/978-981-10-5547-8_58.

[4] A. Dutt, M. A. Ismail, T. Herawan. A systematic review on educational data mining. *IEEE Access*, vol. 5, pp. 15991–16005, 2017. DOI: 10.1109/ACCESS.2017.2654247.

[5] H. Kaur, G. Bathla. Student performance prediction using educational data mining techniques. *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol. 4, no. 12, pp. 93–97, 2018.

[6] G. Lassibille, L. N. Gómez. Why do higher education students drop out? Evidence from Spain. *Education Economics*, vol. 16, no. 1, pp. 89–105, 2008. DOI: 10.1080/09645290701523267.

[7] K. Bhatnagar, S. C. Gupta. Investigating and modeling the effect of laser intensity and nonlinear regime of the fiber on the optical link. *Journal of Optical Communications*, vol. 38, no. 3, pp. 341–353, 2017. DOI: 10.1515/joc-2016-0041.

[8] M. R. M. VeeraManickam, M. Mohanapriya, B. K. Pandey, S. Akhade, S. A. Kale, R. Patil, M. Vigneshwar. Map-Reduce framework based cluster architecture for academic student's performance prediction using cumulative dragonfly based neural network. *Cluster Computing*, vol. 22, no. 1, pp. 1259–1275, 2019. DOI: 10.1007/s10586-017-1553-5.

[9] J. F. Chen, Q. H. Do. A cooperative cuckoo search–hierarchical adaptive neuro-fuzzy inference system approach for predicting student academic performance. *Journal of Intelligent & Fuzzy Systems*, vol. 27, no. 5, pp. 2551–2561, 2014. DOI: 10.3233/IFS-141229.

[10] L. H. Son, H. Fujita. Neural-fuzzy with representative sets for prediction of student performance. *Applied Intelligence*, vol. 49, no. 1, pp. 172–187, 2019. DOI: 10.1007/s10489-018-1262-7.

[11] O. Yildiz, A. Bal, S. Gulsecen. Improved fuzzy modelling to predict the academic performance of distance education students. *The International Review of Research in Open and Distance Learning*, vol. 14, no. 5, pp. 144–165, 2013. DOI: 10.19173/irrodl.v14i5.1595.

[12] K. Srinivas. Prediction of E-learning efficiency by deep learning in E-khool online portal networks. *Multimedia Research*, vol. 3, no. 4, pp. 12–23, 2020. DOI: 10.46253/j.mr.v3i4.a2.

[13] S. L. Karri, L. C. De Silva, D. T. C. Lai, S. Y. Yong. Identification and classification of driving behaviour at signalized intersections using support vector machine. *International Journal of Automation and Computing*, vol. 18, no. 3, pp. 480–491, 2021. DOI: 10.1007/s11633-021-1295-y.

[14] K. Dineva, T. Atanasova. Architectural ML framework for IoT services delivery based on microservices. In *Proceedings of the 23rd International Conference on Distributed Computer and Communication Networks*, Springer, Moscow, Russia, pp. 698−711, 2020. DOI: 10.1007/978-3-030-66471-8_53.

[15] A. Alqwadri, M. Azzeh, F. Almasalha. Application of machine learning for online reputation systems. *International Journal of Automation and Computing*, vol. 18, no. 3, pp. 492–502, 2021. DOI: 10.1007/s11633-020-1275-7.

[16] A. Adhikari, C. N. Aher. SVM: The qualitative and quantitative monolithic predictor. *International Journal of Science and Research*, vol. 2, no. 11, pp. 270–273, 2013.

[17] S. H. Wang, Y. D. Lv, Y. X. Sui, S. Liu, S. J. Wang, Y. D. Zhang. Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. *Journal of Medical Systems*, vol. 42, no. 1, Article number 2, 2018. DOI: 10.1007/s10916-017-0845-x.

[18] A. M. Shahiri, W. Husain, N. A. Rashid. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, vol. 72, pp. 414–422, 2015. DOI: 10.1016/j.procs.2015.12.157.

[19] M. M. Beno, I. R. Valarmathi, S. M. Swamy, B. R. Rajakumar. Threshold prediction for segmenting tumour from brain MRI scans. *International Journal of Imaging Systems and Technology*, vol. 24, no. 2, pp. 129–137, 2014. DOI: 10.1002/ima.22087.

[20] Y. D. Zhang, V. V. Govindaraj, C. S. Tang, W. G. Zhu, J. D. Sun. High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 2012–2021, 2019. DOI: 10.1166/jmihi.2019.2692.

[21] T. T. Dien, S. H. Luu, N. Thanh-Hai, N. Thai-Nghe. Deep learning with data transformation and factor analysis for student performance prediction. *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 711–721, 2020. DOI: 10.14569/IJACSA.2020.0110886.

[22] A. V. Dhumane. Examining user experience of eLearning systems using E-Khool learners. *Journal of Networking and Communication Systems*, vol. 3, no. 4, 2020.

[23] P. Shirsat. Developing deep neural network for learner performance prediction in EKhool online learning platform. *Multimedia Research*, vol. 3, no. 4, pp. 24–31, 2020. DOI: 10.46253/J.MR.V3I4.A3.

[24] A. T. Corbett, J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, vol. 4, no. 4, pp. 253–

278, 1994. DOI: 10.1007/BF01099821.

[25] D. Liu, H. H. Dai, Y. P. Zhang, Q. P. Li, C. P. Zhang. Deep knowledge tracking based on attention mechanism for student performance prediction. In *Proceedings of the 2nd International Conference on Computer Science and Educational Informatization*, IEEE, Xinxiang, China, pp. 95–98, 2020. DOI: 10.1109/CSEI50228.2020.9142472.

[26] R. J. Williams, D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989. DOI: 10.1162/neco.1989.1.2.270.

[27] D. Liu, Y. P. Zhang, J. Zhang, Q. P. Li, C. P. Zhang, Y. Yin. Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction. *IEEE Access*, vol. 8, pp. 194894–194903, 2020. DOI: 10.1109/ACCESS.2020.3033200.

[28] S. Minn. BKT-LSTM: Efficient Student Modeling for knowledge tracing and student performance prediction. [Online], Available: https://arxiv.org/abs/2012.12218, 2021.

[29] B. H. Kim, E. Vizitei, V. Ganapathi. GritNet: Student performance prediction with deep learning. https://arxiv.org/abs/1804.07405, 2018.

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

[31] Q. Liu, Z. Y. Huang, Y. Yin, E. H. Chen, H. Xiong, Y. Su, G. P. Hu. EKT: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 100–115, 2021. DOI: 10.1109/TKDE.2019.2924374.

[32] Y. Su, Q. W. Liu, Q. Liu, Z. Y. Huang, Y. Yin, E. H. Chen, C. Ding, S. Wei, G. P. Hu. Exercise-enhanced sequential modeling for student performance prediction. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, AAAI, New Orleans, USA, pp. 2435–2443, 2018.

[33] O. W. Adejo, T. Connolly. Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, vol. 10, no. 1, pp. 61–75, 2018. DOI: 10.1108/JARHE-09-2017-0113.

[34] S. Sahebi, Y. R. Lin, P. Brusilovsky. Tensor factorization for student modeling and performance prediction in unstructured domain. In *Proceedings of the 9th International Conference on Educational Data Mining*, Raleigh, USA, pp. 502–506, 2016.

[35] W. J. Buchanan, P. Griffiths, R. Macfarlane. Distance measurement methods for improved insider threat detection. *Security and Communication Networks*, vol. 2018, Article number 5906368, 2018. DOI: 10.1155/2018/5906368.

[36] Boot Strap Method, [Online], Available: https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method, February, 2021.

[37] S. Javaid, M. Abdullah, N. Javaid, T. Sultana, J. Ahmed, N. A. Sattar. Towards buildings energy management: Using seasonal schedules under time of use pricing tariff via deep neuro-fuzzy optimizer. In *Proceedings of the 15th International Wireless Communications & Mobile Computing Conference*, IEEE, Tangier, Morocco, pp. 1594–1599, 2019. DOI: 10.1109/IWCMC.2019.8766673.

[38] P. R. Bhaladhare, D. C. Jinwala. A clustering approach for the $\ell$-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm. *Advances in Computer Engineering*, vol. 2014, Article number 396529, 2014. DOI: 10.1155/2014/396529.

[39] I. Benmessahel, K. Xie, M. Chellal. A new competitive multiverse optimization technique for solving single-objective and multiobjective problems. *Engineering Reports*, vol. 2, no. 3, Article number e12124, 2020. DOI: 10.1002/eng2.12124.

**Amlan Jyoti Baruah** received the B. Tech. degree in computer science and engineering from North Eastern Regional Institute of Science and Technology, Arunachal Pradesh, India in 2011, received the M. Tech. degree in computer science and engineering from Kalinga Institute of Industrial Technology (Deemed University), India in 2013. He is also a Ph. D. degree candidate from Jorhat Engineering College, Assam Science and Technology, India. He is currently working as an assistant professor in Department of Computer Science and Engineering, Assam Kaziranga University, India. He has around 8 years of teaching experience.

His research interests include educational data mining, deep learning and artificial intelligence.

E-mail: amlanbaruah2007@gmail.com (Corresponding author)

ORCID iD: 0000-0001-8121-2969

**Siddhartha Baruah** received the B. Sc. degree with honors in physics from Science College Jorhat (Currently known as Jorhat Institute Of Science and Technology (JIST)), India in 1987, the MCA degree in computer application from Jorhat Engineering College, India in 1990, and the Ph.D. degree in computer science from Guwahati University, India in 2010. He is currently working as a professor in Department of Computer Application, Jorhat Engineering College, India. He has around 28 years of teaching experience and 12 years of research experience. He has completed several projects in modernisation and removal of obsolenscence (MODROB) approved by All India Council for Technical Education (AICTE) and played the key role in starting Ph. D. Program in MCA Department of JEC in 2018. He has published several papers as well as attended different international conferences in India and Abroad.

His research interests include embedded system, educational data mining, deep learning and artificial intelligence.

E-mail: sbaruain@gmail.com