Predicting Academic Performance of Deaf Students Using Feed Forward Neural Network and An Improved PSO Algorithm

V. Sathya Durga*
Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Padur, Chennai, India. E-mail: sathyadurga.v@gmail.com

Thangakumar Jeyaprakash
Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Padur, Chennai, India.

Received October 20, 2020; Accepted November 25, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V18SI01/WEB18048

Abstract

Literacy rate of deaf students is very less in India. So there is a need to build an effective academic prediction model for identifying weak deaf students. Many machine learning techniques like Decision tree, Support Vector Machine, Neural Network are used to build prediction models. But the most preferred technique is neural network. It is found out that regression model build with neural networks takes more time to converge and the error rate is quite high. To solve the problems of neural network, we use Particle Swarm Optimization (PSO) for weight adjustment in the neural network. But, one of the main drawback of PSO lies in setting the initial parameters. So, a new PSO algorithm which determines the initial weight of the neural network using regression equation is proposed. The results show that neural network build with the proposed PSO algorithm performs well than neural network build with basic PSO algorithm. The Mean Square Error (MSE) achieved in this work is 0.0998, which is comparatively less than many existing models.

Keywords

Feedforward Network, Linear Regression, Mean Square Error, Particle Swarm Optimization, Prediction.

Introduction

As per census 2011, literacy rate of deaf students is very less in India and only 8.5% of disabled population of India acquire a graduate degree [1]. Many researchers all around the world have build many academic prediction models for students. Some models predicts whether a student will pass an exam or not and some applications predicts the grade a student will acquire at end of an examination. If an academic prediction model for

deaf students is developed it will definitely help the education system of the deaf. Academic prediction models can be of two types, classification model and regression models. In classification model, class labels are predicted. For example, some applications predict whether a students will pass the examination or fail. Here, pass and fail are the class labels. In regression models, numeric values are predicted. For example, the total mark the students acquires in the final exam is predicted. Classification model is evaluated by accuracy and regression model is evaluated by error metrics. So, a good classification model, has more accuracy and a regression model, has less error rate. Many machine learning techniques like Support Vector Machine, Neural Network, Decision tree are used to build academic models. Out of the above mention techniques neural network is the most commonly used technique for both classification and regression applications. But prediction models build with neural network for classification application achieves good accuracy, but regression models fails to achieve less error rate. Another problem faced by the neural network is slow convergence rate. It takes more time to converge.

To solve the above mentioned problems (High error rate and Slow convergence) Particle Swarm Optimization (PSO) for weight adjustment in the neural network is used. PSO is selected because it is simple to implement, has short computation cycle and it converges faster. But the problem in PSO is that, it is difficult to set the initial position of the particles. If initial positions are set to inaccurate values, then PSO will take longer time to converge. So this research work tries to find out an equation, which determines initial optimal position values for the particles in the PSO algorithm. By implementing this enhanced PSO algorithm for weight adjustment, neural network achieves low error rate and convergences fastly. This enhanced neural network is used to build an academic prediction model for deaf students. Finally, interesting patterns discovered from deaf students dataset is also discussed at the end.

This paper contains the following section. Section 1 is the Introduction. Section 2 is the review of literature on students performance prediction systems. Materials and methods used to develop the PSO algorithm and to develop the prediction model for deaf students is discussed in section 3. Section 4 contains results and discussion. Section 5 contains the conclusion of this research work.

Review of Literature

In this section, latest research work on students performance prediction is discussed.

1) Performance Prediction by SVM

Iti and Subhranil builds an academic prediction model for students based on Support Vector Machine. The model works on psychological features of students and classifies the students into low, high or average performer. The accuracy of this work is 95% and it uses radial basis kernel to achieve this result [2]. Olaruntpba and Akinode using SVM technique to build a prediction model for Nigerian students. Using marks obtained in terminal high school exam and three year CGPA, GPA is predicted. The dataset contains 89 student records with 12 features. Accuracy of this work is 97% [3].

2) Performance Prediction by Decision Tree

Paul et al build a student performance prediction model using decision tree technique. Students demographic details, GPA and psychometric attributes are used to predict whether a student will be a honor student or not. Size of the dataset is 50. Accuracy of this prediction model is 90.67% [4]. Mrinal and Vivek build a classification model based on decision tree which is used to predict the final grade an engineering student achieves from his performance in seven semesters examinations. The dataset contains 524 engineering students from Haryana. The decision tree algorithm used in this work is J48. Accuracy of this prediction model is82.5%[5].

3) Performance Prediction by Naive Bayes

Fahad et al builds an prediction model using naive bayes classification and predict GPA of the students. Attendance and assignment grades are used for predicting the GPA of the students. Initially, the dataset consist of 660 records. After pre-processing, the dataset size was reduced to 500. Accuracy of this work is 96.8%[6]. Delai and ester use naive bayes method and build a student's performance prediction model. The model predicts whether a student will pass the exam or fail. Attributes used in this model are attendance, test score, assignment marks, etc. Accuracy achieved by this prediction model is 88% [7].

4) Performance Prediction by Neural Network (Classification Model)

Lau et al use neural network to predict CGPA of students from socio economic details and from entrance exam marks. The dataset contained details of 1000 students. The neural network has 11inputs and two hidden layer. The network is trained with Levenberg – Marquardt algorithm. This prediction model achieves accuracy of 84.8% [8]. Fahim and Jamal build a prediction model with neural network. The model predicts students CGPA with their previous grades, family details and socio economic details. Totally 14 attributes

are used. The dataset contains 120 records. The neural network achieves 88% accuracy [9].

5) Performance Prediction by Neural Network (Classification Model)

Bogdan et al builds an academic prediction model to predict students first year grade with their age and high school marks. The dataset contains 1000 records with 7 features. Tanh and softmax activation function is used. Two hidden layers are used in this research work. The prediction model achieves 2% error rate [10]. Ramanadhan et al builds an prediction model using neural network. The network predicts eight semester grades the student will achieve with their SSLC mark, higher secondary mark and family details. It is a regression model. A new algorithm called lion wolf algorithm is proposed to adjust weight in this network. Error achieved by this network is5.3% [11].

From literature review it is found out there is no prediction models for deaf students and the error rate in neural network for regression application is high. So, there is a need to build a neural network which produce low error rate for regression applications. And the developed neural network will be used to build a prediction model for deaf students.

Methodology with Result of each Phases

The methodology followed to build a neural network for regression applications which produces low error rate and which converges quickly is as follows.

- First, find out an regression equation which determines the initial weight of the network.
- Using the regression equation develop an new PSO algorithm for weight adjustment in the neural network.
- Finally build an prediction model for deaf students using the new PSO algorithm and feed forward neural network.

1) Methodology to Find Out a Regression Equation which Determines Optimal Initial Weight for Neural Network

In this step, a regression equation between number of features of a neural network and the weight at which low MSE is produced is found out.

Step 1: Download students performance prediction dataset from UCI repository. It contains academic achievement of Portuguese students. It has totally 33 features.

Step 2: Pre-process the dataset and check the dataset for missing values and outliners. Next the text value in the dataset is converted into numeric values.

Step 3: Using correlation analysis, the correlation coefficient for all the features is found out. Fifteen features (Age+School+Schoolup+Gaurdian+Studytime+G1+Activities+Traveltime+ Nursery+G2+Absence+Famsup+Famsize+Goout+Higher) which has high correlation coefficient is selected for subsequent processing.

Step 4: Next the weight which produce low MSE for each feature is determined as follows.

First, the 1st feature (Age) is taken. A neural network is build and weights 0 to 1 is given. Weight at which the network achieve low MSE is found out. That weight is noted as the weight which produces low MSE. Next the second feature (Age+School) is added and a neural network is built. Again, weights 0 to 1 is given and weight at which the network produces low error is noted. Next, third feature (Age+School+Schoolsup) is added. A neural network is build and weights 0 to 1 is given. The weight which produces low MSE is noted.

This process continuously iterates. At each iteration, one new feature is added and optimum weight which produces low MSE is found out. At end of this step,list of optimum weights at which neural network produce low MSE is prepared. Table 1displays the list of features and the optimum weights which produce low MSE.

Table 1 Weights which produce low MSE

Feature Names - No of Features (x)	**Optimal Weight (y)**
Age - 1	0.02
Age+School - 2	0.03
Age+School+Schoolup - 3	0.02
Age+School+Schoolup+Gaurdian - 4	0.03
Age+School+Schoolup+Gaurdian+Studytime - 5	0.03
Age+School+Schoolup+Gaurdian+Studytime+G1 - 6	0.13
Age+School+Schoolup+Gaurdian+Study time+G1+Activities - 7	0.10
Age+School+Schoolup+Gaurdian+Studytime+G1+Activities+Traveltime - 8	0.11
Age+School+Schoolup+Gaurdian+Study time+G1+Activities+Traveltime+ Nursery - 9	0.10
Age+School+Schoolup+Gaurdian+Studytime+G1+Activities+Traveltime+ Nursery+G2 - 10	0.18
Age+School+Schoolup+Gaurdian+Studytime+G1+Activities+Traveltime+ Nursery+G2+Absence - 11	0.23
Age+School+Schoolup+Gaurdian+Study time+G1+Activities+Traveltime+Nursery+ G2+Absence+Famsup - 12	0.19
Age+School+Schoolup+Gaurdian+Study time+G1+Activities+Traveltime+Nursery+G2+Absence+Famsup+Famsize- 13	0.18
Age+School+Schoolup+Gaurdian+Studytime+G1+Activities+Traveltime+Nursery+G2+Absence+Famsup+Famsize - 14	0.21
Age+School+Schoolup+Gaurdian +Studytime+G1+Activities+Traveltime+ Nursery+G2+Absence+Famsup+Famsize+ Goout+Higher - 15	0.19

Step 5: Using linear regression, a regression equation between number of features and weights which produced low Mean Square Error is found out.
Sum of X = 120Sum of Y = 1.76Mean of X = 8Sum of Y of = 1.76Sum of squares (SSX) = 280Sum of products (SP) = 4.38Regression Equation = $\hat{y}$ = bX + a b = SP/SSX = 4.38/280 = 0.01564
a = MY - bMX = 0.12 - (0.02*8) = -0.00781$\hat{y}$ = 0.01564X - 0.00781

This regression equation is used in the PSO algorithm to determine initial weight of the network

2) New PSO Algorithm for Initial Weight Determination Using the New Regression Equation

Particle swarm optimization is an optimization technique based on the flocking behavior of the birds. In this method each particle performs search for optimal solution which is influenced by its best achieved position, 'pbest' and its group best position' gbest'. First, each particle is initialized with random position vector 'x' and random velocity vector 'v'.

Using fitness function each particle is evaluated. A comparison is made between the particle's current value with its pbest and with the global fitness value, gbest. When the value achieved is better than the previous solution, the current value is set as pbest. Then the pbest value is compared with the gbest, If it is better than the global best, then the current value of the global best is updated [12]. A PSO algorithm for weight adjustment in neural network is proposed below.

1. For each particle, assign weights using the equation w = 0.01564 X - 0.00781 + rand()/2.Store the weights as pbest. Here, rand() denotes random numbers.
2. For each particle, calculate output of the network and calculate the error.
3. Find minimum error particle and store its weight as gbest.
4. Use velocity equation and find velocity (v) of each particle and add velocity to the current weight and get the new weight.

$$v = v + c1*rand()*(pbest-cr)+c2*rand()*(gbest-cr)$$

Here, c1, c2 are constants and cr is the current value.

5. Use new weight and calculate the output of the network and its error.
6. If the new error is less than previous generated error, save the weight as newpbest.

Find minimum error produced particle and check the value with gbest. If the error is less than the gbest, update gbest to reflect the new one.

7. If the error is not less than the previous error, go to Step 4
8. Continue the process for 100iterations
9. At end of 100 iterations use the weight in gbest to make predictions.

The above mentioned algorithm is implemented for weight adjustment in the neural network which is discussed in the coming section.

3)Develop a Neural Network and Use the Developed PSO for Weight Adjustment

Determine the minimum features to build a prediction model.

Step 1: First, a neural network is built with 5 input features and it is trained with the new PSO algorithm. Using MSE, the network is evaluated. Formula for MSE is as follows.

$$Mse = \frac{1}{n}\sum_{1}^{n}(y_i - \hat{y}_i)^2 \ where, y_i \text{isdesired output}, \hat{y}_i \text{isactual output}$$

Step 2: The performance of the new PSO algorithm is compared with basic PSO in terms of MSE and convergence time. The error rates are noted.

Step 3: Next the sixth feature is added and the neural network is build. Error produced is noted. Likewise all the 15 features is added one by one and the neural network is build and it performance is noted. The following tables, tables 2and table 3tabulates the performance of the neural network build with proposed PSO algorithm and basic PSO algorithm. By comparing the proposed algorithms with PSO algorithm, it is found out that the proposed algorithm outperforms existing PSO algorithm in terms of lower error rate and reduced training time.

Table 2 MSE Error Comparison

No of Features	Error achieved (Basic PSO)	Error achieved (New PSO)
5	0.1205	0.1010
6	0.1627	0.1006
7	0.1233	0.0998
8	0.1281	0.1042
9	01907	0.1209
10	0.1734	0.1284
11	0.1922	0.1410
12	0.1900	0.1518
13	0.1832	0.1499
14	0.2095	0.1467
15	0.1812	0.1584

Table 3 Training Time Comparison

No of Features	Basic PSO (In Seconds)	New PSO (In Seconds)
5	24.1783	24.1722
6	24.5167	24.5130
7	24.5821	24.0976
8	24.5191	24.4850
9	25.4739	24.5546
10	24.8531	24.7968
11	24.7407	24.5093
12	24.6047	24.5088
13	24.8910	24.5431
14	24.6618	24.6078
15	25.0526	24.8734

Table 2 shows the MSE produced by existing PSO algorithm and proposed PSO algorithm. The proposed PSO algorithm achieves low error rate than existing PSO algorithm. Low error rate of 0.0998 is produced with seven features. Table 3 show the training time taken by the neural network to complete its training. It is found that the neural network which uses the proposed algorithm for weight adjustment in the neural network takes less training time. The lowest training time of 24.0976 seconds is achieved with seven features by the proposed PSO algorithm. So, these seven features are used to build the prediction model for deaf students.

The below images shows the convergence of basic PSO algorithm and the proposed PSO algorithm for each feature set. The new PSO algorithm is represented in blue line and basic PSO algorithm is represented in pink line. From the below images it can be inferred that the proposed algorithm converges faster than basic PSO algorithm.

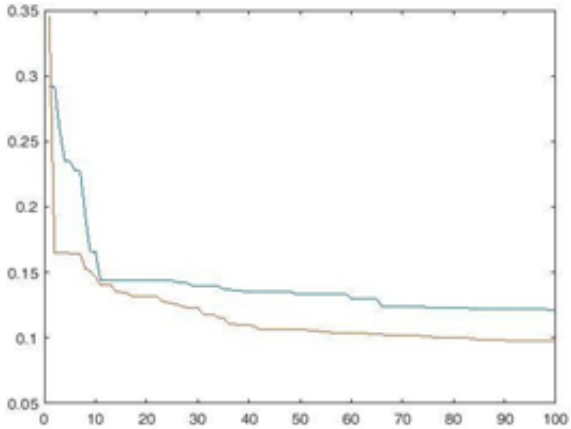

Figure 1 Convergence of 5 Features

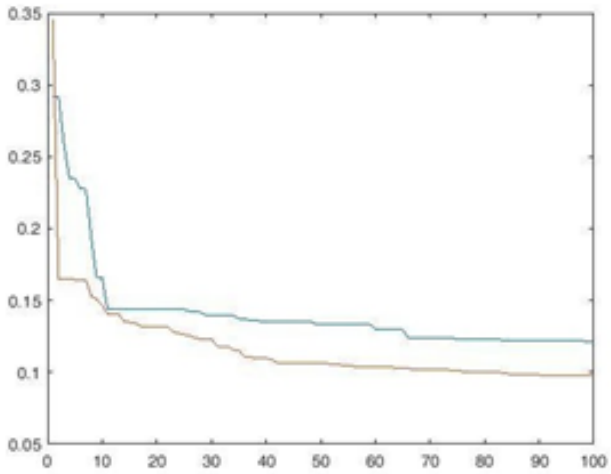

Figure 2 Convergence of 6 Features

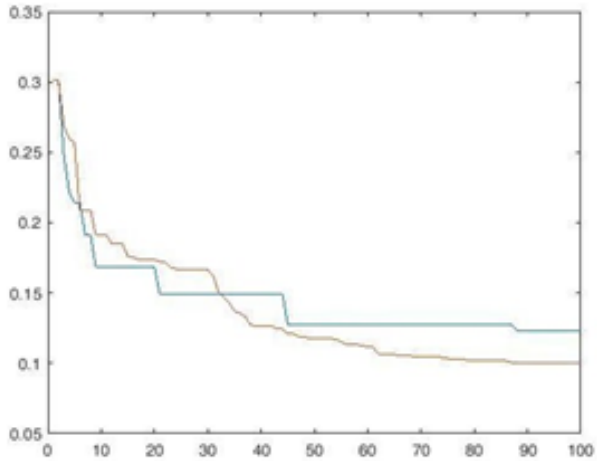

Figure 3 Convergence of 7 Features

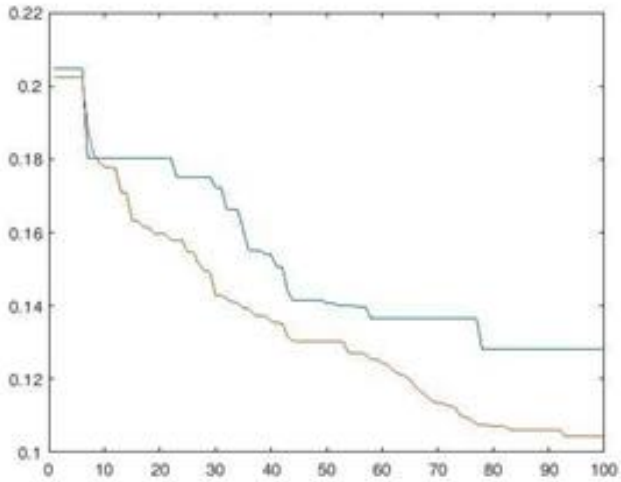

Figure 4 Convergence of 8 Features

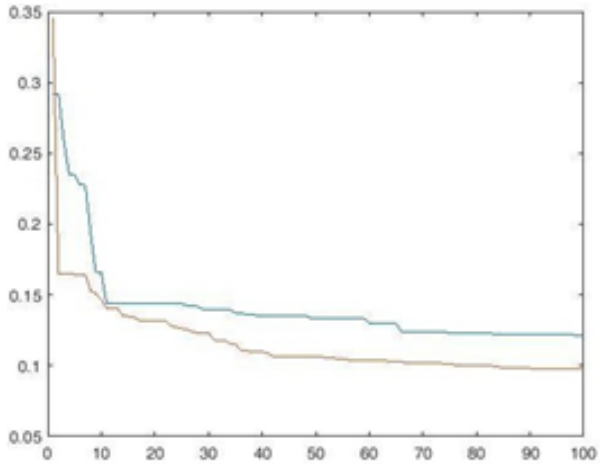

Figure 5 Convergence of 9 Features

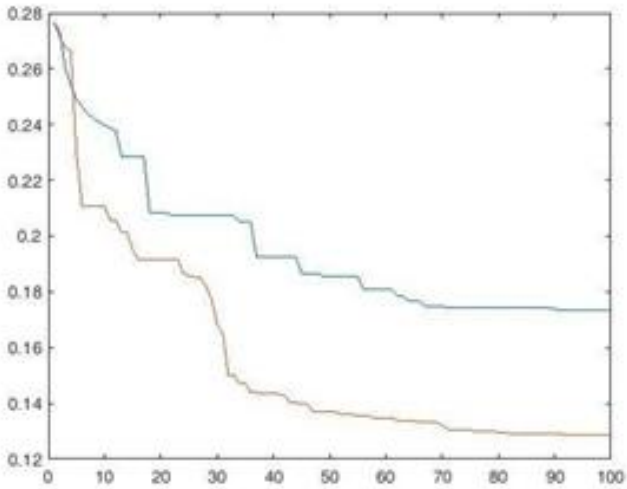

Figure 6 Convergence of 10 Features

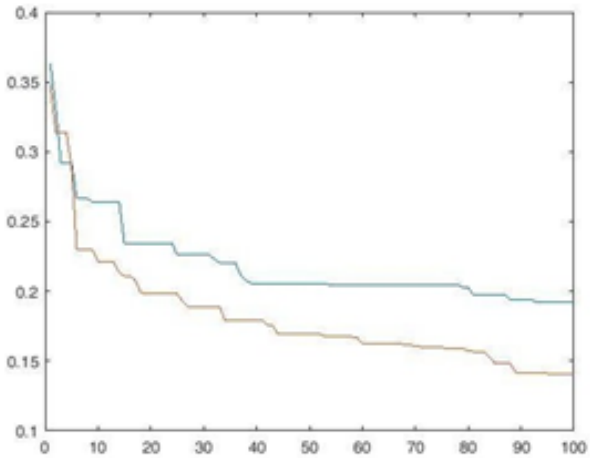

Figure 7 Convergence of 11 Features

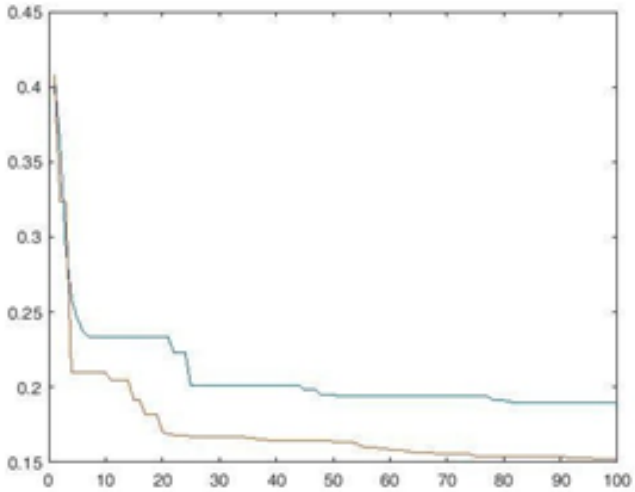

Figure 8 Convergence of 12 Features

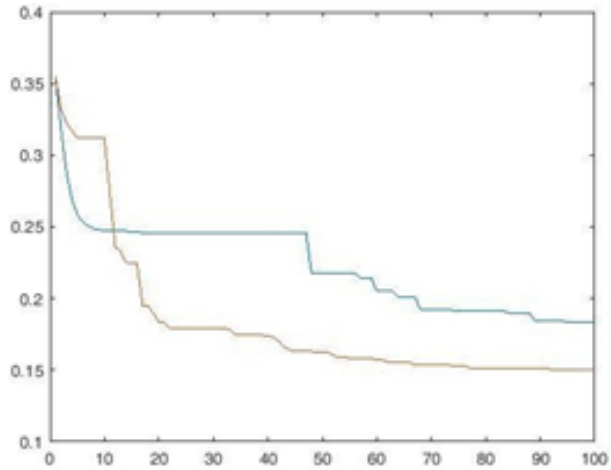

Figure 9 Convergence of 13 Features

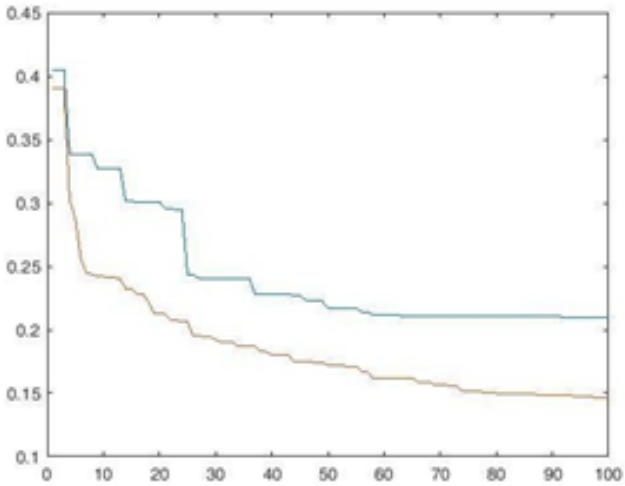

Figure 10 Convergence of 14 Features

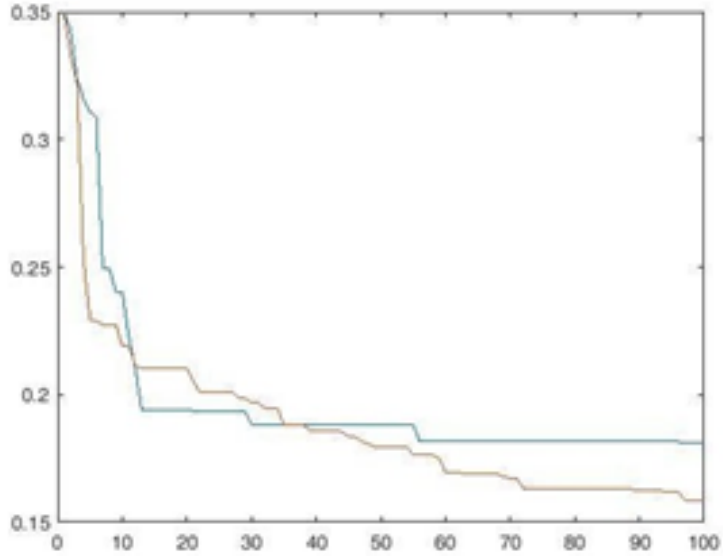

Figure 11 Convergence of 15 Features

4) Building an Prediction Model for Deaf Students Using the New PSO Algorithm and Feed Forward Neural Network

From the previous phase it is found out that seven features namely age, school, school support, guardian, study time, grade 1 mark and extracurricular activities produce low MSE and takes minimum time to converge, which can be inferred from Table 3 and Table 4 respectively. So these features are selected to build a prediction system for the deaf students. The following table shows the various parameters of the prediction system build in this research work.

Table 4

S.No	Parameters	Value
1.	Prediction Technique	Neural Network
2.	Model	Feed Forward
3.	Prediction Type	Regression
4.	Input	7 Units
5.	Hidden Layer	3 Layer
6.	Output	1 Unit
7.	Activation Function	Linear Activation
8.	Training Algorithm	Modified PSO
9.	Particles	50
10.	Iteration	100
11.	C1	1.5
12.	C2	2.5

The prediction model built for these special students is a regression model. It predicts the 10th exam marks from the input features. Data is collected from school of deaf from Chennai and Tiruvanamalai. Instead of grade 1 mark, half yearly marks is collected. With the collected data, a dataset is build. The dataset is used to train and test the neural network. 70:30 split is used in this research work. Matlab 2019a is used to build the

prediction model. A GUI is created using Matlab. In the backend, there is a feed forward neural network which is built with seven features, three hidden layers and one output layer. The neural network is trained with deaf students dataset. Once the network is trained, it is used to make prediction. This research concentrates on predicting the 10th exam mark of deaf student, because tenth standard forms the base of Indian education system. Only if a student passes the tenth standard, he can move up in the education ladder. This is the reason why, the prediction model is used to predict tenth exam marks of the deaf student. The following figure (Fig. 12) shows the GUI of the deaf students prediction system built in this work.

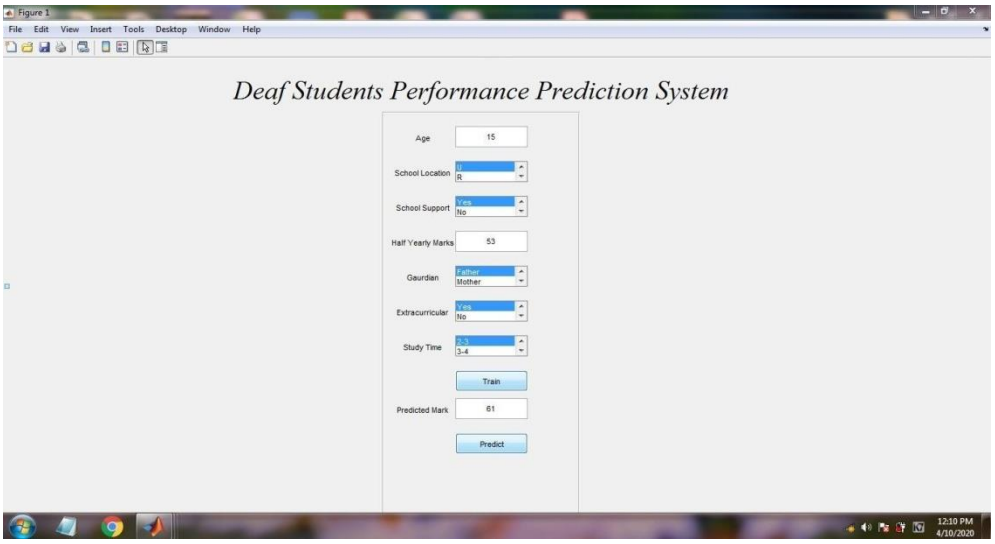

Figure 12 Front of the Deaf students Performance Prediction System

From the figure it can be noted that the system has seven input attributes. Through the front end, attributes values are received and by using the neural network pre trained with deaf students datasets, predictions are made. The train command button, is used to train the neural network and the predict button predicts the tenth exam marks of the deaf student. This prediction model can be implemented in schools of deaf students. If this prediction model is implemented and used, accurate predictions on marks scored by deaf students can be made. Using this system, week learners can be identified earlier and extra coaching can be given to them. Real time implementation of this system in schools of deaf, will aid in increasing the literacy rate of deaf students in India.

Discussions and Conclusion

The literacy rate of deaf students is very less in India. So there is a need to identify weak deaf students as early as possible. In this research work, A prediction model for deaf students which predicts the 10th exam marks of deaf students was built. This model can

be implement in schools of deaf and can be used to identify deaf students who will get low marks in 10th exams. For building this prediction model, a new PSO algorithm which determines initial weight for the neural network using regression equation has been developed. The neural network which implements this modified PSO algorithm for weight adjustment in the neural network performs well in terms of reduced MSE and faster convergence. The reason for better performance is that the regression equation which is used in the PSO algorithm is formulated in advance between number of features in the neural network and weights which produces low MSE. So, when this equation is used, faster convergence is achieved. The error achieved in this work is 0.0998. As future enhancement, prediction model for blind students can be build and other optimization algorithms can be used.

Acknowledgments

I thank Hindustan Institute of Technology and Science, Chennai for providing conducive environment to carry out the research work.

References

Disabled_persons_in_India_2016. PDF File.

Burman, I., & Som, S. (2019). Predicting Students Academic Performance Using Support Vector Machine. *In Proccedings of Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates*, 756-759.

Oloruntoba, S.A., & Akinode, J.L. (2019). Student Academic Performance Prediction Using Support Vector Machine. *International Journal of Engineering Sciences & Research Technology*, *6*(12), 588-598.

Estrera, M., Natan, E., Rivera, T., & Faith, B.C. (2017). Student Performance Analysis forAcademic Ranking Using Decision Tree Approach in University of Science and Technology of Southern Philippines Senior High School. *International Journal of Engineering and Techniques*, *3*(5), 147-153.

Mrinal, P., & Vivek, K.S. (2013). A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction. *International Journal of Computer Applications*, *61*(13), 1-5.

Fahad, R.M., Nareena, S., Shoaib, S., & Safeeullah, S. (2017). Using naïve bayes algorithm to students' bachelor academic performances analysis. *In 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1-5.

Delali, K.D., & Esther, G. (2017). Students Grades Predictor using Naïve Bayes Classifier – A Case Study of University of Education, Winneba. *International Journal of Innovative Research in Science, Engineering and Technology*, *6*(12), 19301-19308.

Lau, E.T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Science Journal*, *1*(982).

Md. Fahim, S., Md. Jamal, U., & Sajal, H. (2016). Predicting Students Yearly Performance using Neural Network: A Case Study of BSMRSTU. *In 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 524-529.

Bogdan, O., Raluca, D., & Stefan, C. (2013). Predicting students' results in higher education using neural networks. *InInternational Conference on Applied Information and Communication Technologies*, 190-193.

Ramanathan, L., Geetha, A., Khalid, M., & Swarnalatha, P. (2016). Student Performance Prediction Model Based on Lion-Wolf Neural Network. *International Journal of Intelligent Engineering and System*, *10*(1), 114-123.

De Almeida, B.S.G., & Leite, V.C. (2019). Particle swarm optimization: A powerful technique for solving engineering problems. *In Swarm Intelligence-Recent Advances, New Perspectives and Applications. IntechOpen*.