Journal of Scientific & Industrial Research
Vol. 82, November 2023, pp. 1152-1164
DOI: 10.56042/jsir.v82i11.40

Predicting Student Performance with Adaptive Aquila Optimization-based Deep Convolution Neural Network

Jiayi Lu[1,2], Vineeta Singh[3]*, Suruchi Singh[3], Alok Kumar[3], Saurabh Pandey[4], Deepak Kumar Verma[3] & Vandana Dixit Kaushik[5]

[1]Arts School of Shanghai University of Sport, Shanghai 200 438, China

[2]Center for Post-doctoral Studies in Nanjing University of the Arts, Nanjing 210 013, China

[3]Department of Computer Science and Engineering, School of Engineering and Technology (Formerly known as UIET Kanpur), Chhatrapati Shahu Ji Maharaj University, Kalyanpur, Kanpur 208 012, Uttar Pradesh, India

[4]IT & EMPC Department, Vardhman Mahaveer Open University, Kota, Rajasthan, India

[5]Department of Computer Science & Engineering, HBTU East Campus, Nawabganj, Kanpur, Uttar Pradesh 208 002, India

Received 22 March 2023; revised 09 October 2023; accepted 19 October 2023

Predicting student performance is the major problem for enhancing the educational procedures. A level of student's performance may be influenced by several factors like job of parents, sexual category and average scores obtained in prior years. Student's performance prediction is a challenging chore, which can help educational staffs and students of educational institutions to follow the progress of students in their academic activities. Student performance enhancement and progress in educational quality are the most vital part of educational organizations. Presently, it is essential for an educational organization to predict the performance of students. Existing methods utilized only previous student performances for prediction without including other significant behaviors of students. For addressing such problems, a proficient model is proposed for prediction of student performance utilizing proposed Adaptive Aquila Optimization-allied Deep Convolution Neural Network (DCNN). In this process, data transformation is initiated using the Yeo-Johnson transformation method. Subsequently, feature selection is performed using Fisher Score to identify the most relevant features. Following feature selection, data augmentation techniques are applied to enhance the dataset. Finally, student performance is predicted through the utilization of a DCNN, with a focus on fine-tuning the network parameters for optimal performance. This fine-tuning is achieved through the use of the Adaptive Aquila Optimizer (AAO), ensuring the network is poised to deliver the best possible results in predicting student outcomes. Proposed AAO-based DCNN has achieved minimal error values of Mean Square Error, Root Mean Square Error, Mean Absolute Error, Mean Absolute Percentage Error, Mean Absolute Relative Error, Mean Squared Relative Error, and Root Mean Squared Relative Error, respectively.

Keywords: Adaptive concept, Aquila optimizer, DCNN, KNN, Student performance prediction model

Introduction

Within the realm of Knowledge Discovery in Databases (KDD),[1,2] there exists an emerging focus on Educational Data Mining (EDM).[3] EDM represents a progressive trend dedicated to unearthing valuable patterns and deriving meaningful insights from educational information systems.[4] These systems encompass a wide range of educational data, including but not limited to admissions, registrations, and course management systems. This methodology extends its reach across diverse educational levels, from primary schools to universities. EDM offers a novel approach to harnessing the potential knowledge residing within educational datasets, ultimately contributing to improved educational practices. This research field involved researchers concentrating over creating helpful information to assist educational institutions for managing the students or for facilitating students for dealing with education, better deliverables and performance improvement.[5] EDM is considered to assess data developed in an educational group utilizing different systems. Its main objective is to develop systems for improving experience in learning and efficacy of institutions.[6,7] Due to educational processes digitization, universities are developing enormous quantity of data relevant to students in electronic structure. It is significant for them to efficiently[8] alter this huge compilation of data into knowledge that assists administrators and teachers for analyzing it to improve decision making.[9] Analyzing students data and information for

*Author for Correspondence
E-mail: vineeta.singh.cs@gmail.com

categorizing students, mostly concentrates on understanding and assessing of students educational data,[10] which signifies their performances in education and develops certain rules, categorizations and predictions for assisting students in their prospective academic performance.

In educational data mining i.e. EDM, prediction of student performance is major consideration issue in training and education field. Subsequently this outcome facilitates pupils for selection of courses while opting for fitted study plan. Furthermore, prediction of student's performance[11] assists educational administrators and teachers to reveal about monitoring students and support them to finish their programs with superior outcomes.[12] The educational administrators can provide them suitable guidance to evade poor outcomes and also prepare students by identifying the dependencies for last exams after obtaining the results. Each year a number of students are lagging behind due to flaw of appropriate monitoring and guidance. Hence, it would be very obliging for students and educational administrators.[13,14] Furthermore, EDM and deep learning[15,16] plays a vital part in finding feeble students of an educational institution in an academic way and assist them by developing various suggestion systems for enhancing their performance. Thus, the technologies guide the students for their prospective planning by creating valuable patterns from their information history. Presently, there are several techniques for predicting a performance of students in which data mining is the major accepted approaches that is expansively utilized in an area of education. In addition, researchers have not been adequately compared deep learning techniques[17] with other classical techniques.

The dynamic landscape of education grapples with the challenges of burgeoning classroom sizes and diverse learning needs, posing significant hurdles for educators in tailoring teaching methods to individual students. Accurate predictions offer a transformative potential by enabling timely interventions, personalized support, and efficient resource allocation, ultimately enhancing the student learning experience and academic success.[18] Explainable student performance prediction via personal attention was the main base of research, moreover finding the cause why an student fail.[19] Education, with its multifaceted goals of enhancing student learning, improving teaching methodologies, and providing tailored support, aims to empower students for success in an ever-evolving world. However, it faces challenges such as high student-teacher ratios, diverse learning needs, and the urgency of timely interventions. Embracing data-driven approaches, predictive analytics emerges as a game-changing concept, allowing educators to forecast student performance and adapt instruction to individual needs.[20] Deep learning techniques, particularly Convolution Neural Networks (CNNs), stand out for their ability to decipher intricate patterns within diverse data sources. In our study, we introduce the AAO-based DCNN, emphasizing its potential to address the critical need for precise student performance prediction and advance educational analytics by developing a predictive model using the Aquila network.

This research work devises an efficient technique for student performance prediction applying proposed AAO-based Deep CNN. In this proposed technique source data has been fed to the data transformation stage that is processed via Yeo-Johnson. A transformed data is further injected in feature selection module for selecting the features utilizing fisher score. Thereafter, the selected features are subjected to data augmentation phase for increasing the dimensionality of data using oversampling technique. Finally, prediction of student performance is conducted employing DCNN, wherein devised AAO is utilized for tuning the network. However, the proposed AAO is presented by integration of Adaptive concept with AO. A significant contribution of this paper is revealed below:

Adaptive Aquila Optimization: Adaptive Aquila Optimization is a relatively new meta-heuristic optimization algorithm inspired by the hunting behavior of eagles, particularly the Aquila genus. This algorithm mimics the hunting strategy and foraging behavior of eagles in its approach to solving optimization problems. It is a bio-inspired method for solving complex optimization problems across various domains. In Aquila Optimization, the term adaptive refers to the algorithm's ability to adjust its parameters and strategies based on the problem it's solving.

Proposed AAO-based DCNN: An efficient technique for student performance prediction is proposed utilizing AAO based DCNN. Here, deep CNN classifier is trained utilizing newly proposed optimization, namely AAO that is an incorporation of adaptive concept and AO.

Literature Review

Here, we have added a review of the literature on old and modern methods for forecasting student performance that encourages the investigators to create a successful method for doing so.

Turabieh *et al.*[1] proposed an enhanced Harris Hawks Optimizer with k-Nearest Neighbors (KNN) for the prediction of student performance. This method controlled the diversity of population and prevented the earlier convergence of HHO by passing present populations with novel solution, when every solution was belonged to single cluster and trapped in local optimum but still it did not simulated the improvement of HHO over several original industrial issues. Delianidi *et al.*[18] modeled the performance of student as a dynamic issue and comparison was done on two main classes of dynamic neural structures for their solution, like Time Delay Neural Networks (TDNN) as well as Recurrent Neural Networks (RNN) for tracing the knowledge status of student. It was proven that RNN is better when compared to TDNN technique in every datasets but, an extensive experiment with many benchmark datasets was not conducted. Niu *et al.*[19] introduced a newly designed a prediction technique for assessing the student performance possessed with personalized attention, abbreviated as ESPA were presented for mapping the prediction of student performance issue in education to the suggestion system. This model predicted the reasonable failure of students but it did not construct high accurate and time model of students for effective way of representation. Sood *et al.*[20] utilized a hybridization technique comprising of cluster-based linear discriminate analysis i.e. ANN for providing the future plans of students with motive comments. This technique was utilized for assisting the teachers and students to enhance the performance of students but it did not verify the validity in finding probable failure level students. Hussain & Khan[21] focused on grade and marks prediction, which could have enhanced educational quality. The advantage lay in its potential to inform administrative and teaching staff improvements. However, a disadvantage may have been the challenge of ensuring data quality and interpretability in complex models.

Major Challenges

Diverse problems faced by traditional techniques for prediction of student performance are expounded below-

HHO-based KNN algorithm[1] was devised for predicting the performance of student. However, it did not examine various types of injection techniques on basis of distributed solution history.

ESPA was designed by employing integration of students profile and previous information of corresponding courses. Even so, it was founded that the real environmental data was extremely imbalanced.[19]

In addition, only vital features are regarded for student performance prediction but other features, like cultural and social attributes of students and time taken to finish a particular chore are not considered.[20]

Proposed Methodology for Predicting Student Performance

Predicting the student performance is a significant chore to many educational institutions, teachers and students themselves. It is utilized to construct earlier alerting systems and personalized suggestion systems for enhancing the learning knowledge of students. An effective approach using AAO-based deep CNN is proposed for forecasting the student's performance. In the beginning, input data is obtained from a database and sent to data transformation, where a Yeo-Johnson process is carried out. Thereafter, feature selection is carried out using fisher-score and then, data is augmented. Finally, performance of students is predicted utilizing deep CNN, where the proposed AAO is used for training the network. Howsoever, the introduced AAO is designed by an incorporation of adaptive concept and AO. We have explained through Fig. 1 the diagrammatic view of student's performance prediction utilizing introduced AAO based Deep CNN.

Input Data

Here we are taking inputs from the dataset[22] and mathematical representation involves like below,

$$X = \{D_1, D_2 .... D_i ,..... \ D_a\} \qquad \ldots (1)$$

Here, the total count of training samples involved in training database is signified by a, whereas X represents training dataset and D_i indicates the i^{th} input data.

Data Transformation using Yeo-Johnson

The input data D_i having dimension of $u \times v$ when subjected to data transformation, thus formed transformed data using Yeo-Johnson is expressed as T_i with same dimension of $u \times v$. Data transformation

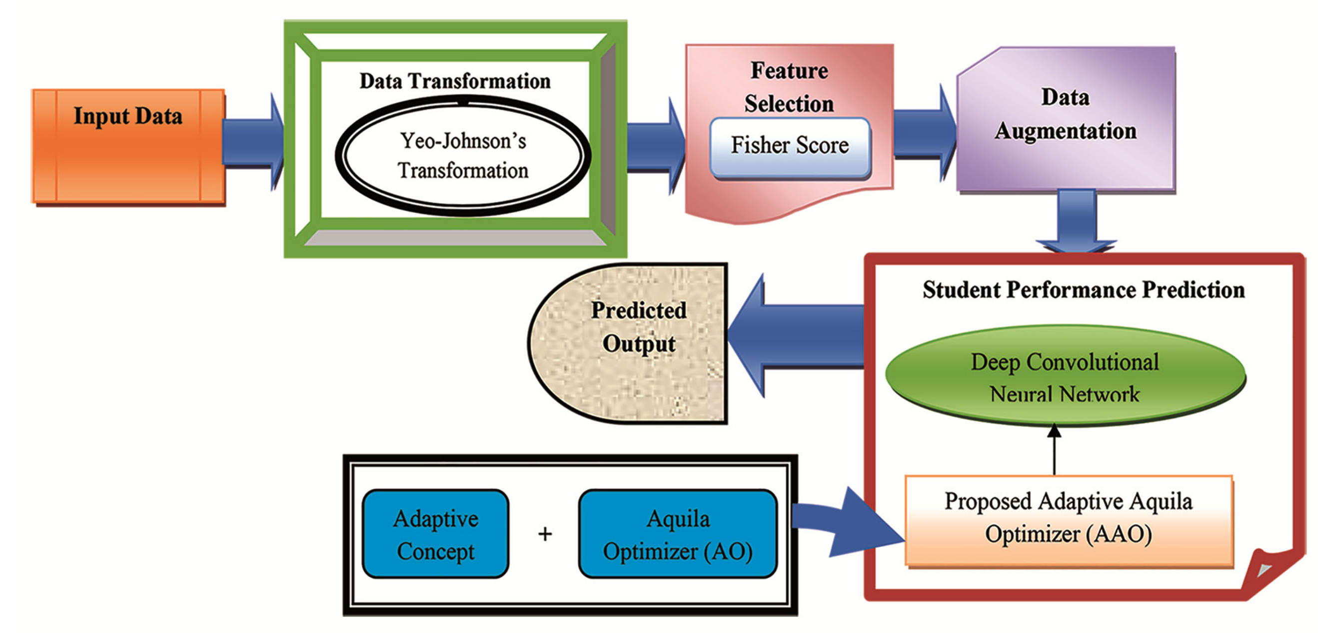


Fig. 1 — Diagrammatic view of Student's Performance Prediction utilizing Introduced AAO based Deep CNN

is also termed as consolidation of data, in which selected data is transformed into other forms suitable for further procedure. Here, Yeo-Johnson[23] is employed for smoothing circumstance to integrate the transformations for negative and positive examination. Yeo-Jhonson transformation is expressed by,

$$T_i = \begin{cases} \dfrac{(R+1)^{\tau}-1}{\tau}, & \tau \neq 0 \ and \ R \geq 0, \\ Ln\left(R+1\right), & \tau = 0 \ and \ R \geq 0, \\ \dfrac{-\left(\left(-R+1\right)^{2-\tau}-1\right)}{2-\tau}, & \tau \neq 2 \ and \ R < 0, \\ Ln\left(-R+1\right), & \tau = 2 \ and \ R < 0 \end{cases} \quad \ldots (2)$$

where, Yeo-Johnson transformation takes an input variable R, T_i is transformed variable and this transformation of variable is suitable for validating right and left tilt during $\tau < 1 and \tau > 1$ whereas the linear affiliation is attained when $\tau = 1$. Moreover, transformations of Yeo-Johnson can seize a log mean standardization properties afterwards an inverse transformation since T_i is invertible.

Fisher Score Coefficient-based Feature Selection

One of the primary objectives of feature selection is to identify a subset of input variables by eliminating irrelevant or redundant features that do not contribute to the predictive accuracy. This process is crucial for enhancing the overall quality of the model and expediting the learning process. In this study, feature selection is carried out using the Fisher score,[24] a widely adopted technique. The Fisher score method independently evaluates each feature based on their individual scores, adhering to Fisher's criteria, which leads to the selection of an optimal subset of features. The transformed data T_i is then given to feature extraction phase and the expression of fisher score for i^{th} feature selection is given below.

$$S\left(z^i\right) = \frac{\sum_{l=1}^{k} p_l \left(R_l^{\,i} - \eta^i\right)^2}{\left(\lambda^i\right)^2} \quad \ldots (3)$$

Here,

$$\left(\lambda^i\right)^2 = \sum_{l=1}^{k} p_l \left(\lambda^i{}_l\right)^2 \quad \ldots (4)$$

where, $S\left(z^i\right)$ denotes the fisher score for a variable z^i, k represents the number of categories, p_l denotes prior probability or weight associated with the l^{th} class, $R_l^{\,i}$ denotes the mean of the i^{th} feature within the l^{th} class, η^i denotes reference or global mean of the i^{th} feature across all classes and λ^i denotes the scaling factor associated with the variable z^i. When fisher score is calculated for each feature, top ranked features is denoted as u and w sets the threshold for the significance of those features based on their Fisher's scores. Features that have a score greater than or equal to $u \times w$ are considered

significant and are selected as the top features. w controls the level of significance required for a feature to be considered among the top features. Essentially, it serves as a scaling factor for the Fisher's scores, helping to determine which features are considered the most discriminative.

Data Augmentation

The data augmentation is commonly utilized for training data to decrease the risk of over fitting. The chosen feature F_i of dimension $u \times w$ is next subjected to a process of data augmentation, which is carried out via an oversampling approach known as SMOTE (Synthetic Minority Over-sampling approach). The augmented data obtained is denoted by A_i and the data size is incremented to $r \times w$. Here, u is enhanced to size r by means of oversampling. Hence, within each column, we calculate the maximum and minimum values, and subsequently, random samples are generated within the same interval.

Student Performance Prediction using Deep CNN

After data augmentation, the augmented data is subjected to prediction step and the student performance is predicted utilizing Deep CNN. However, the Deep CNN is trained using the proposed AAO, a novel optimization technique that combines adaptive concepts with the AO.

Architecture of Deep CNN: A CNN can be conceptualized as an Artificial Neural Network (ANN) comprising convolutional, non-linear, fully connected, and pooling layers, thus giving rise to the structure of a deep CNN.[25] Design of Deep Convolutional Neural Network is represented through Fig. 2. Layers of Deep CNN are elucidated as here:

i) Convolutional Layers: In this process, multiple filters traverse across the input data. The output layer is generated by aggregating the element-wise products of these filters with the accessible input field. An expression is given by,

$$\left(A_o{}^q\right)_{s,t} = \left(W_o{}^q\right)_{s,t} + \sum_{b=1}^{c_1^{b-1}} \sum_{\beta=c_1^q}^{c_1^q} \sum_{w=-c_2^q}^{c_2^q} \left(\alpha_{0,b}{}^q\right)_{w,\beta} * \left(A_i{}^{q-1}\right)_{s+\beta,t+w} \quad \ldots (5)$$

Here, $*$ signifies the convolutional operator, fixed feature map is expressed through $\left(A_o{}^q\right)_{s,t}$ and q^{th} convolutional layers output is centered as (s,t). Here input for k^{th} convolutional layer is generated through an output of the prior $(q-1)^{th}$ layer. Considering convolutional layers weights be $\alpha_{0,b}{}^q$ that is the weight of q^{th} convolutional layer whereas $W_o{}^q$ is the bias of q^{th} convolutional layer. By assuming f convolutional layers, $(1 \le k \le f)$ similarly b, β and w

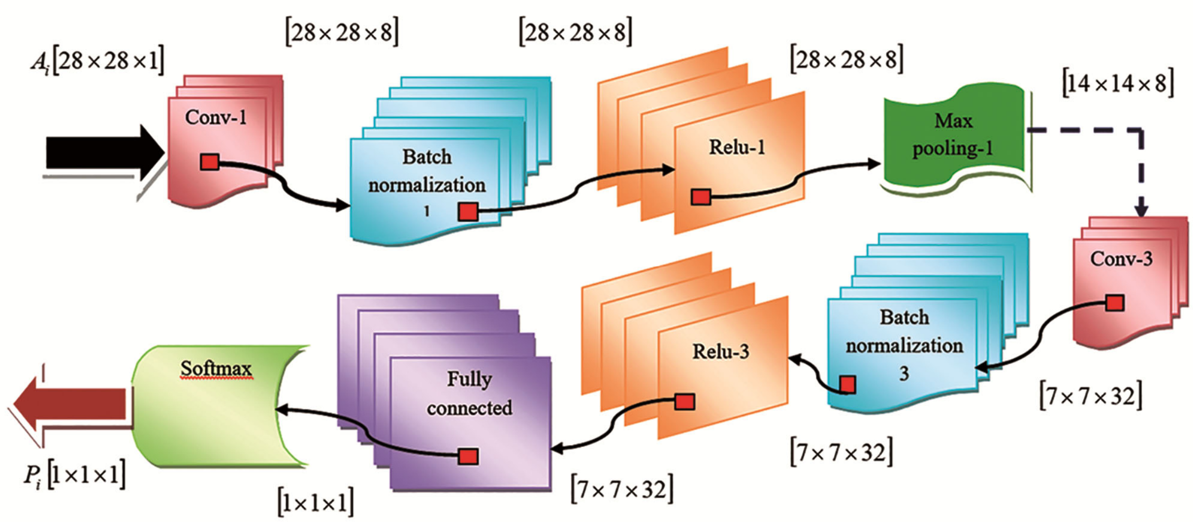


Fig. 2 — Deep CNN Architecture

represents a feature maps, this serves as the output of each convolutional filter individually.

An element-wise activation function is used by the rectified linear unit, or ReLU layer, whereas an activation function of $(k-1)^{th}$ prior layer is output obtained via k^{th} ReLU layer denoted by Eq. (6),

$$A_o^{\,q} = Afn\left(A_o^{\,q-1}\right) \quad \ldots (6)$$

A consequence of ReLU layer relates a deep CNN speed that is augmented and gives the capability to act with several networks. ReLU is commonly utilized nonlinearity in several fields. The ReLU can be expressed by,

$$\operatorname{Re}LU = \begin{cases} 0, & if\ A_i < 0 \\ A_i, & if\ A_i \geq 0 \end{cases} \quad \ldots (7)$$

where, A is a parameter, which can be any real number and i is a parameter that controls the exponentiation of A

ii) Pooling Layer: This layer approximately decreases inputs dimension though it is non-parametric it executes a permanent functions.

iii) Fully Connected Layers: Subsequently, the data from the pooling layer is transmitted to the fully connected layer. This layer functions as a classifier and is crucial for capturing the non-linear combination of features. It can be expressed as follows, where the resulting output from the deep CNN is denoted as P_i.

$$F_o^{\,q} = \mu(A_o^{\,q}) \text{ with } \sum_{b=1}^{c_1^{b-1}} \sum_{\beta=c_1^{\,q}}^{c_1^{\,q}} \sum_{w=-c_2^{\,q}}^{c_2^{\,q}} \left(\alpha_{0,b}^{\ \ q}\right)_{w,\beta} \left(A_i^{\,q-1}\right)_{s+\beta,t+w} \quad \ldots (8)$$

iv) Aquila Position Encoding: To obtain an optimal solution, deep CNN is tuned until attaining an optimum solution, where η is denoted as learning parameter of CNN.

v) Fitness Function: A fitness function is defined as a differentiation among aimed output as well as an output acquired from deep CNN. It is formulated as,

$$\delta = \frac{1}{a} \sum_{i=1}^{a} \left[T_i - P_i\ \right]^2 \quad \ldots (9)$$

where, a indicates a total count of samples, T_i and P_i reveals target output as well as a deep CNN output.

Proposed Adaptive Aquila Optimization (AAO)

Aquila is one of the most famous prey birds whereas its distinct actions can be experimented naturally. It is mainly studied across global due to its bravery in hunting and very clever and skilled hunters next to humans. Aquila Optimizer (AO) [26] creates the behavior of Aquila during hunting, which shows the action of hunt in each step. Aquila-genus eagles, specifically, are the inspiration for the adaptive Aquila optimization meta-heuristic optimization algorithm. This method solves optimization problems by imitating the foraging and hunting behavior of eagles. It is an approach inspired by biology for resolving challenging optimization issues in numerous fields. In Aquila Optimization, the term adaptive refers to the algorithm's ability to adjust its parameters and strategies based on the problem it's solving. The advantage of adaptiveness is that it enables Aquila to effectively tackle a wide range of optimization problems by tailoring its approach to the specific problem characteristics, ultimately improving efficiency and solution quality.

Step 1: Solution Initialization: While a rule of optimization begins with the population of candidate solutions (R) as shown in Eq. (10), AO is a population-based technique. The finest solution is obtained as optimum solution in every iteration and it is signified by,

$$R = \left\{R_1, R_2, \ldots\ldots\ldots\ldots, R_i, \ldots\ldots R_n\right\} \quad \ldots (10)$$

Here, Y signifies a present candidate solution that is produced arbitrarily by Eq. (11), R_i expresses the position of i^{th} solution, total count of candidate solution of population is denoted by n and d reveals the problem dimension. It is given as,

$$R_{ig} = rand \times \left(U_g - L_g\right) + L_g, \quad i = 1,2,3\ldots\ldots\ldots n \ \ g-1,2,3,\ldots\ldots\ldots\ldots. d \quad \ldots (11)$$

Here, $rand$ mentions random number, U_g and L_g indicates g^{th} upper and lower bound of given problem, which are specific limits or constraints within which the variables or parameters of the problem should operate. These bounds are defined for each variable or parameter in an optimization problem to ensure that the solution remains within a feasible and meaningful range.

Step 2: Evaluate Objective Function: In this step, a variation among aimed output and an output attained from deep CNN is computed using Eq. (9).

Step 3: Expanded Exploration (R_1): In this method, AO expansively explores from higher soar to find out a prey region of search space and it is represented by,

$$R_1(h+1) = R_{best}(h) \times \left(1 - \frac{h}{H}\right) + \left(R_M(h) - R_{best}(h) * rand\right) \quad \ldots (12)$$

Here, $R_1(h+1)$ is the subsequent iteration solution of h that is produced by initial search method (R_1). R_{best} is the best solution obtained until h^{th} iteration, which shows the approximate location of prey. $\left(1-\frac{h}{H}\right)$ is utilized to control expanded search by means of number of iterations. $R_M(h)$ indicates the location mean value of present solution connected at h^{th} iteration. $rand$ represents a random value in between 0 as well as 1, the current iteration is denoted by h while H depicts the highest iteration count.

Step 4: Narrowed Exploration (R_2): AO narrowly discovers the chosen region of targeted prey in preparation for attack. It is expressed by,

$$R_2(h+1) = R_{best}(h) \times Levy(N) + R_G(h) + (x-y) * rand \quad \ldots (13)$$

Here, $R_2(h+1)$ is the subsequent iteration solution of h that is produced by secondary search method (R_2). N is a dimension space and levy flight distribution function has been depicted via $Levy(N)$. $R_G(h)$ depicts a randomly selected solution, which is taken in $[1 \;\; n]$ range at i^{th} iteration and x as well as y are utilized to represent a spiral appearance in a search.

$$x = f \times \cos(\theta) \quad \ldots (14)$$

$$y = f \times \sin(\theta) \quad \ldots (15)$$

Here

$$f = f_1 + V \times B_1 \quad \ldots (16)$$

f_1 is the value ranging among 1 and 20 for permanent count of search sequences. Here V is a very small value set as 0.00565 while B_1 denotes an integer number in between 1 to search space with length denoted as d.

Step 5: Expanded Exploitation (R_3): Here AO employs a selected region of the target to attain nearer of prey and assault. It is expressed by,

$$R_3(h+1) = \left(R_{best}(h) - R_M(h)\right) \times \varepsilon - rand + \left((U-L) \times rand + L\right) \times \varphi \quad \ldots (17)$$

Here, $R_3(h+1)$ is the subsequent iteration solution of h that is produced by thirdly search method (R_3) and $R_{best}(h)$ signifies the fairly accurate position of prey until i^{th} iteration. ε and ϕ are an exploitation adjustment parameters set to a least value (0.1).

Step 6: Narrowed Exploitation (R_4): At last, AO assaults the prey in the final position and is presented by,

$$R_4(h+1) = Q \times R_{best}(h) - \left(J_1 \times R(h) \times rand\right) - J_2 \times Levy(N) + rand \times J_1 \quad \ldots (18)$$

Here, $R_4(h+1)$ is the subsequent iteration solution of h that is produced by fourth search method (R_4). Q indicates quality function utilized to balance a search strategies that is computed utilizing Eq. (19), J_1 signifies several movements of AO, which are employed to follow the prey at time of elope that is created wielding Eq. (20). J_2 represents to chase the prey at the time of elope from the first position (1) to the last position (h) that is developed utilizing Eq. (21). $R(h)$ is the present solution at h^{th} iteration.

$$Q(h) = h^{\frac{2 \times rand(\;) - 1}{(1-H)^2}} \quad \ldots (19)$$

$$J_1 = 2 \times rand(\;) - 1 \quad \ldots (20)$$

$$J_2 = 2 \times \left(1 - \frac{h}{H}\right) \quad \ldots (21)$$

Therefore, $rand$ is made adaptive and an update equation is represented by,

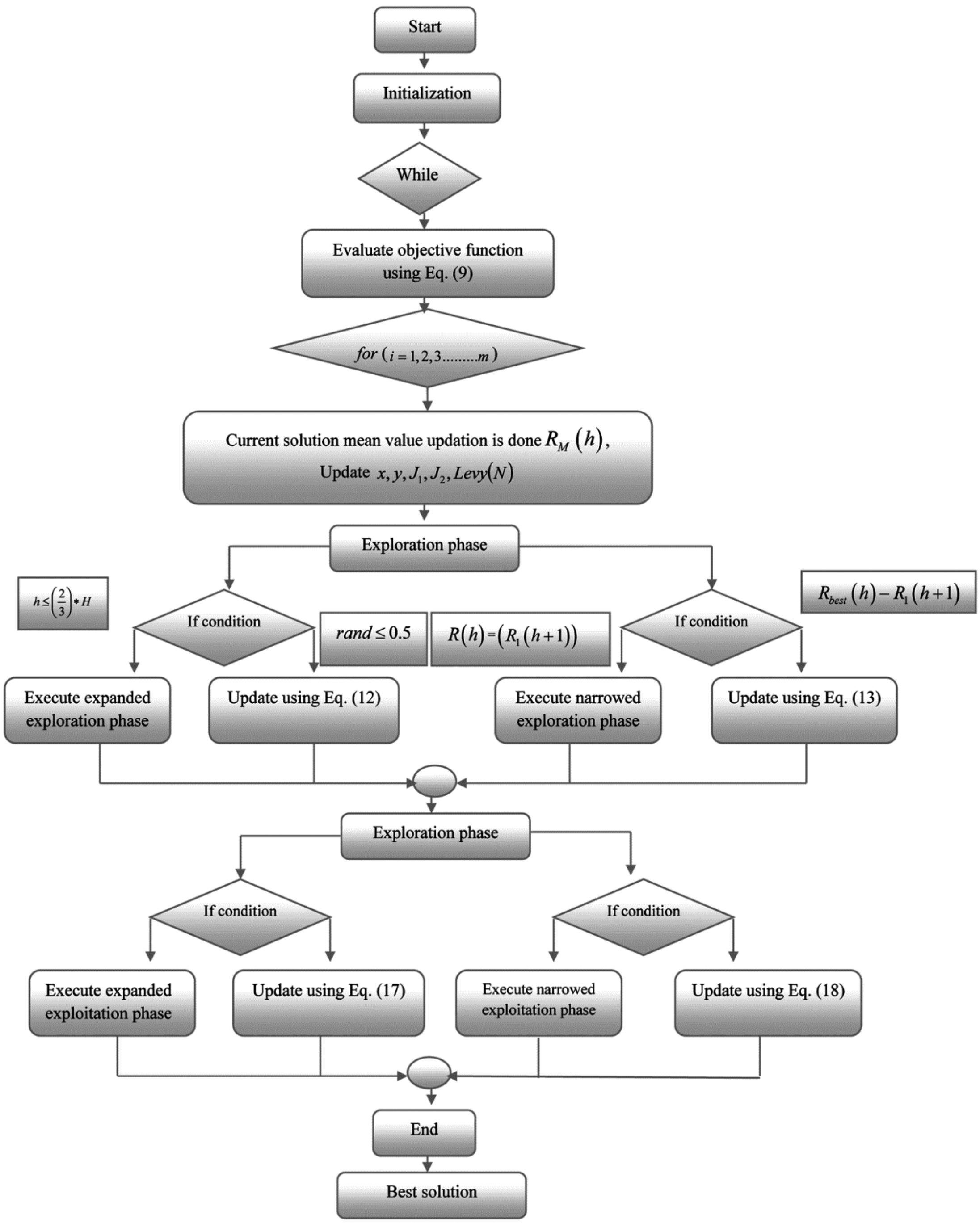
$$rand = 1 - \frac{\left(C * \left(1 - \frac{h}{H}\right)\right)}{n} \quad \ldots (22)$$

Here, h denotes current iteration, H signifies maximum iteration, m indicates population size and C reveals parameter constant having a value of 0.5 whereas $rand$ value ranges from 0 and 1.

Step 7: Termination: Aforementioned steps are repeated until it attains an optimum solution. The pseudo code of AAO is represented through Algorithm 1

Experimental Results and Discussions

Performance measurements are used in this section to characterize the results of the newly introduced AAO-based deep CNN. AAO-based deep CNN is implemented using the MATLAB program on a computer with 4 GB of RAM, an Intel Core i3 processor, and Windows 10.



Algorithm 1 — Flow Diagram along with Pseudo Code of the Proposed AAO (**Input:** R_h, ε, ϕ ; **Output:** R_{h+1})

Dataset Description

Student Performance: This dataset pertains to student performance in secondary education from two Portuguese schools. It comprises of a range of attributes, involving demographic information, student grades, social factors as well as school-related features. Data collection involved questionnaires as well as school reports. Here it involved two distinct datasets corresponding to performance in two subjects: Portuguese language (por) as well as mathematics (mat).

Evaluation Measures

***Mean Square Error (MSE)*:** MSE estimates error in statistical frameworks via an average squared difference in between observed as well as predicted values utilizing Eq. (9).

Root Mean Square Error (RMSE): RMSE is a squared root of the MSE.

$$RMSE = \sqrt{\frac{1}{a}\sum_{i=1}^{a}\left[T_i - P_i\ \right]^2} \quad \ldots (23)$$

Mean Absolute Error (MAE): A calculation of errors between paired observations denoting the same fact is called an MAE.

$$MAE = \frac{1}{a}\sum_{i=1}^{a}\left|T_i - P_i\ \right| \quad \ldots (24)$$

Mean Absolute Percentage Error (MAPE): A typical statistic for assessing forecasting or prediction accuracy is MAPE. It calculates the standard deviation of the percentage difference between expected and actual values.

$$MAPE = \frac{1}{a}\sum_{i=1}^{a}\left(\frac{\left|T_i - P_i\right|}{P_i}\right)\times 100 \quad \ldots (25)$$

Mean Absolute Relative Error (MARE): By calculating the average relative difference between anticipated values and actual values, MARE is a statistic used to evaluate the accuracy of predictions or forecasts.

Mean Squared Relative Error (MSRE): The MSRE error metric measures the average of the squared relative differences among anticipated values and actual values to evaluate the accuracy of forecasts or predictions.

$$MSRE = \frac{1}{a}\sum_{i=1}^{a}\left(\frac{\left|T_i - P_i\right|}{T_i}\right)^2 \quad \ldots (26)$$

Root Mean Squared Relative Error (RMSRE): By evaluating the square root of the average of the squared differences between anticipated values and actual values, the RMSE error measure is used to evaluate the accuracy of predictions or forecasts.

$$RMSRE = \sqrt{\frac{1}{a}\sum_{i=1}^{a}\left(\frac{\left|T_i - P_i\right|}{P_i}\right)^2} \quad \ldots (27)$$

Comparative Methods and Analysis

The performance measures of introduced AAO-based deep CNN are assessed and comparison is done with existing methods like KNN,[1] TDNN+RNN,[18] ESPA[19] and CLDA+ANN.[20] A comparison analysis is performed in terms of changeable training data and value of K-fold for aforementioned performance measures.

Analysis based on Training Data

A detailed comparative evaluation of the created AAO-based deep CNN is shown in Fig. 3, while altering the percentage of training data from 50% to 90% and accounting for various evaluation variables. Specifically, when examining the MAE in Fig. 3(a), the AAO-based deep CNN exhibits a noteworthy performance, achieving an impressively low MAE of 0.178 with 90% of the training data, indicating minimal error when compared to existing methods. Furthermore, when assessing the MAPE in Fig. 3(b), the proposed AAO-based deep CNN attains a value of 15.98, outperforming alternative approaches. Similarly, the MARE in Fig. 3(c) highlights the excellence of the AAO-based deep CNN, with a MARE of 0.160 in comparison to other methods. Additionally, when considering the MSE as depicted in Fig. 3(d), the AAO-based deep CNN achieves an MSE of 5.593, demonstrating its superior accuracy compared to existing techniques. Furthermore, Fig. 3(e) illustrates the MSRE where the AAO-based deep CNN excels with a value of 4.475, surpassing other methods. In the context of the RMSE, as shown in Fig. 3(f), the proposed AAO-based deep CNN records an RMSE of 2.365, indicating its remarkable precision compared to existing methodologies. Lastly, Fig. 3(g) demonstrates the RMSRE, in which the AAO-based deep CNN achieves a low value of 2.129, further emphasizing its superiority over other existing approaches

Analysis based on K-Fold

In Fig. 4, a comprehensive comparative evaluation of the developed AAO-based deep CNN is presented, taking into account various evaluation metrics while varying the proportion of k-fold from 5 to 9. Specifically, when examining the MAE in Fig. 4(a), the AAO-based deep CNN exhibits a noteworthy performance, achieving an impressively low MAE of 0.192 with 9-fold of the training data, indicating minimal error when compared to existing methods. Furthermore, when assessing the MAPE in Fig. 4(b), the proposed method attains a value of 17.250, outperforming alternative approaches. Similarly, the MARE in Fig. 4(c) highlights the excellence of the proposed method, with a MARE of 0.173 in comparison to other methods. Additionally, when considering the MSE as depicted in Fig. 4(d), the AAO-based deep CNN achieves an MSE of 7.041, demonstrating its superior accuracy compared to existing techniques. Furthermore, Fig.

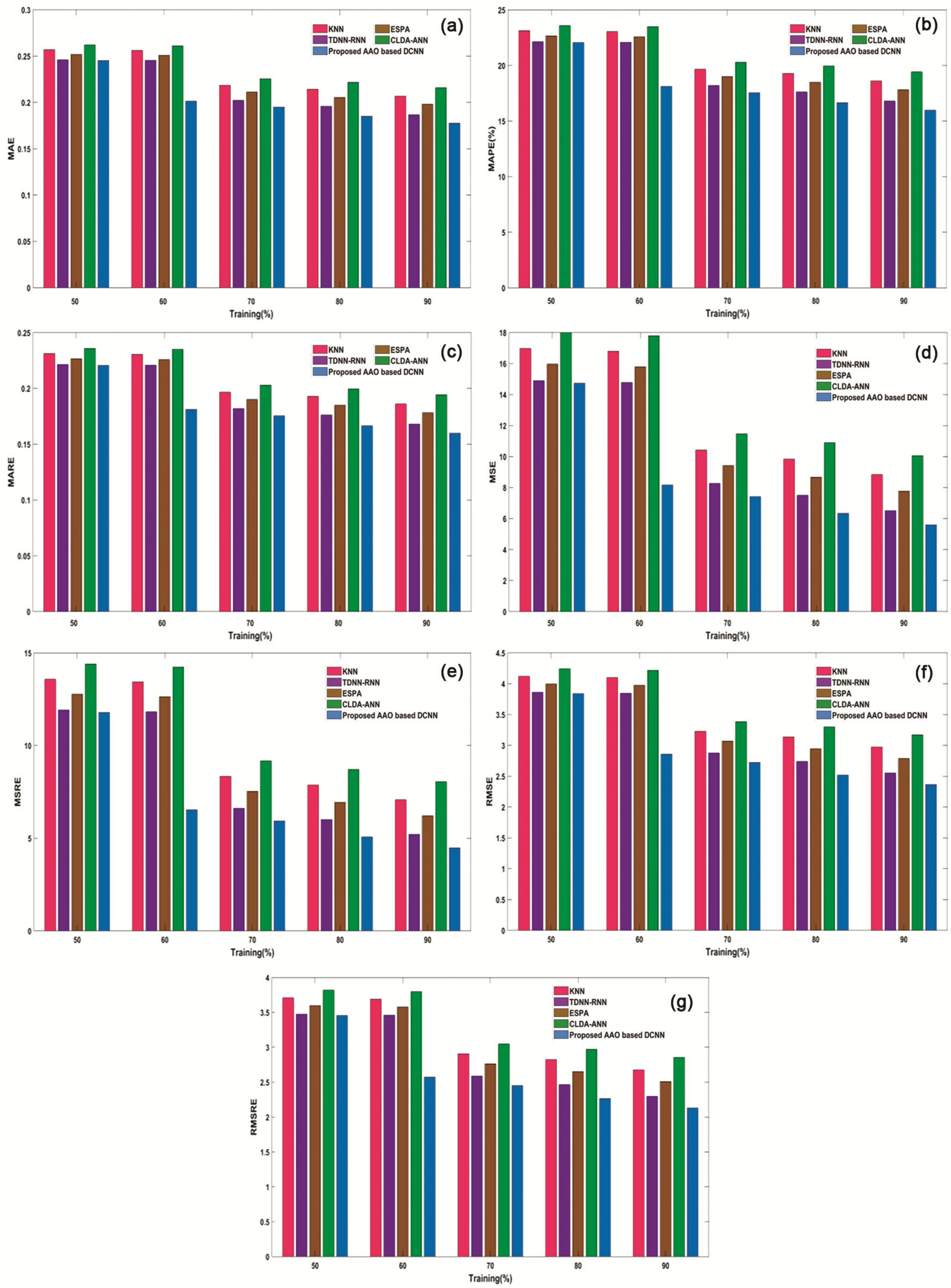


Fig. 3 — Comparative estimation relied on training percentage (a) MAE (b) MAPE (c) MARE (d) MSE (e) MSRE (f) RMSE (g) RMSRE

4(e) illustrates the MSRE where the AAO-based deep CNN excels with a value of 5.633, surpassing other methods. In the context of the RMSE, as shown in Fig. 4(f), the proposed AAO-based deep CNN records an RMSE of 2.654, indicating its remarkable precision compared to existing methodologies. Lastly, Fig. 4(g) demonstrates the RMSRE, in which the AAO-based deep CNN achieves a low value of 2.388, further emphasizing its superiority over other existing approaches.

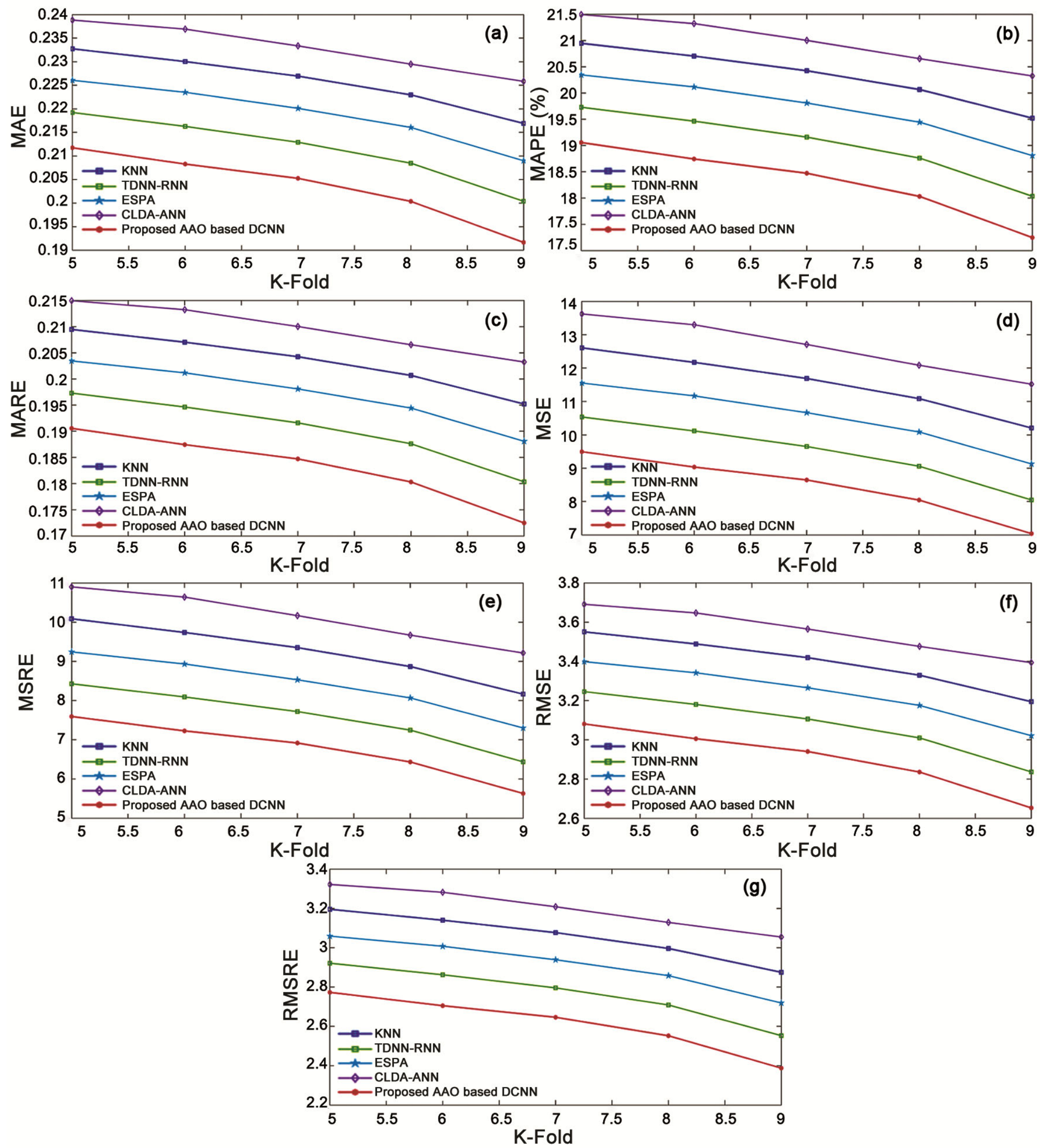


Fig. 4 — Comparative Estimation relied on k-fold (a) MAE (b) MAPE (c) MARE (d) MSE (e) MSRE (f) RMSE (g) RMSRE

Table 1 — Comparative analysis of proposed model for training percentage

Methods	Training percentage (90%)						
	MAE	MAPE	MARE	MSE	MSRE	RMSE	RMSRE
KNN	0.245	22.061	0.221	14.728	11.782	3.838	3.454
TDNN-RNN	0.201	18.124	0.181	8.166	6.533	2.858	2.572
ESPA	0.195	17.546	0.175	7.410	5.928	2.722	2.450
CLDA-ANN	0.185	16.650	0.166	6.331	5.065	2.516	2.265
Proposed	0.178	15.976	0.160	5.593	4.475	2.365	2.129

Comparative Discussion

Comparisons of the proposed AAO-based DCNN for training percentage and k-fold are provided in Tables 1 and 2. The proposed AAO-based DCNN has achieved minimal error value in case of 0.178 MAE, 15.976 MAPE, 0.160 MARE, 5.593 MSE, 4.475 MSRE, 2.365 RMSE, and 2.129 RMSRE for considered 90% of training data.

Table 2 — Comparative analysis of proposed model for k-fold

Methods	k-fold (9)						
	MAE	MAPE	MARE	MSE	MSRE	RMSE	RMSRE
KNN	0.245	22.061	0.221	14.728	11.782	3.838	3.454
TDNN-RNN	0.201	18.124	0.181	8.166	6.533	2.858	2.572
ESPA	0.195	17.546	0.175	7.410	5.928	2.722	2.450
CLDA-ANN	0.185	16.650	0.166	6.331	5.065	2.516	2.265
Proposed	0.178	15.976	0.160	5.593	4.475	2.365	2.129

Conclusions

In traditional time, educational staffs predict the student's performance based upon their experience regarding the nature of students and characteristics. Understanding and improvement of learning state of students is a significant intention in educational organizations. Therefore, prediction models are very necessary for an effectual prediction of student performance. Performance evaluation generally intended to offer feedback to students for improving learning quality and thus could augment the productivity of educational institutions because; students are the foundation of learning procedures. The major aim of student performance prediction is to evaluate a particular goal of student in handling learning materials effectively. This research proposes a proficient technique for prediction of student performance with an introduced AAO-based deep CNN. The proposed model has achieved minimal values of MAE, MAPE, MARE, MSE, MSRE, RMSE, and RMSRE for 90% of training data and 9-fold. In future, we intend to reiterate this model for predicting better student performance.

References

1 Turabieh H, Azwari S A, Rokaya M, Alosaimi W, Alharbi A, Alhakami W & Alnfiai M, Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance, *Computing*, **103(7)** (2021) 1417–1438, doi: 10.1007/s00607-020-00894-7.

2 Dixit P, Nagar H & Dixit S, Student performance prediction using case based reasoning knowledge base system (cbr-kbs) based data mining, *Int J Inf Educ Technol*, **12(1)** (2022) 30—35.

3 Singh V & Kaushik V D, Concepts of data mining and process mining, in process mining techniques for pattern recognition, *CRC Press*, (2022) 1–17.

4 Yousafzai B K, Khan S A, Rahman T, Khan I, Ullah I, Ur Rehman A & Cheikhrouhou O, Student-performulator: student academic performance using hybrid deep neural network, *Sustainability*, **13(17)** (2021) 9775.

5 Saa A A, Educational data mining & students' performance prediction, *Int J Adv Comput Sci Appl*, **7(5)** (2016) 212–220.

6 Dutt A, Ismail M A, & Herawan T, A systematic review on educational data mining, *IEEE Access*, **5** (2017) 15991–16005.

7 Wang Y, Ding A, Guan K, Wu S & Du Y, Graph-based ensemble machine learning for student performance prediction, (2021) *arXiv e-prints*, arXiv–2112.

8 Srinivas K, Prediction of E-learning efficiency by deep learning in E-khool online portal networks, *Multimedia Res*, **3(4)** (2020) 12—23, https://doi.org/10.46253/j.mr.v3i4.a2.

9 Asif R, Merceron A, Ali S A & Haider N G, Analyzing undergraduate students' performance using educational data mining, *Comput Educ*, **113** (2017) 177–194.

10 Gijbels D, Watering G V, Dochy F & Bossche P V, The relationship between students' approaches to learning and the assessment of learning outcomes, *Eur J Psychol Educ*, **20(4)** (2005) 327–341.

11 Shirsat P, Developing deep neural network for learner performance prediction in EKhool online learning platform, *Multimedia Res*, **3(4)** (2020) 24–31, doi: 10.46253/j.mr.v3i4.a3.

12 Dien T T, Luu S H, Thanh-Hai N & Thai-Nghe N, Deep learning with data transformation and factor analysis for student performance prediction, *Int J Adv Comput Sci Appl*, **11(8)** (2020) 711–721.

13 Amra I A A & Maghari A Y, Students performance prediction using KNN and Naïve Bayesian. In *2017 8th Int Conf Info Technol (ICIT) IEEE*, (2017) 909–913.

14 Hasan H R, Rabby A S A, Islam M T & Hossain S A Machine learning algorithm for student's performance prediction. In *2019 10th Int Conf Comput, Commun Netw Technol (ICCCNT)*, IEEE (2019) 1–7.

15 Kim B H, Vizitei E & Ganapathi V, GritNet: Student performance prediction with deep learning, (2018) *arXiv preprint arXiv:1804.07405*.

16 Wang X, Mei X, Huang Q, Han Z & Huang C, Fine-grained learning performance prediction via adaptive sparse self-attention networks, *Inf Sci*, **545** (2021) 223—240, https://doi.org/10.1016/j.ins.2020.08.017.

17 Clare B, Promoting deep learning: A teaching, learning and assessment endeavour, *Soc Work Educ*, **26(5)** (2007) 433–446.

18 Delianidi M, Diamantaras K, Chrysogonidis G & Nikiforidis V, Student performance prediction using dynamic neural models, (2021) *arXiv preprint arXiv:2106.00524*.

19 Niu K, Cao X & Yu Y, Explainable student performance prediction with personalized attention for explaining why a student fails, (2021) *arXiv preprint arXiv:2110.08268*.

20 Sood S & Saini M, Hybridization of cluster-based LDA and ANN for student performance prediction and comments evaluation, *Educ Inf Technol*, **26(3)** (2021) 2863–2878.

21 Hussain S & Khan M Q, Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning, *Ann Data Sci*, **10(3)** (2023) 637–655.

22 Student Performance Data Set, "https://archive.ics.uci.edu/ml/datasets/student+performance," Accessed on June 2022.

23 Othman S A & Ali H T, Improvement of the nonparametric estimation of functional stationary Time series using Yeo-Johnson Transformation with Application to Temperature Curves, *Adv Math Phys*, (2021) 1–6.

24 Gu Q, Li Z & Han J, Generalized fisher score for feature selection, (2012) *arXiv preprint arXiv:1202.3725*.

25 Tu F, Yin S, Ouyang P, Tang S, Liu L & Wei S, Deep convolutional neural network architecture with reconfigurable computation patterns, *IEEE Trans Very Large Scale Integr. (VLSI) Syst*, **25(8)** (2017) 2220–2233.

26 Abualigah L, Yousri D, Abd Elaziz M, Ewees A A, Al-Qaness M A & Gandomi A H, Aquila optimizer: A novel meta-heuristic optimization algorithm, *Comput Ind Eng*, **157** (2021) 107250.