Prediction of Student's Performance with Deep Neural Networks

Meryem Karlık *drmeryemk@gmail.com*
Department of Foreign Languages,
Silk Road International University of Tourism,
Samarkand, Uzbekistan

Bekir Karlık *bekir.karlik@mail.mcgill.ca*
McGill University, Neurosurgical Simulation
Research & Artificial Intelligence Learning Centre,
Montréal, QC, Canada

Abstract

The performance of education has a big part in people`s life. The prediction of student's performance in advance is very important issue for education. School administrators and students` parents impact on students' performance. Hence, Academic researchers have developed different types of models to improve student performance. The main goal to reveal of this study is to search the best model of neural network models for the prediction of the performance of the high school students. For this purpose, five different types of neural network models have been developed and compared to their results. The data set obtained from Taldykorgan Kazakh Turkish High School (in Kazakhstan) students was used. Test results show that proposed two types of neural network model are predicted students` real performance efficiently and provided better accuracy when the test of today's and future's samples have similar characteristics.

Keywords: Deep Learning, Prediction, Student Performance, Fuzzy Clustering Neural Networks.

1. INTRODUCTION

The educational system always needs to improve the quality of education to achieve the best results and reduce the percentage of failure. The exist of facing difficulties and challenges of educational systems need to analyze effectively. Moreover, in order to improve the prediction process selection of the best prediction technique is also very important. By using machine learning algorithms, it is easy to solve the lack of prediction accuracy, improper attribute analysis, and insufficient datasets.

Forecasting performance evaluation is a regression problem. A regression task is a task to predict the value of a continuous output parameters in basis of a certain input dataset. Always the regression problem arises when predicting one or more than one values, which can continuously change within certain value interval. In the last decades, a number of research papers are presented and comprised different machine learning models for predicting student performance at various levels. Yusof et al [1] have presented a fuzzy inference system (ANFIS) model for evaluation of student's performance and Learning Efficiency. This work was focused on a systematic approach in assessing and reasoning the student's performance and efficiency level in the programming technique course. Sevindik [2] has also presented an assessment model based on the adaptive neuro-fuzzy inference system (ANFIS) for the prediction of the students` academic performance. The data used in his work was collected from the students of computer and instructional technology education. This Data sets consist of exam 1, exam 2, project, final and attitude points of the students at graphics and animation course for education in one semester. Similarly, Alanzi [3] has proposed a cascade correlation neural network model trained by the Quick Propagation algorithm for performance prediction of students. Her data set consists of student's parameters such as grade point average, score of medical college admission test,

and structure of personal interview. Liveris et al [4] have proposed a user-friendly software tool based on neural networks for predicting the students' performance in the Math course. Oladokun et al [5] have presented a neural network model based on the Multilayer Perceptron Topology was developed and trained using data spanning five generations of graduates from an Engineering Department of University of Ibadan, in Nigeria.

Sabourin et al [6] have described a preliminary investigation of self-regulatory, and more specifically metacognitive, behaviors of students in a game-based science mystery. In their study, Data was collected from 296 middle school students interacting with Crystal Island. This data was used to classify students into low, medium, and high self-regulated learning behavior classes by using different supervised machine learning classifiers. Karlik [7-8] has determined English learning abilities of students among affective factors which analyzed by using Neural Networks in freshman class of two university students in Bosnia and Herzegovina.

Moreover, Kabakchieva [9] has focused on the development of machine learning models for predicting student performance, based on their personal, preuniversity and university-performance characteristics. In her proposed study, several well-known supervised machine learning algorithms such as decision tree classifier, neural network, and Nearest Neighbor classifier were used and compared. Ashraf et al [10] have presented a comparative study of various machine learning algorithms for predicting student performance. They have used decision trees [ID3, c4.5, and CART], Bayesian network classifier, Naïve Bayes classifiers, MLP, NB and J48 algorithm, logistic regression, conventional neural network, clustering/classification, association rule, and NBtree classification algorithms.

Parveen and Quadri [11] have presented a review of studies previously done by different authors on student performance by using different well-known machine learning methods in 2019. However, to date, no deep learning method has been used. In this study, we present different deep neural network models to predict student performance. The main goal of this study is to develop and compare for the prediction of the performance of High School students by using the best ANN algorithms. Dataset was collected with the survey from Taldykorgan Kazakh Turkish High School students in Kazakhstan.

2. WHAT`S DEEP NEURAL NETWORK?

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological neural systems. It is composed of a large number interconnected processing elements (or called as artificial neurons). The processing elements in back-propagation neural networks are arranged in layers that contain an input layer, an output layer, and a number of hidden layers generally [12]. The most popular ANN consists of a back-propagation algorithm in a supervised learning paradigm in which the generalized delta rule was used in updating the weight values [13].

Many neural network structures and training algorithms exist in the literature. Deep learning is a field of study involving machine learning algorithms and artificial neural networks that consist of one or more hidden layers. In other words, Deep learning is a class of machine learning algorithms that use a cascade of many layers of nonlinear processing units for feature extraction and transformation as seen in Fig. 1.

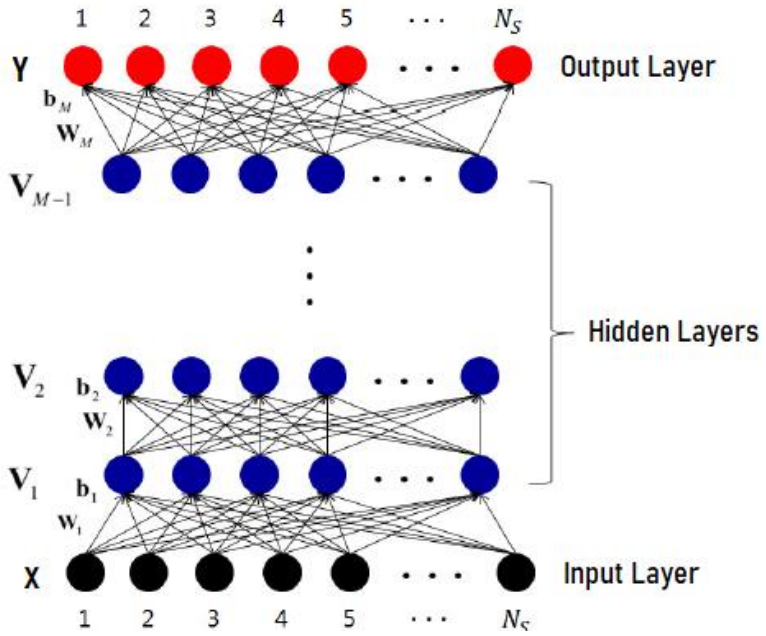


FIGURE 1: The Architecture of Deep Neural Network.

According to Le Chun et al [14]; Deep learning discovers intricate structure in large data sets by using the back-propagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep learning is also called as deep structured learning, hierarchical learning, deep machine learning, or Deep Neural Networks (DNN). DNN is mostly used for big data problems or insoluble problem by conventional ANN.

Types of DNN are; Back-propagation to train multilayer architectures, Convolutional neural networks (or Le-Net) [15], and Recurrent neural networks. Recently, some new DNN algorithms have been developed by different researchers. These are Back-propagation Algorithm with Variable Adaptive Momentum as supervised learning [16], and as unsupervised algorithms, the Parameter-Less self-organizing map (PLSOM) [17], PLSOM2 [18], and Robust adaptive learning approach to self-organizing maps [19-20]. DNN was inspired by the limitations of the conventional neural networks algorithms especially being limited to processing data in raw form, and by updating the weights of the simulated neural connections on the basis of experiences, obtained from past data [21]. DNN has been successfully used in diverse emerging domains to solve real-world complex problems with may more deep learning models, being developed to date. To achieve these state-of-the-art performances, the DNN architectures use activation functions, to perform diverse computations between the hidden layers and the output layers of any given deep neural learning architecture [22-23].

Another efficient way of neural network architecture is to use hybrid models. There is a various hybrid classifier algorithm such as Adaptive Neuro-Fuzzy Inference System (ANFIS), Fuzzy Clustering Neural Network (FCNN), Particle Swarm Optimization based Neural Network (PSONN), Principal Component Analysis based Neural Network (PCANN), Vector Quantization Neural Network (VQNN), etc. Fuzzy Clustering Neural Network (FCNN) is one of the best hybrid models which consists of unsupervised fuzzy c-means and supervised Backpropagation neural network [24-25].

In the last decades, neural networks algorithms are used to make predictions of student performance from observations of survey data. Five different types of neural network models were used in this study. These are; Multi-Layered Perceptron (MLP) structured as 9:9:1 with unipolar sigmoid activation function (named ANN1), MLP structured as 9:9:3 (named ANN2),

FCNN consisted of a Fuzzy C-means and MLP structured as 9:10:1, DNN1 structured 9:9:9:1 and DNN2 structured as 9:9:9:3. Where, 9 is a number of the neuron for input layer, for both number of the neuron of hidden layers are 9, and number of the neuron of output layer is 3 or 1. If output neuron is 1; output values are described as 1, 2 and 3 numerically. For the other, output neurons are described (100), (010) and (001) logically.

All used software has been developed by Karlik [26-28]. All Proposed ANN models have a unipolar sigmoid activated function. Optimum learning rate and momentum coefficient were found as 0,3 and 0,2 respectively.

3. METHODOLOGY OF PREDICTION OF STUDENT`S PERFORMANCE

This chapter presents all used methodologies which is starting from data collection, training and testing results of each used neural networks, comparing of their results.

3.1 Data Collection

The data set was collected from 11th grade students of Taldykorgan Kazakh Turkish High School in Kazakhstan between 2012 and 2013. The input variables selected are those which can easily be obtained from students by survey questions. These questions are:

Q1- Do you live in dormitory? (Yes or No)
Q2- Is your father alive? (Yes or No)
Q3- Is your mother alive? (Yes or No)
Q4- Do your parents live together? (Yes or No)
Q5- During last 5 years, how many times teachers visit your home?
Q6- Did your teacher help you enough during your education in school? (give a mark out of 10)
Q7- Give a mark to your Class-teacher? (give a mark out of 10)
Q8- Did you get your expectations in this school? (give a mark out of 10)
Q9- What is your national university exam (N.U.E) result? (out of 100)

TABLE 1 shows some samples of dataset for the first 10 students. Where each rows represents a student, and each column represents their answers for survey questions. All dataset was normalized between 0 and 1. The output variables represents the performance of each student as low, middle, and high [29].

TABLE 1: Sample of input variables for the first 10 students.

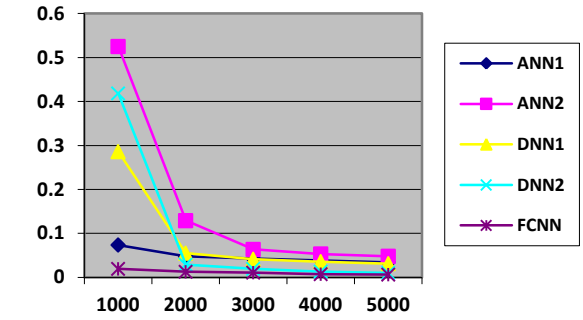
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Student 1	1	1	1	0	4	8	9	10	83
Student 2	1	1	1	1	2	7	8	7	76
Student 3	1	1	1	0	2	2	10	10	94
Student 4	1	1	1	1	3	10	10	10	92
Student 5	1	1	1	1	1	10	10	10	92
Student 6	1	1	1	0	2	6	10	10	92
Student 7	1	1	1	1	6	8	10	10	86
Student 8	1	1	1	1	1	10	10	10	97
Student 9	1	1	1	1	6	10	9	10	88
Student 10	1	1	1	1	3	8	10	10	86

The 10% KDD`99 dataset has 494.021 records with many duplicates. In our experiments we removed the duplicated data and the number of records has been dropped to 145.585. Then we converted the attributes with text data to numeric values and applied normalization by scaling each attribute between 0 and 1.

4. EXPERIMENTAL RESULTS

In this study, two ANN, one hybrid, and two DNN algorithms were used for predicting student`s performance. All of these models have different multi-layered perceptron (MLP) architectures and

trained with back-propagation supervised learning algorithm. The survey in students, keeping in mind the particular problem, is used as input of these MLP architectures. Our input variables are; staying dormitory, parent's life, exam mark, student's mark to his teacher, expectation of student and teacher`s support. The target (or desired) outputs of MLP were the level of performance of students as high, medium and low performances. Four different neural network models were used for survey data set which obtained from high school students. Figure 2 shows Comprising results for each ANN models according to iteration (from 1000 to 5000)



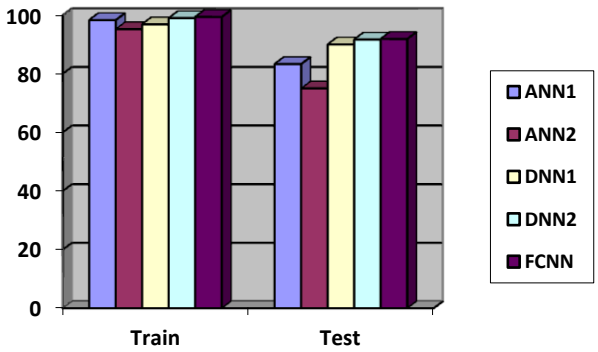
.

FIGURE 2: Comprising MSE error according to iteration.

As seen in Fig. 2, DNN has less performance in the beginning. But it is being better than the others while increasing the iteration. Moreover, beginning of both ANN models are better than both DNN models. But after 2000 iterations DNN are became much better. The best performance was obtained from beginning to the last iteration for FCNN model as seen Fig. 2.

TABLE 2 shows training and test accuracies for four neural network models. FCNN model has not only the best training accuracy (99,38%) but also the best test accuracy (91,877%). The second- best training accuracy and test accuracy were found as 99,38% and 91,877 respectively for DNN2. Similarly, training accuracy was 98,9737% and test accuracy were 91,6666 for DNN2, training accuracy was 97.0% and test accuracy was 90,0% for DNN1. But DNN1 and DNN2 training time took too long because of their complicated MLP structures. ANN1 can be useful to solve such a problem. Moreover, its MLP structure is simple and training is faster than the other models. The training accuracy was 98,3079% and test accuracy were 83,3333% for ANN1. Similarly, training accuracy was 95,2051% and test accuracy was 75,0% for ANN2.

TABLE 2: Comparison of accuracies for all algorithms.



One by one all students are trained and they are evaluated as high, middle and low performance. After analyzing each test dataset, we observed the following results:

If student lives in dormitory, both parents are alive and satisfied with the school are affected to them positively. Moreover, the teacher visits student and expectation from school are less affected. However, if the student has lost one of his (or her) parents or parents not live together, these situations are affected to students negatively.

5. CONCLUSION AND DISCUSSION

In this study, an evaluation of students` performance by using conventional artificial neural networks and deep neural networks have been presented. For this purpose, five different neural networks models named ANN1, ANN2, DNN1, DNN2, and FCNN were proposed. For testing, we used twelve students` survey input variables. The best performance was obtained from FCNN model which has been recognized student's performance successfully with this model. We have also obtained very high performance from DNN2 model. ANN1 and ANN2 conventional neural network models could not show high performance because of the complexity of the dataset.

One of students couldn't show his\her real performance regarding our test results. It can be necessary to add an extra reason for the student's performance. Because there are some criterions which are bad effect such as health problem, a student absent, passed away one of student`s parent, etc. which can be negative effects.

TABLE 3 shows comparison of various neural network model results between our study and developed the other authors' studies in the literature as defined in ref [11]. The higher accuracies of ordinary neural network models are 89.6% (Huang, 2013), 83.33% (Proposed ANN1), and 86.11% (A.T.M. Shakil Ahmad, 2017). As seen the result in Table 3, it can be observed that both proposed deep neural networks model (FCNN, DNN1, and DNN2) are better accuracies than ordinary ANN models. The best models are proposed FCNN and DNN2 having accuracies of 91.877% and 91.67% respectively.

Student's performance is very important point during human life. By the correction of this problem one person will be gained to the life with real high performance. For solving this problem, the other machine learning methods can be also used for future work

TABLE 3: Percentage accuracy of used neural network algorithms.

Authors	Attributes	Accuracy
(Kaur, 2015)	Personal and Academic Information	75.0%
(A.T.M. Shakil and Ahmad, 2017)	Demographic, Psychological and Academic Information	86.11%
(Mueen, 2016)	General, Forum and Academic Information	81.4%
(Huang, 2013)	Midterm and Final Exam Scores	89.6%

(Kotsiantis, 2004)	Demographic, Marks in Assignments	72.26%
(Rustia, 2018)	Subject Areas and LET Results	65.67%
(Agrawal, 2015)	SS Grade, Living Location, Med. of Teaching	70.0%
Proposed ANN1	Personal, Academic Info, Living Location, and NUE Results	83.33%
Proposed ANN2	Personal, Academic Info, Living Location, and NUE Results	75.0%
Proposed DNN1	Personal, Academic Info, Living Location, and NUE Results	90.0%
Proposed DNN2	Personal, Academic Info, Living Location, and NUE Results	91.67%
Proposed FCNN	Personal, Academic Info, Living Location, and NUE Results	91.877%

6. REFERENCES

[1] Norazah Yusof, Nur Ariffin Mohd Zin, Noraniah Mohd Yassin, Paridah Samsuri, Evaluation of Student's Performance and Learning Efficiency based on ANFIS, International Conference of Soft Computing and Pattern Recognition, pp. 460-465, 2009.

[2] Sevindik Tuncay, Prediction of student academic performance by using an adaptive neuro-fuzzy inference system, *Energy Education Science and Technology Part B-Social and Educational Studies*, vol. 3, Issue: 4, pp. 635-646, October 2011.

[3] J. K. Alenezi, M. M. Awny and M. M. M. Fahmy, The Effectiveness of Artificial Neural Networks in Forecasting the Failure Risk: Case Study on Pre-medical Students of Arabian Gulf University, International Conference on Computer Engineering & Systems, pp. 135-138, 14-16 Dec. 2009.

[4] Ioannis E. Livieris, Konstantina Drakopoulou, Panagiotis Pintelas, Predicting Students' performance using artificial neural networks, 8th PanHellenic Conference with International Participation Information and Communication Technologies in Education, pp. 321-328, September 2012, Volos, Greece

[5] Victor Oladokun, A T Adebanjo, O E Charles-Owaba, Predicting Students Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course, he Pacific Journal of Science and Technology, vol. 9, no. 1, pp. 72-79, May-June 2008.

[6] Jennifer L. Sabourin, Lucy R. Shores, Bradford W. Mott & James C. Lester, Understanding and Predicting Student Self-Regulated Learning Strategies in Game-Based Learning Environments, *International Journal of Artificial Intelligence in Education*, vol. 23, pp. 94–114, 2013.

[7] Meryem Karlik, Artificial Neural Networks Methodology for Second Language Learning, LAP LAMBERT Academic Publishing, May 5, 2017.

[8] Meryem Karlık & Azamat Akbarov, Investigating the Effects of Personality on Second Language, Learning through Artificial Neural Networks, *International Journal of Artificial Intelligence and Expert Systems*, vol. 7, Issue. 2, pp. 25-36, 2016.

[9] Dorina Kabakchieva, Student Performance Prediction by Using Data Mining Classification Algorithms, *International Journal of Computer Science and Management Research*, vol. 1, Issue 4, pp. 686-690, November 2012.

[10] Aysha Ashraf, Sajid Anwer, and Muhammad Gufran Khan, A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques, *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 44(1), pp. 122-136, 2018.

[11] Zeba Parveen, Mohatesham Pasha Quadri, Classification and Prediction of Student Academic Performance using Machine Learning: A Review, *International Journal of Computer Sciences and Engineering*, vol.7, Issue.3, pp.607-614, 2019.

[12] Ozbay Yuksel, Karlik Bekir, A Fast Training Back-Propagation Algorithm on Windows, 3rd International Symposium on Mathematical & Computational Applications, pp. 204-210, 4-6 September, 2002, Konya, Turkey.

[13] Karlik Bekir, Differentiating Type of Muscle Movement via AR Modeling and Neural Networks Classification of the EMG, *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 7, no.1-3, pp. 45-52, 1999.

[14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, Deep learning, *Nature*, vol. 521, pp. 436–444, 2015.

[15] Karlik Bekir, Uzam Murat, Cinsdikici Muhammet, and Jones A. H., Neurovision-Based Logic Control of an Experimental Manufacturing Plant Using Convolutional Neural Net Le-Net5 and Automation Petri Nets, *Journal of Intelligent Manufacturing*, vol. 16(4-5), 527-548, 2005.

[16] Hameed Alaa Ali, Karlik Bekir, Salman M. Shukri, Back-propagation Algorithm with Variable Adaptive Momentum, *Knowledge-Based Systems*, vol.114, pp.79–87, December 2016.

[17] Erik Berglund and Joaquin Sitte, The parameter-less self-organizing map algorithm, *IEEE Transactions on Neural Networks*, vol. 17(2), pp. 305–316, March 2006.

[18] Erik Berglund, Improved PLSOM algorithm, *Applied Intelligence*, vol. 32, Issue 1, pp 122–130, February 2010.

[19] Hameed Alaa Ali, Karlik Bekir, Salman M. Shukri, Eleyan Gulden, Robust Adaptive Learning Approach of Self-Organizing Maps, *Knowledge-Based Systems*, vol. 171 – May 1, pp. 25-36, 2019.

[20] Hameed Alaa Ali, Ajlouni Naim, Karlik Bekir, Robust Adaptive SOMs Challenges in a Varied Datasets Analytics, Springer Nature Switzerland, Vellido et al. (Eds.): WSOM 2019, AISC 976, pp. 110–119, https://doi.org/10.1007/978-3-030-19642-4_11

[21] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[22] Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall, Activation Functions: Comparison of Trends in Practice and Research for Deep Learning, arXiv.org > cs > arXiv:1811.03378

[23] Karlik Bekir and Olgac Ahmed Vehbi, Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks, *International Journal of Artificial Intelligence and Expert Systems*, vol. 1, no. 4, pp. 111–122, 2011.

[24] Ceylan Rahime, Ozbay Yuksel, Karlik Bekir, Comparison of Type-2 Fuzzy Clustering Based Cascade Classifier Models for ECG Arrhythmias, *Biomedical Engineering: Applications, Basis and Communications (BME)*, vol. 26, no. 6, 2014, 1450075

[25] Karlik Bekir, The Positive Effects of Fuzzy C-Mean Clustering on Supervised Learning Classifiers, *International Journal of Artificial Intelligence and Expert Systems*, vol. 7(1), pp. 1-8, 2016.

[26] Karlik Bekir, Myoelectric control using artificial neural networks for multifunctional prostheses, PhD Thesis, Yıldız Technical University, Istanbul, Turkey, March 1994.

[27] Karlik Bekir, Tokhi Osman, Alci Musa, A Fuzzy Clustering Neural Network Architecture for Multi-Function Upper-Limb Prosthesis, *IEEE Trans. on Biomedical Engineering*, vol. 50, no:11, pp.1255-1261, 2003.

[28] Karlık Bekir, Koçyiğit Yücel, Korürek Mehmet, differentiating types of muscle movements using a wavelet based fuzzy clustering neural network, *Expert Systems*, vol. 26 (1), pp. 49-59, 2009.

[29] Alaka Halil, Master Thesis, Evaluation of Student Performance in Kazakhistan than Using Neural Networks, Kazakhstan Suleyman Demirel University, Almaty, Kazakhstan, July 2013.