



Student Performance Prediction Model Based on Lion-Wolf Neural Network

Ramanathan L^{1*}

Angelina Geetha²

Khalid M³

Swarnalatha P⁴

¹*School of Computer Science & Engineering, VIT University, Vellore, Tamilnadu, India.*

²*Department of CSE, B.S. Abdur Rahman University, Chennai, Tamilnadu, India*

³*Director, NITMAS & TNU, Kolkata, India*

⁴*School of Computer Science & Engineering, VIT University, Vellore, TamilNadu, India*

*Corresponding author's Email: lramanathan@vit.ac.in

Abstract: Early prediction of student performance helps to take action for better achievements of students. To achieve the better education standard, several attempts have been made to predict the performance of the student, but the prediction accuracy is not acceptable. To accomplish the enhanced prediction, neural network (NN) based method is proposed. In this paper, an approach to predict student academic performance in college education based on Lion-Wolf artificial neural network is proposed. Lion algorithm and Grey Wolf optimizer is integrated to develop a Lion-wolf training algorithm to find the optimal weight for every neuron in NN. The proposed prediction model is validated based on Mean Squared Error (MSE) & Root Mean Square Error (RMSE) with the existing NN based prediction model. The experimental results show that performance of proposed prediction model is improved compared to existing prediction model with MSE of 5.25 and RMSE of 2.3.

Keywords: Educational data mining, Lion optimization, Grey wolf optimizer, Prediction model, Entropy measure.

1. Introduction

The knowledge obtained through the application of Educational Data Mining (EDM) [2-5] can be used for various things such as offer suggestions to academic planners in higher education institutes to enhance their decision-making process, take actions for underperforming students [6-7]. Because of benefit provided to learners i.e. students, EDM has emerged as a state of art research area in recent years [2-5]. Student performance prediction model is one of the oldest and significant applications provided by EDM. Performance prediction model relies on different factors [8]. Many of factors are considered for EDM. Some of them are society factors, school factors, college factors, individual factor, family factors and so on. However, the factors considered of prediction are not consistent [9]. For effective prediction of student performance model, all factors related to the student must be included [15]. In early years, Statistical Package for Social Sciences has been used for student performance prediction. Artificial Neural Network-

based prediction model came into the picture recently [10-12]. The NN-based (Neural Network), performance prediction models are considered as demographic background inputs to the model. NN based methods are developed to increase the prediction accuracy compared to the regression-based prediction model.

The accuracy of NN-based modeling depends on training algorithm. Training algorithm in NN is considered as crucial since they will provide the right output for input based on neuron weight. And so, optimal weight selection for the training of NN is considered as an optimization problem. Many of Metaheuristic Optimization algorithms have been adapted for training ANN such as Genetic Algorithm, Cuckoo search, Firefly [13, 14], etc. Even though they provide better output for corresponding input in prediction models, the accuracy is not adequate. The better accuracy can be obtained by the optimization algorithm with fast convergence, minimal complexity, speedy solution attainment, etc. The main objective of this paper is to predict one to eight-semester performances of

college students. By predicting the semester performance of students in college education, the students with fewer semester marks can be taken with extra care providing or guiding the student for a better and safe future. For prediction with better accuracy, a new optimization algorithm based training algorithm has been developed in this paper.

The main contribution of this paper is given as follows:

Lion-Wolf training algorithm: Lion-Wolf Optimization algorithm is the main contribution of this work for optimal weight update in training algorithm of Neural Network for prediction of student semester marks in college education. Lion-Wolf algorithm is the integration of Lion Optimization algorithm and Grey Wolf optimizer.

The organization of this paper is as follows; motivation for the research is provided in section 2 briefing literature and challenges. In section 3, proposed student performance prediction model is explained with neat block diagrams. The proposed Lion-Wolf training algorithm for NN is deliberated in section 4. The results and discussion of proposed student performance prediction model is presented in section 5. Finally, section 6 concludes the paper.

2. Motivation

In this section, the motivation for the research on student performance prediction model using the artificial neural network is presented.

2.1 Literature Survey

In this section, the literature of existing research papers related to student performance prediction model is discussed. Cristoba Romero *et al.* [21] have demonstrated a prediction model. The purpose of the prediction model was to predict the final performance of first-year students based on their participation in online discussion forums. They have used multiple linear regression models and SVM classifier for prediction. They found that data collected for prediction was limited so that generalization of prediction approach is not possible. Since the prediction is made based on clustering approach; it is tedious and time-consuming. S.B. Kotsiantis presented another regression method based prediction model in [2]. The purpose of prediction model was to predict student's marks in the distant learning system. The prediction performance was better compared to conventional regression method, but wide-ranging conclusions seemed to be a problem.

Other than regression model which uses clustering phenomena for performance prediction,

Classifier and NN-based prediction model came into existence. Behrouz Minaei-Bidgoli *et al.* [3] have developed a prediction model based on multiple classifiers. The purpose of prediction model was student classification based on grade. The prediction performance was better compared to other classifier based approaches since multiple classifiers are combined. And also genetic algorithm was applied to weighting feature vectors aiding in enhanced prediction. Accuracy in the prediction of about 94% was obtained, but complexity in a combination of multiple classifiers such as Quadratic Bayesian classifier, 1-nearest neighbour (1-NN), k-nearest neighbour (k-NN), Parzen-window, multi-layer perceptron (MLP), and Decision Tree is not tolerable. The performance of multi-classifier approach decreases with increase in attribute selected for prediction.

Annika Wolff *et al.* [4] have applied decision tree classifier and SVM method for prediction of student performance. The purpose of prediction model was to identify the student who will benefit the tutor interventions. The decision classifier and SVM are integrated in a way for prediction with features as assessment scores. The attribute selection based on demographic, assessment data is indistinct aiding as a great compromise, but the complexity in prediction was not acceptable because of combined action of SVM and Decision tree classifiers. Another method incorporating decision tree classifier and naïve Bayes classifier was presented by Camilo.E. Lopez G *et al.* [5]. The purpose of prediction model was to predict low-performance students in first-year enrolments. The model used two classifier techniques, one for better understanding and other to quality assess the predicted data. They have obtained the prediction with accuracy ranging from 50-57%. The performance of enrolment based prediction approach decreases, when the enrollment data collected exceeds.

Also, Ernesto Pathros Ibarra García *et al.* [6] have also focused on a prediction based on classifiers. The purpose of prediction model was to predict first-year student performance. They have used Naïve Bayes classifier for classification. And they attained the prediction with 70% classification accuracy. But the factor considered for prediction was limited.

2.2 Problem Definition

The problem formulation for student performance prediction model is reflected below. The utmost intention of the proposed student

performance prediction model is to predict the semester marks of separate students based on data collected from various influencing factors related to students. Let X_i be data collected from S students based on family, schooling, environmental and individuality factors. The collected data is represented by,

$$X = \{X_1, X_2, \dots, X_N\} \tag{1}$$

where, N represent the number of data samples collected. From the collected data, the best features for the prediction are selected based on entropy measures. The entropy value for feature set is calculated and the feature factors with minimal entropy measure are selected as best feature. The best feature selected for prediction is given as,

$$X_F = \{X_{F_1}, X_{F_2}, \dots, X_{F_M}\} \tag{2}$$

where, M represents the best feature selected. The feature sets are given as input to the neural network to predict the semester performance. In a neural network, optimal weight for prediction is selected using proposed Lion-Wolf training algorithm. The solution vector representing the optimal weight of neuron in NN is given by,

$$W = \{W_1, W_2, \dots, W_0\} \tag{3}$$

The output of the proposed NN based prediction model is given by $SEM1, SEM2, SEM3, SEM4, SEM5, SEM6, SEM7$ and $SEM8$ are the marks of students in each semester.

3. Proposed Student Performance Prediction Model

In this section, detailed description about the proposed student performance prediction model is discussed. The overall block diagram of proposed prediction model is given in figure 1.

3.1 Data Collection

Data collection is the initial step in proposed prediction model. Data related to influential student factors such as individuality, environment, schooling, and family are collected.

For data collection, the questionnaire is initially prepared and distributed among the students and answers are collected. The students for data collection are selected from the different living areas such as village, town, city and metropolitan. Students are questioned with prepared questioned based on considered influencing factors related to family, schooling, family and environment. The answers to concerned questions are collected in excel sheet. In addition, academics of students on the topic of HSC and SSLC score are collected from mark sheets provided to college during admission. Based on necessity, parents are also questioned for data collection. Upon data collection, based on the requirement of NN, the collected data's are grouped into the input and the output variables.

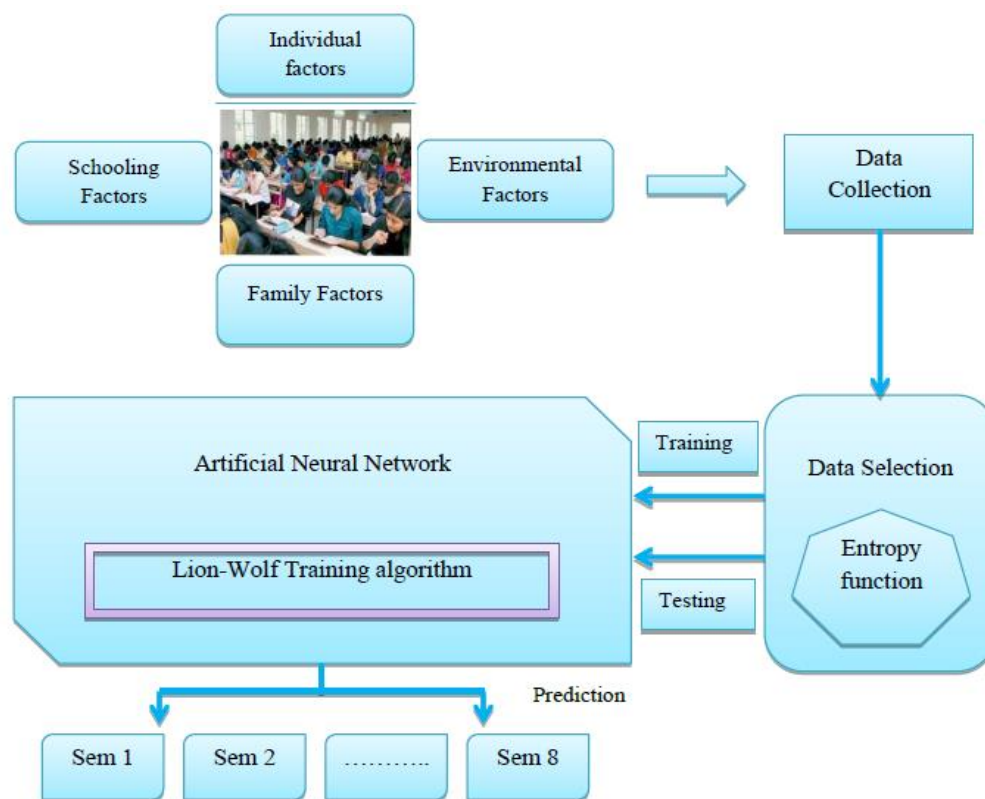


Figure.1 Block diagram of proposed student performance prediction model

The input variables are selected based on four varying factors connected to the student. The input variables are collected based on a questionnaire, from student application cards, academic records, etc. These collected factors are transformed into a format suitable for network analysis. The output variables represent the performance of a student. The output variables are semester marks of conforming students. Based on collected input, the semester performance of the student in college is predicted using which the performance of underperforming students can be further improved. Assume that X_i represents data collected from students S_s .

3.2 Data selection

Data selection is the second step in proposed student performance prediction model. Data selection is the fundamental block of the prediction model since the complexity in prediction is avoided by selection of relevant features from bulk data sets. In data selection, best feature for prediction is selected based on entropy function. Best feature signifies unique feature for semester performance prediction in our approach. The data selection on application of entropy measures is discussed below;

a) Entropy Function:

Entropy is a measure of information obtained when the outcome is observed [16]. Entropy is a measure of uncertainty of a random variable [17]. In this prediction model, certain feature or factors of respective students are selected based on entropy measure. Entropy value of each factor associated with students is calculated with respective to actual class values. The Shannon entropy [18] function is given by;

$$H = -\sum p \log p \quad (4)$$

where, p is probability value. Here, entropy measure is calculated as attribute vs. class function. The attribute represents the feature factor of students and class represents a number of target output i.e. semester marks. For each of feature factors, entropy values are calculated and it can be described by;

$$F(X_i) = \frac{1}{m} \sum_{q=1}^m \sum_{i=1}^{X_f} P_i^q \log P_i^q \quad (5)$$

where, P_i^q is the probability of getting q^{th} value from feature set. X_f is unique features selected. For all feature, entropy measure is calculated. Based on calculated measure values, feature factors with minimal entropy value are selected as the best feature for prediction of semester performance.

3.3 Prediction

The selected features X_f are given to neural network as input. In the neural network, based on given input semester marks of concerned students are predicted. The prediction of semester marks of concerned students is made based on ground truth values of collected student data. NN is a supervised predictor, based on provided knowledge, the semester marks of students are predicted by influencing factors. The predicted outputs of neural network are given by *SEM 1*, *SEM 2*, *SEM 3*, *SEM 4*, *SEM 5*, *SEM 6*, *SEM 7* and *SEM 8*.

4. Design of Neural Network

The Artificial neural network designed for prediction of semester marks of college students is discussed in this section. ANN is models based on biological neural networks. It is one of noteworthy intelligent machine learning technique [19]. In this paper, data selected from collected data of students are used to predict semester performance of students.

4.1. Architecture

The architecture of proposed NN for student performance prediction model is presented in this section. In NN, neurons are organized into a sequence of layers. The NN consist of three types of layers; input layers, output layers, and hidden layers. The input layer is not a logical layer it only passes input and has neither input weights nor any activation functions. The NN used in our prediction model is feed forward network which activates the output function from the input layer to output layer. The Network architecture of proposed NN for semester mark prediction is shown in figure 2.

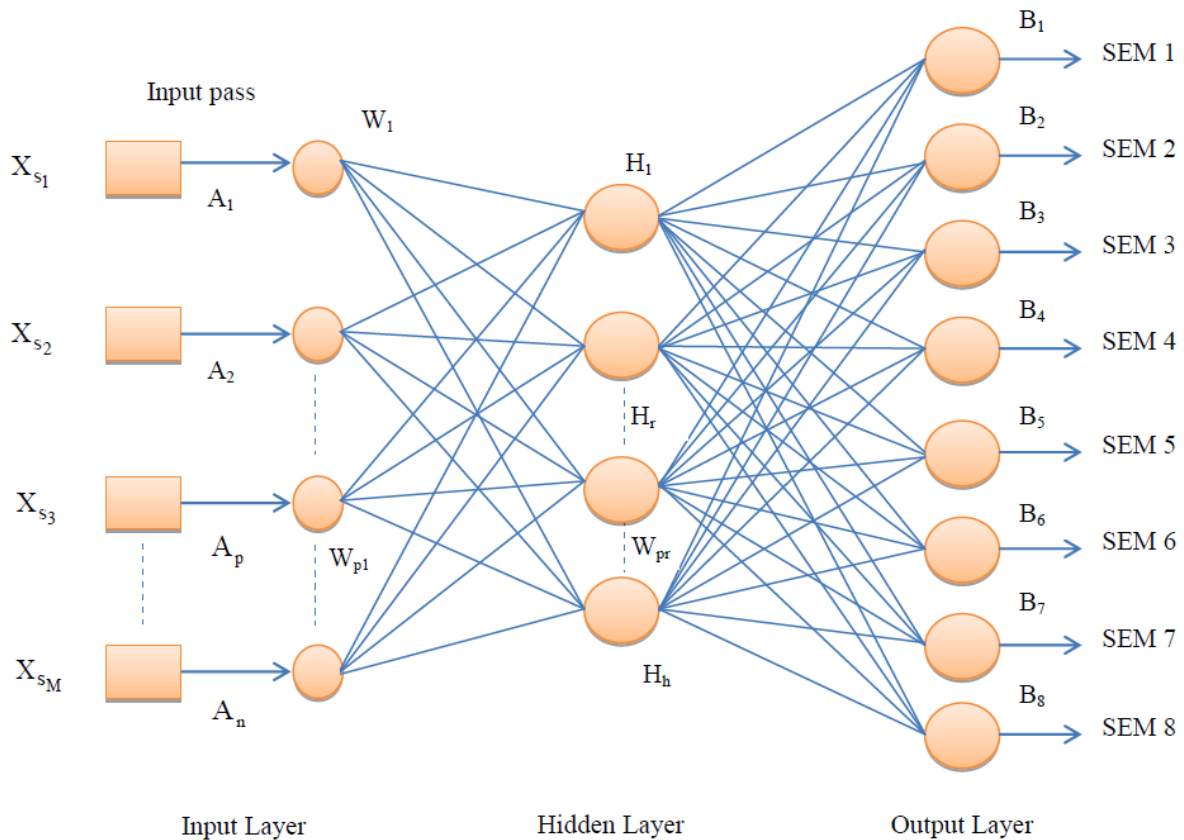


Figure.2 ANN architecture

In figure 2, n , h and m denotes the number of input neuron, hidden neuron and output neuron in input, hidden and output layer respectively. W_{pr} is input weight, and W_{pr} is hidden weight matrices in networks. X_F is input factors selected for a neural network which is passed on to the input layer. A is input vector of the network, H is output vector of hidden layer and B is output vector of the network. θ is a bias of hidden and output layers. Determination of a number of neurons and number of layers in a neural network is the initial step. It is considered important because if smaller the neurons and layers in the network, the processing ability of network are low. Similarly, if a number of neurons and layers is large, training time of network slow down. Either increase or decrease in neuron and layer affects network performance. In the proposed prediction approach, the number of hidden layers and number of neurons is adjusted and selected based on experimentation i.e. network with better prediction.

The information flows only in forward direction, the output of hidden neuron can be described by;

$$H_r = \text{Fn}(h_r) \tag{6}$$

Where, $\text{Fn}(\cdot)$ is node transfer or activation function. The activation function used is same for all

nodes in NN. Usually, non-linear functions such as sigmoid function is chosen as node transfer function. [3] Activation function used in NN is a sigmoid function. The output of hidden neuron as per sigmoid function is given by;

$$H_r = \frac{1}{1 + (\exp(-h_r))}; \quad r = 1,2,3,\dots,h \tag{7}$$

$$h_r = \sum_{p=1}^n (W_{pr} \cdot A_p) - \theta_r \tag{8}$$

Here, h is the number of hidden layer, W_{pr} is connection weight between hidden neuron and input, θ is a bias of hidden neuron, A is an input of node and n is the number of hidden neuron. The output of NN can be described by;

$$B_q = \frac{1}{1 + (\exp(-b_q))} \tag{9}$$

$$b_q = \sum_{r=1}^h (W_{r,q} \cdot H_r) - \theta_q; \quad q = 1,2,3,\dots,m \tag{10}$$

Here, H_r is the output of hidden neuron, W_{rq} is weight between hidden and output neuron, θ_q is a bias of output neuron.

4.2. Training algorithm

Training algorithm of NN is considered important because processing ability of network depends on connection weights of neuron in network obtained through learning. Training of NN is to find a set of weights for neurons in the network. In training, biases for all neurons are also found out.

The input to neural network is represented by,

$$A = \{A_1, A_2, A_3, \dots, A_n\} \quad (11)$$

The output response of network is given as,

$$B = \{B_1, B_2, B_3, \dots, B_m\} \quad (12)$$

Also, the desired response of network is given as,

$$C = \{C_1, C_2, C_3, \dots, C_m\} \quad (13)$$

where, n is the number of input features, and m is the number of output. The weight must be selected in the network that gives desired values as the output. During training, weight set which gives response equal to desired response of network is selected as optimal weight. The termination criteria utilized in weight selection and update is Means Square Error function. NN is trained by minimizing the error function. In our prediction model, error function used is Mean square error (MSE) function. MSE is given by;

$$MSE = \frac{1}{nm} \sum_{p=1}^n \sum_q^m (C_{pq} - B_{pq}) \quad (14)$$

Where, n denotes the number of inputs to network, m denotes the number of FNN outputs, C is target value, and B is the actual response of NN. In proposed student performance prediction model, NN predicts the semester performance of student, so output of NN is represented by;

$$\begin{aligned} B_1 &= SEM_1, B_2 = SEM_2, B_3 = SEM_3, B_4 = SEM_4, \\ B_5 &= SEM_5, B_6 = SEM_6, B_7 = SEM_7, B_8 = SEM_8 \end{aligned} \quad (15)$$

4.2.1. Lion-Wolf Optimization algorithm

In this section, proposed Lion-Wolf optimization algorithm for neural network training is deliberated. Lion-Wolf optimization algorithm is developed by integrating Grey Wolf optimizer [20] and Lion optimization algorithm [1]. The proposed optimization algorithm is majorly based on grey wolf optimizer, but the position update of grey wolf optimizer is hybridized with lion optimization algorithm increasing convergence rate and avoiding the local optimal problem. Grey wolf optimizer is a metaheuristic optimization algorithm which is based on leadership behaviors of grey wolf [20]. The hunting behavior of the grey wolf is used for optimization. The concept of prey hunting is extended in search space for best solution attainment which is nothing but the prey.

a) Solution Encoding

In proposed Lion-Wolf optimization algorithm, the solution is grey wolf population. The possible solution is represented in solution encoding. The grey wolf population represents vectors of neuron weight for training. The length of solution i.e. number of grey wolfs population is equal to a number of weights required in training of NN. Let W be grey wolf population, and o be the number of grey wolfs.

$$W = \{w_1, w_2, \dots, w_o\} \quad (16)$$

Each of solution in population is equal to a number of neuron weights to be adjusted in the network. The values are initially selected randomly, and the Lion-Wolf optimization algorithm took over the part generating an optimal weight for neurons in the network.

b) Fitness Evaluation

The fitness function used in proposed Lion-Wolf Optimization training algorithm for the neural network is deliberated in this section. The fitness evaluation function used is Mean Square Error. For each of solution in grey wolf population which is nothing but an optimal vector of connection weight, training is performed in NN. After training, the Mean square error function is calculated between actual response and desired response of the network. Mean Square Error function value is given by;

$$MSE (W_{pq}) = \frac{1}{nm} \sum_{p=1}^n \sum_q^m (C_{pq} - B_{pq}) \quad (17)$$

Where, n is a number of input, m is a number of output, C is the desired response value of NN and B is the actual response of NN for considered solution vector. For all solution in grey wolf population, MSE is calculated and solution with best fitness i.e. which has response equal to that of desired response is chosen optimal weight.

c) Lion-Wolf training algorithm

The steps involved in proposed Lion-Wolf optimization algorithm is discussed below;

Each solution in grey wolf population represents solution vector containing weights for the training of algorithm. Optimal weight for training which produces target output of NN is intended. Optimization using Lion-Wolf Optimization algorithm selects the optimal weight. The position and solution update of Lion-Wolf optimization algorithm updates solution vector over consecutive iterations.

Step 1. Parameter Initialization

Initialize Coefficient vectors for position updating of grey wolfs. Coefficient vectors \vec{A} and

\vec{C} are based on the value of \vec{a} . The Coefficient vectors are calculated as follows;

$$\vec{A} = 2 \vec{a} r_1 - \vec{a} \quad (18)$$

$$\vec{C} = 2 \vec{r}_2 \quad (19)$$

Where, r_1 and r_2 are random vectors in the interval [0, 1]. In addition, a number of iterations (N_{max}) for optimization is also initialized.

Step 2. Population Initialization

Primarily, the grey wolf population is initialized randomly. The grey wolf population is represented by;

$$G_u = \{ \vec{G}_1, \vec{G}_2, \dots, \vec{G}_U \} \quad (20)$$

where, U represents the size of the population.

Step 3. Fitness Calculation

For each of search agents in grey wolf population, fitness is calculated using the fitness equation given in fitness evaluation section. Based on the fitness values, the first three best solutions are considered as;

$$Fit_1 = \vec{G}_\alpha; Fit_2 = \vec{G}_\beta; Fit_3 = \vec{G}_\delta \quad (21)$$

Where, alpha, beta and delta wolves are the top three grey wolves sorted according to the fitness function.

Step 4. Lion Fusion

To further reduce the local optimal problem in grey wolf optimizer, fertility evaluation of lion optimization algorithm is fused with GWO. According to Lion algorithm [32], Female Fertility evaluation equation is given by;

$$\vec{G}_{Female+} = \begin{cases} \vec{G}_{Female+}^k; & \text{If } l = k \\ \vec{G}_{Female+}^l; & \text{Otherwise} \end{cases} \quad (22)$$

Here, we are assuming that l is equal to k . so the above equation becomes;

$$\vec{G}_{Female+}^1 = \vec{G}_{Female+}^1 \quad (23)$$

To find the value of updated female Lion, min maximum bound equation is utilized, and it is given by;

$$\vec{G}_{Female+}^l = \min[\vec{G}_{max}^l, \max(\vec{G}_{min}^l, \nabla_l)] \quad (24)$$

$$\nabla_l = [\vec{G}_{Female+}^l + (0.1v_2 - 0.05)(\vec{G}_{male} - v_1 \vec{G}_{Female+}^l)] \quad (25)$$

Here, l is vector element of updated female lion generated within the interval [0, L]. L is the length of the solution, ∇ is female update function, v_1 and v_2 are random integers generated between interval [0, 1].

Step 5. Lion-Wolf hybridization

For grey wolf optimizer, the first and second best fitness function is epitomized by,

$$Fit_1 = \vec{G}_\alpha; Fit_2 = \vec{G}_\beta \quad (26)$$

Similarly, for Lion optimization algorithm [2], the first and second best fitness function is epitomized by;

$$Fit_1 = \vec{G}_{Male}; Fit_2 = \vec{G}_{Female} \quad (27)$$

The second best solution in both optimization algorithms is equal, so it is given by;

$$\vec{G}_\beta = \vec{G}_{Female} \quad (28)$$

Substituting the female lion into the beta wolf, we get,

$$\vec{G}_{\beta+}^l = \min[\vec{G}_{max}^l, \max(\vec{G}_{min}^l, \nabla_l)] \quad (29)$$

$$\nabla_l = [\vec{G}_{\beta+}^l + (0.1v_2 - 0.05)(\vec{G}_\alpha - v_1 \vec{G}_{\beta+}^l)] \quad (30)$$

Step6. Position Updation

Omega search agents in grey wolf population are position updated based on the position of search agents alpha, beta, and gamma. The position updation formula proposed in this optimization algorithm is given by;

$$\vec{G}(t+1) = \frac{\vec{G}_1 + \vec{G}_2 + \vec{G}_3 + \vec{G}_4}{4} \quad (31)$$

Where, \vec{G}_1 is based on the position of first best search agent, \vec{G}_2 and \vec{G}_4 are based on the position of second best search agents, \vec{G}_3 is based on third best search agent. The values are given by;

$$\vec{G}_1 = \vec{G}_\alpha - A_1 \cdot D_\alpha \quad (32)$$

$$\vec{G}_2 = \vec{G}_\beta - A_2 \cdot D_\beta \quad (33)$$

$$\vec{G}_3 = \vec{G}_\delta - A_3 \cdot D_\delta \quad (34)$$

$$\vec{G}_4 = \vec{G}_{\beta+} \quad (35)$$

Where, D is based on encircling behavior of grey wolf and it is given by;

$$D = \left| \vec{C} \cdot \vec{G} - \vec{G}(t) \right| \quad (36)$$

Based on search agents, the encircling position of grey wolf varies.

Step7. Solution Update

After position update, update the value of coefficient vectors, thereby changing encircling position of grey wolves. The values of coefficient vectors are varied by changing the value of \vec{a} . The value of \vec{a} is reduced from 2 to 0 over the course of iterations in optimization. The value of

\vec{A} corresponds to the solution attainment and it is related by;

$$\left| \frac{\vec{A}}{A} \right| = \begin{cases} > 1; \text{Divergeaway} \\ < 1; \text{Converge} \end{cases} \quad (37)$$

Upon parameter update and position update, the fitness of population is calculated based on fitness evaluation.

Step8. Iteration

Position update and solution update is iterated till the termination condition is achieved. The best \vec{G}_α is selected as the optimal solution. The best \vec{G}_α is solution vector containing optimal weights for a neural network with low error function i.e. MSE.

5. Results and discussion

The result and discussion of proposed student performance prediction model is discussed in this section.

a) Experimental Setup

The experimentation of proposed student performance prediction model predicting student semester marks in college is performed in a personal computer with the following specification; i) Windows 8 Operating System ii) Intel Core i-3 processor iii) 2GB Physical memory. The software tool used for implementation of proposed student performance prediction model is Weka in Java.

b) Performance Measures

The performance measures considered for performance evaluation are Mean Squared Error and Root Mean Squared Error.

c) Comparative Methods

The performance validation of proposed student performance prediction model is compared to existing ANN based student performance prediction model using optimization algorithms such as genetic algorithm, Grey Wolf optimizer, and Lion Optimization algorithm as training algorithm.

5.1 Performance Evaluation

The performance evaluation of proposed student performance prediction model is presented in this section. The performance of proposed NN in student performance prediction model is validated over comparative method using MSE and RMSE measure.

i) Analysis based on varying percentage of training data

Figure 3 depicts analysis based on varying percentage of training data. The analysis curve is a

plot between the percentage of training data and MSE & RMSE as depicted in figure 3a and 3b respectively. The experimentation for varying percentage of training data is performed with 30 features and one hidden layer. With 60% of training data to the network, existing GA, GWO, and LION attained MSE value of 9.125, 9.575, and 9.125 respectively, whereas the proposed LION-WOLF attained the MSE value of 8.625 which is lower compared to the existing system. Lower the MSE value better the performance of NN in the prediction of student performance. With the increase in the percentage of training data, the MSE value of comparative system decreases. However, MSE value of proposed LION-WOLF is lower than all other comparative systems at 70%, 80% and 90% of training data. Moreover, with 60% of training data, RMSE value attained by GA, GWO and LION is 3, 3.1 and 3.02 respectively, whereas the RMSE value attained by proposed LION-WOLF is 2.9 which is lower than RMSE value of existing systems. Similarly, for 70% of training data, existing GA, GWO and LION achieved RMSE value of 2.9 respectively which is greater than RMSE value 2.6 attained by LION-WOLF. The best case MSE value of 5.25 is attained by proposed LION-WOLF with 90% of training data and best case RMSE value of 2.3 is attained by proposed LION-WOLF with 90% of training data. From the analysis, it is clear that proposed LION-WOLF perform well for a different percentage of training data, thereby abetting in enhanced performance prediction of students. From the figure 3, the proposed Lion-wolf outperformed the existing methods by showing the minimum MSE and RMSE. The reason for improvement happened for the lower percentage of training data and also higher percentage. This performance is achieved due to the better convergence of the Lion-wolf algorithm by finding the training weights shortly. The inclusion of Lion based weight update finds the training weights accurately so that the training required for the neural network requires less time and thereby improve the result.

ii) Analysis based on varying hidden layers

The experimentation for analysis over varying the hidden layers in the network is performed with 60% of training data and 30 features. Figure 4 represents analysis for varying number of hidden layers. The evaluation based on MSE measure is depicted in figure 4a, and RMSE measure is depicted in figure 4b. At the same instance,

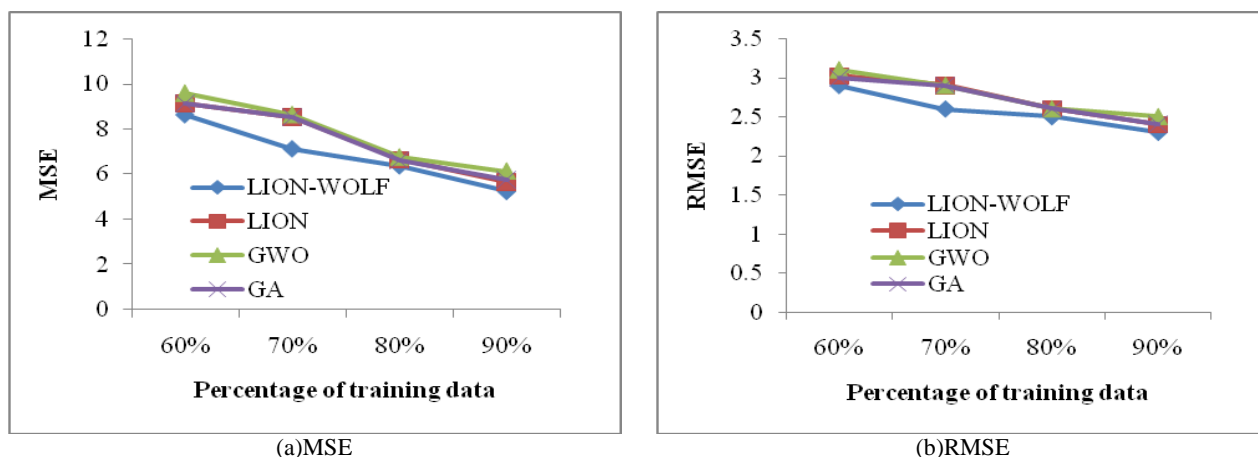


Figure.3 Analysis curve for varying percentage of training data

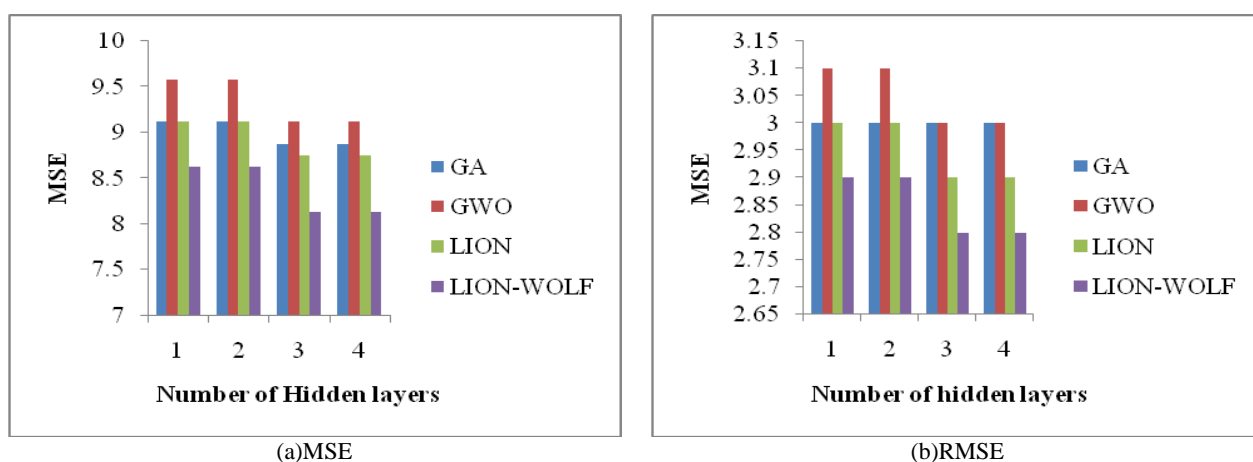


Figure.4 Analysis curve for varying number of hidden layers

proposed LION-WOLF achieved the MSE value of 8.625 which is lower than the value attained by existing systems. With 3 hidden layers in NN, the MSE value of 8.125 is attained by proposed LION-WOLF, whereas the existing GA, GWO and LION reached MSE value of 8.875, 9.125 and 8.75 respectively. The best case MSE value of 8.125 is attained by proposed LION-WOLF with hidden layer size 3 and 4 respectively. The worst case MSE value of 9.575 is attained by GWO with hidden layer size 1 and 2 respectively. Consequently, in NN with one hidden layer, the existing GA, GWO and LION achieved the RMSE value of 3, 3.1 and 3 respectively, whereas proposed LION-WOLF attained RMSE value of 2.9 which is lower than the value reached by existing systems. Similarly, with a hidden layer of size 2, the proposed LION-WOLF attained RMSE value of 2.9, but existing GA, GWO and Lion attained the RMSE value of 3, 3.1 and 3 respectively. The best case RMSE value of 2.8 is achieved by proposed LION-WOLF in NN with 3 and 4 hidden layers respectively

GWO achieves the worst case RMSE value of 3.1 with 1 and 2 hidden layers. From the analysis, it is clear that proposed LION-WOLF perform well with the decrease in a number of hidden layers. However it is better compared to existing systems. When the performance is analyzed through the hidden layers, the proposed Lion-Wolf shows the better performance for lower and higher hidden layers. This performance improvement happened due to the useful finding of the training weights by the proposed training algorithm. The fusion of Lion formula into the grey wolf optimization improves the convergence and thereby the training error was reduced. The better performance in training by the proposed algorithm shows the good performance in the semester mark prediction.

6. Conclusion

This research work presented in this paper was an attempt to predict the semester performance of college students using learning capability of artificial neural network. The proposed prediction model was developed based on some input variables

related to individual, schooling, environmental and family factors of students in which best factor for prediction is selected using entropy function. The selected features were given to Lion-Wolf NN for prediction of semester marks. Lion-Wolf Optimization algorithm was proposed for the training of neural network for optimal weight selection. The experimentation of proposed student performance prediction model was validated over existing training algorithm. The prediction results show that potential efficacy of proposed Lion-Wolf NN as a prediction tool was superior compared to existing ANN based student performance prediction model with minimal MSE of 5.25 and minimal RMSE of 2.3. In future, the training of neural network can be improved by including various objective functions instead of a mean square error.

References

- [1] B.R.Rajakumar, "Lion algorithm for standard and large scale bilinear system identification: A global optimization based on Lion's social behaviour." *2014 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2014.
- [2] S.B.Kotsiantis, "Use of machine learning techniques for educational proposes: a decision support system for forecasting students grades", *Artificial Intelligence Review*, vol. 37, no.4, pp. 331–344, 2012.
- [3] B.Minaei-Bidgoli and W.F.Punch, "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System", *In proceedings of Genetic and Evolutionary Computation Conference*, pp. 2252-2263, 2003.
- [4] A. Wolff, Z.Zdrahal, D.Herrmannova, and P.Knoth, "Predicting Student Performance from Combined Data Sources", *In Educational Data Mining*, pp. 175–202, Springer, 2014.
- [5] C.E.L.Guarín, E.L.Guzman and F.A.González, "A model to predict low academic performance at a specific enrolment using DATA mining", *IEEE Journal Learning technologies*, vol. 10, no. 3, pp. 119-125, 2015.
- [6] E.P.I.García and P.M.Mora, "Model prediction of Academic Performance for First year Students", *In proceedings of International Conference on Artificial Intelligence*, pp. 169-174, 2011.
- [7] G.Lassibilille and L.N.Gomez, "Why do higher education students dropout?", *Evidence from Spain, Educational Economics*, vol. 16, no. 1, pp. 89-105, 2008.
- [8] J.Touron, "The determination of factors related to academic achievement in the university: implications for the selection and counselling of students.", *Higher Education*, vol. 12, no. 4, pp. 399-410, 1983.
- [9] S.Malvandi and A.Farahi., "Provide a method for increasing the efficiency of learning management systems using educational data mining", *Indian Journal of Science and Technology*, vol. 8, no. 28, pp. 1-10, 2015.
- [10] Z. Ibrahim and D.Rusli "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression," in *21st Annual SAS Malaysia Forum*, 2007.
- [11] Z.Ibrahim, N.M.Rusli and R.M Janor, "Predicting Students' Academic Achievement: Comparison between Logistic Regression, Artificial Neural Network, and Neuro-Fuzzy," in *International Conference on Computers*, 2008.
- [12] P.M.Arsad, N.Buniyamin and J.A.Manan, "A Neural Network Students' Performance Prediction Model (NNSPPM)", *In proceedings of IEEE International Conference on Smart Instrumentation, Measurement and Applications*, 2013.
- [13] D.J.Montana and L.Davis, "Training Feedforward Neural Networks Using Genetic Algorithms", *Machine Learning*, pp. 762-767, 1989.
- [14] I.Brajevic and M.Tuba, "Training Feed-Forward Neural Networks Using Firefly Algorithm" *Recent Advances in Knowledge Engineering and Systems Science*, pp. 156-161, 2013.
- [15] P.R.Shetgaonkar, "Predicting the impact of different Variables on Students Academic Performance using Artificial Intelligence", *International Journal of Computer Science and Information Technologies*, vol. 6, no. 2, pp. 1367-1370, 2015.
- [16] H.M.Lee, C.M.Chen, J.M.Chen, and Y.L. Jou, "An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy", *IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics*, VOL. 31, no. 3, pp. 426-432, 2001.
- [17] C.E.Shannon, "A mathematical theory of communication," *Bell System Technical Journal*., vol. 27, pp. 379–423, 1948.
- [18] B.Kosko, "Fuzzy entropy and conditioning," *Information. Science*, vol. 40, pp. 165–174, Dec. 1986.
- [19] J.F.Chen, H.N.Hsieh, and Q.H.Do, "Predicting Student Academic Performance: A Comparison of Two Meta-Heuristic Algorithms Inspired by Cuckoo Birds for Training Neural Networks", *Journal Algorithms*, vol. 7, pp. 538-553, 2014.
- [20] S.Mirjalili, S.M.Mirjalili and A. Lewis, "Grey Wolf Optimizer", *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [21] C.Romero, M.I. López, J.M.Luna and S.Ventura, "Predicting students' final performance from participation in on-line discussion forums", *Computers and Education*, vol. 68, pp. 359-472, 2013.