WILEY | Hindawi

Research Article

Performance Prediction for Higher Education Students Using Deep Learning

Shuping Li and Taotang Liu ⓘD

Mudanjiang Normal University, Mudanjiang 157011, China

Correspondence should be addressed to Taotang Liu; 2015053@mdjnu.edu.cn

Predicting students' performance is very important in matters related to higher education as well as with regard to deep learning and its relationship to educational data. Prediction of students' performance provides support in selecting courses and designing appropriate future study plans for students. In addition to predicting the performance of students, it helps teachers and managers to monitor students in order to provide support to them and to integrate the training programs to obtain the best results. One of the benefits of student's prediction is that it reduces the official warning signs as well as expelling students because of their inefficiency. Prediction provides support to the students themselves through their choice of courses and study plans appropriate to their abilities. The proposed method used deep neural network in prediction by extracting informative data as a feature with corresponding weights. Multiple updated hidden layers are used to design neural network automatically; number of nodes and hidden layers controlled by feed forwarding and backpropagation data are produced by previous cases. The training mode is used to train the system with labeled data from dataset and the testing mode is used for evaluating the system. Mean absolute error (MAE) and root mean squared error (RMSE) with accuracy used for evolution of the proposed method. The proposed system has proven its worth in terms of efficiency through the achieved results in MAE (0.593) and RMSE (0.785) to get the best prediction.

1. Introduction

Today, increasing importance is given to predicting student performance due to the great importance of this issue in the development of countries around the world because it depends entirely on the educational process that leads to the production of a generation capable of taking the responsibility of leading this country and its march towards development in all aspects of life (scientific, economic, social and military, etc.). Also, the evaluation of students' performance is a reflection of the efficiency of educational institutions which is responsible for developing successive generations in line with the different stages of the lives of people in every country. Therefore, focusing on the development of the educational process is one of the utmost necessities that push governments represented by educational institutions to make tremendous and painstaking efforts to push the educational process towards continuous and escalating development.

Future knowledge can be obtained through prediction. The higher the amount of data is, as in large databases, the better the prediction is produced; this process is known as data mining which is used to identify hidden information by exploring different data sources related to different fields such as commercial, social, medical, and educational [1]. The knowledge presented by different resources of educational data can be analyzed to extract desired information. A new discipline termed by Educational Data Mining (EDM) was developed as a method of discovering valuable information [2]. The best learning environment results from sharing statistics and exploration with deep learning. The importance of EMD has increased rapidly in the present day because of the increase in the collected data, according to the educational data acquired from different e-learning systems,

as well as the development of traditional educational systems. The power of EMD is represented by various data in different fields and how they are connected together. It concerns about the extraction of features to support the progression of educational process from huge data provided from institute [3]. Unlike the investigation of traditional database, which can answer questions, such as who is the student who failed in the exam? EMD can answer more deep questions such as the prediction of the result of the student (if he will pass or fail in the exam).

Academic institutions try to build their student's model to predict both characteristics and performance of each student individually [4]. Therefore, the researchers who are involved with the EDM field use different techniques of data mining in order to assess lecturers, to conduct their educational organization. Because the prediction of student's performance in current educational systems is not given the necessary importance, these systems suffer from a lack of efficiency. The process of predicting the lessons that the student is interested in and knowing his activity in educational institutions leads to raising educational efficiency. Through machine learning and EDM techniques, the evaluation process of students is carried out continuously in many educational institutions. These evaluation systems are very feasible to improve the student performance and then the whole educational process [5].

The researchers found that deep learning can be used in many different areas, such as pattern recognition, image processing, object detection, and natural language processing. In learning management system, data mining can be used to obtain better and more accurate results. This research proposes a deep learning approach to build a model prediction of student performance by implementing convolutional neural network on a one-dimensional data ( ) in addition to long short-term memory (LSTM) to predict the student performance in next semesters based on his performance in the previous semesters. In this research, data preprocessing techniques (such as mini-max scaler and quantile transforms) were introduced to increase results' accuracy in addition to other factors such as incentive activity grade and entrance English testing.

The aim of this study is to extract new features and find their weights to build a neural network system of variable nodes and hidden layers according to derived weights from features. Then, these features and their weights are used after building the system to predict information regarding students.

2. Deep Learning Types

2.1. Conventional Neural Network.
CNN's algorithm has been widely used in all fields of image recognition due to the ability of the algorithm to recognize various full behaviors [6]. Thus, its use was extended to include the education and learning prediction process. In this sense, similar to the neural network, CNN generally consists of connected and multiple neurons in a hierarchical form, and through training, the layered structure is completed.

DNN differs from on network to another in connection such as deep brief network, backpropagation, and sparse autoencoder; each layer in the network can share the weight for each neuron, so weight can control the layers in the network.

CNN is used for student behavior by extracting new features on specific time point, and the feature considers the characteristic from education status [7].

2.2. Recurrent Neural Network (RNN).
It considers one of the neural network algorithms and performs well with sequence of data. One of the best feature in this algorithm is the ability to memorize the previous state to be used in the current state or the next state [8].

In addition to the hidden layer, there are dynamic input and output layer; inside the hidden layer there are input and output cases represented by the output weight from one node to another. The advantage of such algorithms during the training is suitable for prediction due to connection and feedback paths in hidden layers.

2.3. Long Short-Term Memory (LSTM).
Its model is defined as the variant of the RNN. The worth of this algorithm is the self-loop created in the hidden layers, and it automatically generates the paths during running the system in addition to the generate short path in each iteration. It is like DNN but different in updating the weight that affects the sort path in neural network.

This type uses previous data from history to extract useful information (features) to achieve better student behavior prediction [9].

2.4. Deep Neural Network (DNN).
It is a neural network with more than one hidden layer. Its model performs better with complicated data and nonlinear function. This type of deep learning can adapt with any improvement to the hidden layers during training and the training going through backpropagation algorithm. Since DNN is good with scalable data in prediction model that uses complex data, it is considered suitable for education deep learning prediction [10, 11].

3. Related Work

A critical increase in predicting student performance requirements is a result of the increasing interest by universities to raise the level of student performance and education to keep pace with the development taking place in various aspects of current life [12]. Various techniques were applied to predict student performance; each of these techniques belongs to different areas such as artificial neural network, machine learning, and collaborative filtering.

Effectiveness of transfer from deep neural network learning was investigated in order to predict the performance of a higher-education student by [13]. Experiment was organized based on five compulsory courses dedicated for two undergraduate systems. Empirical results show that

the prognosis of students at risk of failure can be achieved with good accuracy in the majority of cases.

The system of student performance prediction based on *MyMediaLight* which is an open source recommendation system was proposed by [14]. This system is applied for GPA database of the student that was collected previously from university. A technique of biased matrix factorization (BMF) was proposed by authors to predict student performance to assess them and to choose appropriate courses.

Techniques' combination was suggested in [15] to develop a new prediction for a student performance model. The gray model and the Taylor approximation method were combined together to achieve the best results by computing approximations several times to improve the predictive accuracy of two gray models. The results obtained from this research can help both education administrators and educators to choose better solutions to raise the performance of a student who is experiencing instability in the learning process, as well as the matrix factorization, restricted Boltzman machine, and collaborative filtering techniques was used by [16] to analyze systematically the data that has been collected from the academic management system. The obtained results show that the better technique among the previously mentioned techniques is the restricted Boltzmann machines.

Due to its effectiveness and simplicity, the collaborative filtering algorithm is used in the recommendation system. However, effectiveness of such techniques is limited due to the data sparsity, and it restricts the further improvement of prediction results. Thus, there is more interesting on the model that consist of combination of deep learning and historical data prediction algorithm. To achieve more accurate latent features, a model based on the quadratic polynomial regression model was proposed in [17]; in this model, traditional algorithm of matrix factorization was improved; then, the input data into deep neural network is the latent features. Implementation of the proposed model on three different datasets show significant improvement in efficiency prediction compared to the traditional models. The new model proposed in [18] consists of combination between deep learning and collaborative filtering model. The feedback in the neural network during the prediction process works in the form of simulations of the interaction process between the student and the educational institution. The preprocessed features will used as the input of neural network. The proposed model is implemented on ten million samples of *MovieLens* dataset and dataset of one million samples of *Movielens* to verify the performance of the proposed model which obtained very good results. Other approaches were suggested to improve prediction of student performance which is found in [19, 20].

Numerous research studies were proposed previously; these research studies take into account the issue of prediction of student performance by using the theory of machine learning; there is still room for improvement which is the student performance prediction factor analyzing based on data the transformation technique and the explanation model. The main aim of this research proposed a new approach by taking into account deep learning technique represented by long short-term memory (LSTM) with the use of time-based features.

4. Why Deep Neural Network?

Many advantages allow us to choose DNN, and there are many reasons listed below:

DNN models provide high accuracy with the result when comparing with other methods such as regression

DNN models can learn complex nonlinear with also sequence data in additional to data from updated function

DNN models easily assess the accuracy and significance by using mean square error (MSE) and R^2

DNN models easily handle the nonparametric method without prior knowledge of distribution and *I/O* mapping function

DNN models are flexible when dealing with vast data for prediction

DNN models are easy to update according to the changing environment to be dynamic and suitable

Deep learning is used to understand the behavior of the data and predict accordingly, so this information can be useful in the near future, as illustrated in Figure 1.

5. Deep Learning

With the development in data science and modern technology, such as big data and high-performance computers, an opportunity is provide for machine learning to understand data and behavior of it through complex systems. Machine learning gives the machine ability to learn in different algorithms without strict orders from a certain program or limited instruction [21].

Deep learning can be defined as technique of machine learning to learn useful features directly from given different media or problems. Many layers are exploited by deep learning for nonlinear data processing of unsupervised or supervised feature extraction for classification and pattern recognition [22]. Deep learning motivation is greatly reduced by artificial intelligence (AI) area, which simulates the ability of the human brain in terms of analyzing, making decision, and learning. Deep learning goal is to emulate the approach of hierarchical learning of extracting features by the human brain directly from unsupervised data.

The core of deep learning is the hieratically computing the features and representation of information, such as defining of the features starting from the lowlevel to the high level. With huge data obtained from previous student performance, the standard techniques of machine learning do not work well when run directly due to ignoring the nature of data behavior. In deep learning, features are extracted automatically from given student data. The characteristics of this method of features considered part of learning system [23].

Characterization of input such data used as a feature is the key issue to success of processing prediction of the future
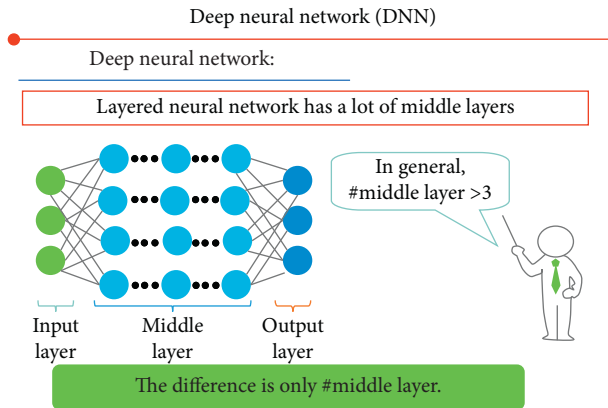
FIGURE 1: The true understanding of using DNN in stock market prediction.

state. There is a limitation for extracted features' student performance such as CGPA of previous semesters and number of credit the student has earned from previous semesters [24]. For this reason, we can use deep learning by its feature extraction to solve limitation in such systems.

As mentioned before, the main difference between machine learning and deep learning is the features' selection method, as shown in Figure 2.

Features in deep learning will be generated automatically to simulate the appropriate results [25]. Different hidden layers participate in making decision by using the feedback from certain layer to the previous layer to get better result [23]. DL enables computers to be able to perform complex calculations by relying on simpler calculations to optimize computer efficiency. It is difficult for a computer to understand complex data such as collection data from literature or a series of data of a complex nature, so we use deep learning algorithms instead of usual learning methods [26].

6. Data Explanation

To test the proposed model, a real data was collected from a multidisciplinary university; therefore, the proposed model can be used by other universities. The collected data contains courses, student, marks, and other information, from 2007 to 2019, with 4,699 subjects (courses), 83,993 students, and 3,828,879 records. These datasets describe data distribution with sample information, in addition to the training and testing sample number. Training ratio and total sample number are also considered.

The dataset represents student performance for 16 academic units (faculty/institute/college). The data is divided into two unequal parts. The main part (data collected from 2007 to 2016) is used for training, while the remaining sample part (data collected from 2017 to 2019) is used for testing. Dataset of Economics Education represents the highest dataset percentage with a value of 18%, while the lowest dataset percentage belonged to Physics Education with 0.9% value.

Mark level distribution for training and testing samples for entire dataset is shown in Figures 3 and 4, respectively.

89.7% of mark level of training dataset was equal or greater than medium grade, while the percentage of the testing dataset that is equal or greater than medium grade is 88.6%. The distribution of student's mark level is the same for all academic units in the university. For example, the distribution of the mark level of engineering technology is illustrated, respectively, in Figures 5 and 6.

7. Proposed Method

In order to predict the student performance for the next courses from the performance of previous courses, the research exploits the collected data in the training process for the proposed method.

After collecting real data from multidisciplinary university, a data preprocessing will take place to remove redundant attribute, noise, etc. Then, data will be divided into two sets, according to the date the data was obtained: the first dataset (data obtained from 2007 to 2016) is used for training, while the second dataset (data obtained from 2016 to 2019) is used for testing the proposed method. The testing process is used to evaluate the accuracy of the proposed prediction model.

7.1. Dataset Preprocessing and Transformation. Due to the rich information included by the collected database, preprocessing is essential to tackle some undesired issues such as redundant attribute and noise, which are described by the following steps:

First step: clear redundant attributes such as course name, lecturer name, and student name.

Second step: clear redundant or noise records such as the courses which have been registered by the student but its exam, exemption courses, etc. have never been completed.

Third step: some universities ignore courses when the total number of registered student is less than 15. In our case, these ignored courses are consider as noise and will be neglected.

Fourth step: convert string or text values into numeric values.

The learning model input attributes were selected after making analysis on entire input data; see Table 1, which shows samples of datasets used in the literature. These selected attributes were chosen based on experimental results and some previous models of student performance prediction.

Because there is a different distribution for the various attributes, the proposed prediction model will use quantile transformation (QTF) with min-max scaler (MMS) to generate and convert all values of the range of vales where algorithms of deep learning are converged.

A nonlinear transformation QTF can be considered as one of the strongest preprocessing techniques, due its significant reduction of outlier effects. Values of unseen/new data (such as validation/test data) which are higher or lower than the fitted bounds will be set to the range of output
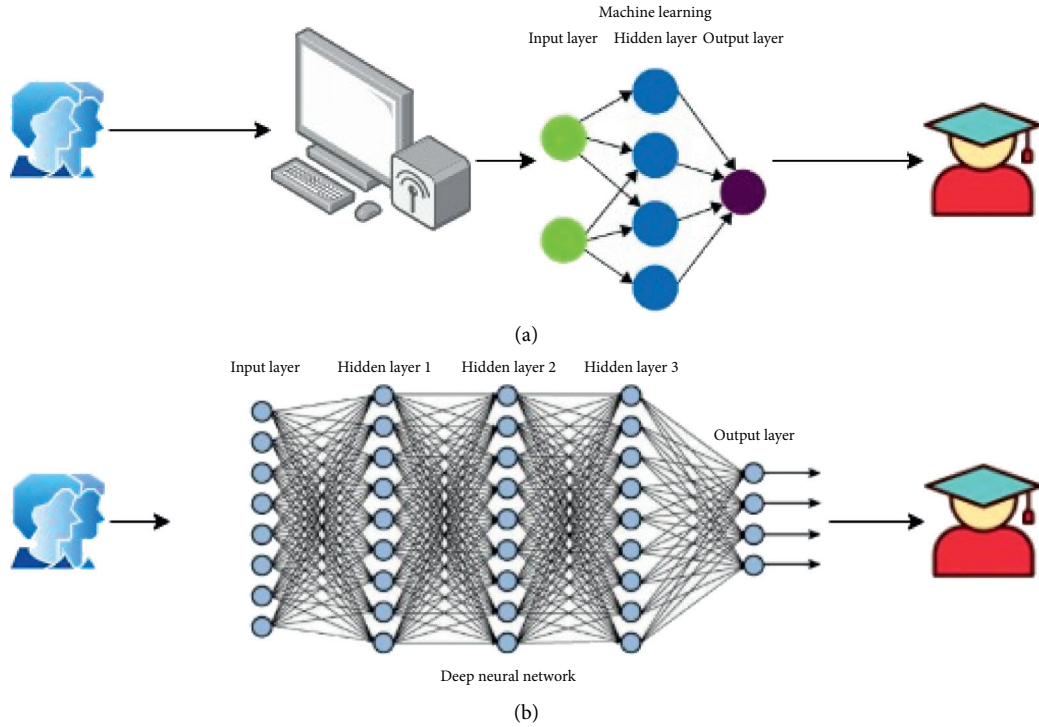
(a)

(b)

FIGURE 2: Difference of simple neural network and deep learning in feature combination techniques.
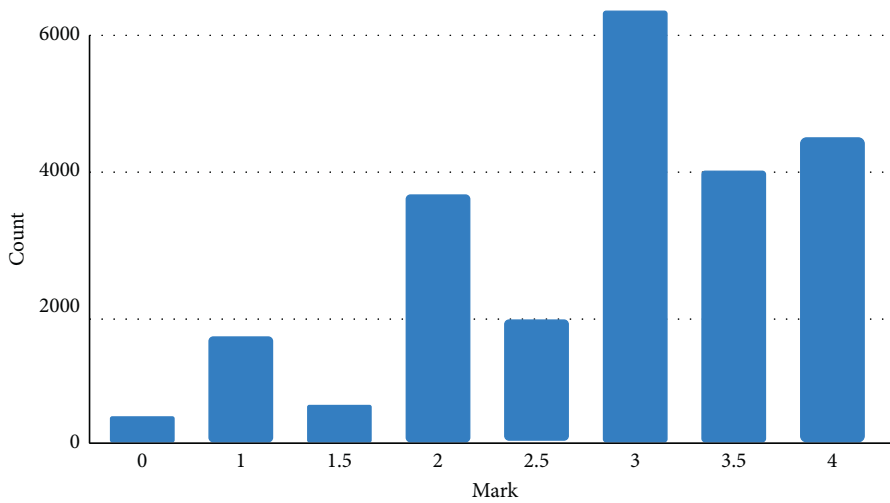

FIGURE 3: Training of full dataset.

distribution. Before data transformation, there is a significant difference between the distribution and range of each feature. The range of QTF data transformation with all of the features will be between 0 and 1, as shown in Figure 6; the result of each feature scaler, in addition to its distribution, will be enabled to be more normal distribution.

In Figure 7, the general framework illustrated contains four main stages. The first stage considers collected data that comes from the standard dataset; the dataset consists more than 70 cases related to 7000 IDs. Then, the preprocessing stage of manipulating data and extracting useful informative data is obtained from them. Weight is derived from features.

These features input to the neural network and then create multiple hidden layers to process the data. Evaluation includes the training mode and the testing mode with evaluation criteria.

MMS will be used to create bins for each image. By using equations (1) and (2), formulas each feature will be scaled into the given range:

$$X_{std} = \frac{X - X_{.min}}{X_{.max} - X_{.min}}, \tag{1}$$

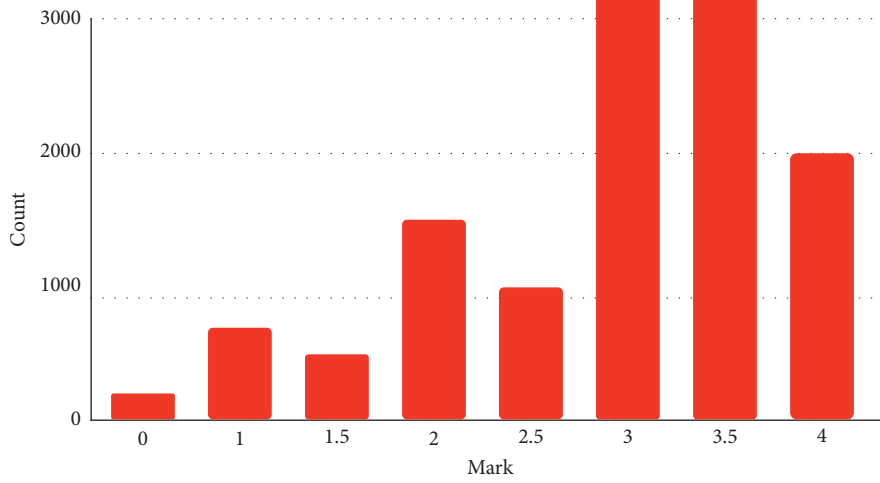$$X_{scaled} = X_{std} * (\max - \min) + \min. \tag{2}$$
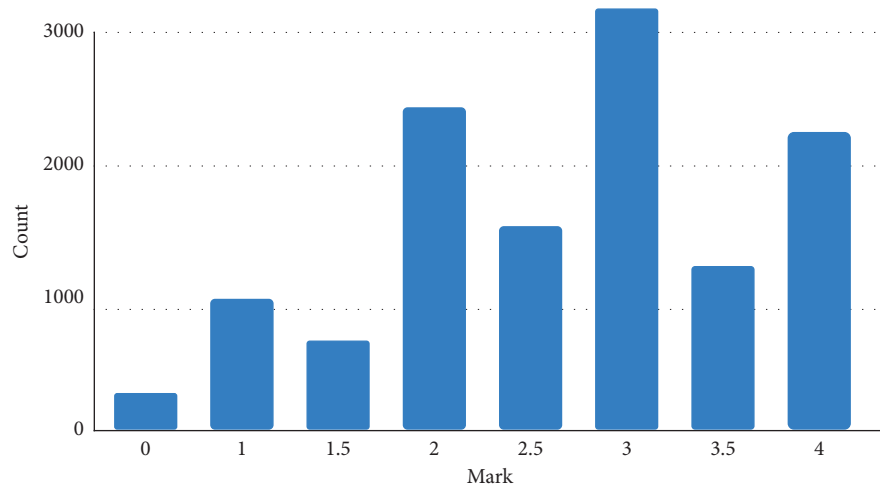
FIGURE 4: Testing of full dataset.

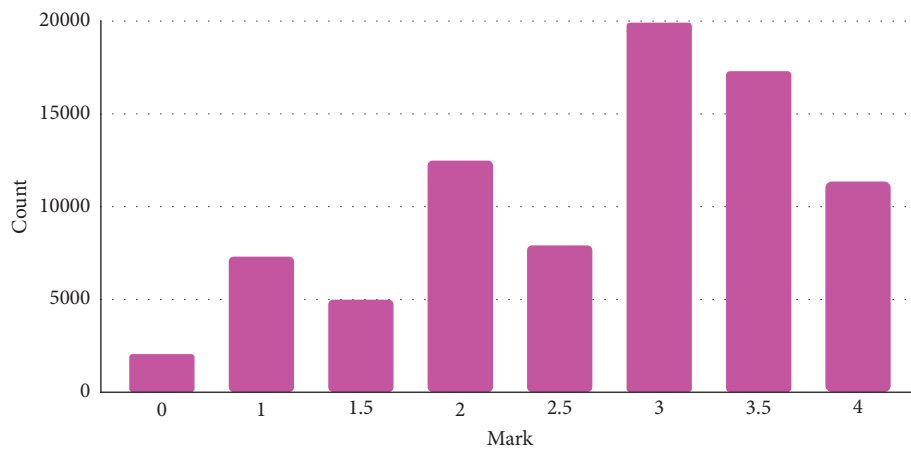FIGURE 5: Mark level distribution on training engineering technology dataset.

FIGURE 6: Mark level distribution on testing engineering technology dataset.

TABLE 1: Literature dataset with their training.

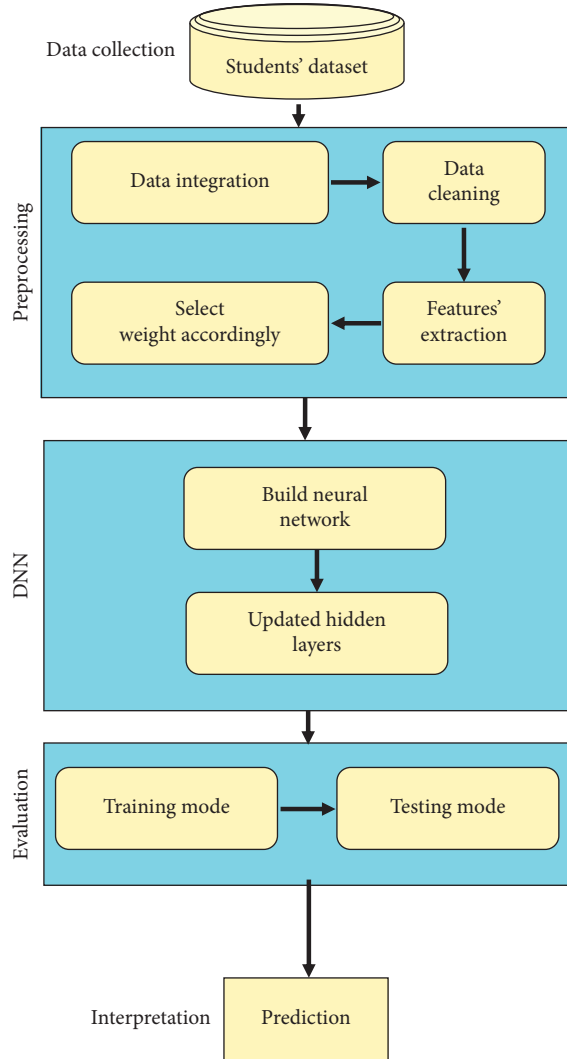| Dataset	Training (%)	Size
Education	75	375,025
Environment and natural rec.	61	208,456
Economics	74	691,256
Foreign languages	67	183,245
Law	73	208,325
Political science	53	642,512
Agriculture	63	476,213

FIGURE 7: General flowchart of the proposed system.

These algorithms show efficient performance in classification field [26]. In this research, the experimental results show promising achievements compared with the original data in task of regression. By learning from the training set, the scaler is applied to the testing set.

7.2. The Proposed Model. Two algorithms which are deep learning and linear regression were used to achieve the student performance prediction model. 1D data vector of 21 features is received by 1D CNN; then, it will be passed

through the stack of one convolutional layer with 64 nodes each with 3 kernels. Then, after each convolution, the rectified linear unit (ReLU) activation function will implemented, as shown in Figure 8.

LSTM includes 64 Tanh unit and one-time step, as shown in Figure 8. Using Sigmoid function illustrated by equation (3) on both 1D CNN and LSTM produces output ranged from 0 to 1. To simulate the student grade which ranged from 0.0 to 4.0 the output of equation (3) will be multiplied by 4.0:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}. \tag{3}$$

8. Experimental Results

8.1. Performance Metrics and Learning Settings. Root mean square propagation (RMSprop) algorithm or Adam algorithm is deployed for all deep learning model networks [6] as an optimizing function with 0.0001 learning rate. A size of batch 16,000 is obtained until 500 epochs. To get overfitting reduction, this research performs stopping with 5 epoch patience. Learning should be stopped if after each consecutive epoch the loss does not reduce. In addition, the learning of scaler algorithm is coming from the training set and also transforms both the training set as well as the test set.

Performance of regression can be evaluated by MAE and RMSE with average of 5 times of running on the test set. The MAE and RMSE can be calculated by using equations (4) and (5), respectively:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |(y_i - \widehat{y}_i)|, \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}, \tag{5}$$

where y_i is the true student grade value (of scale from 0.0 to 4.0) and $\widehat{y}_i$ is the student predicted grade value.

Experimental results can be shown as follows. Section 7.2 will illustrate various scaler results, where QTF is shown as an appropriate technique to preprocess data for tasks of regression; after that, the scaler which has been selected is run with RMSprop and Adam, as two optimizers for comparison, as illustrated in Section 7.3. The research is performed by using both liner regression and deep learning with the optimizer function and best scaler on all of 16 datasets of different departments, and the prediction is performed on the merged 16 datasets as one dataset.

8.2. Deep Learning Performance Enhanced by Scalers. Results are obtained by various scalers; it is clear that deep learning performance can be improved by scalers. Among the used scalers, QTF is the best, and it reveals that the highest performance on fifteen out of sixteen datasets for 1D

Operation	Data dimensions			Weights (n)	Weights (%)
Input	####	21	1		
conv1D	\ | /	----------------------		256	17.4%
relu	####	19	64		
flatten	| | | | |	---------------------		0	0.0%
	####	1216			
Dense	XXXX	---------------------		1217	82.6%
Sigmoid	####	1			

FIGURE 8: Architecture of the proposed method.

TABLE 2: Prediction performance of optimal function.

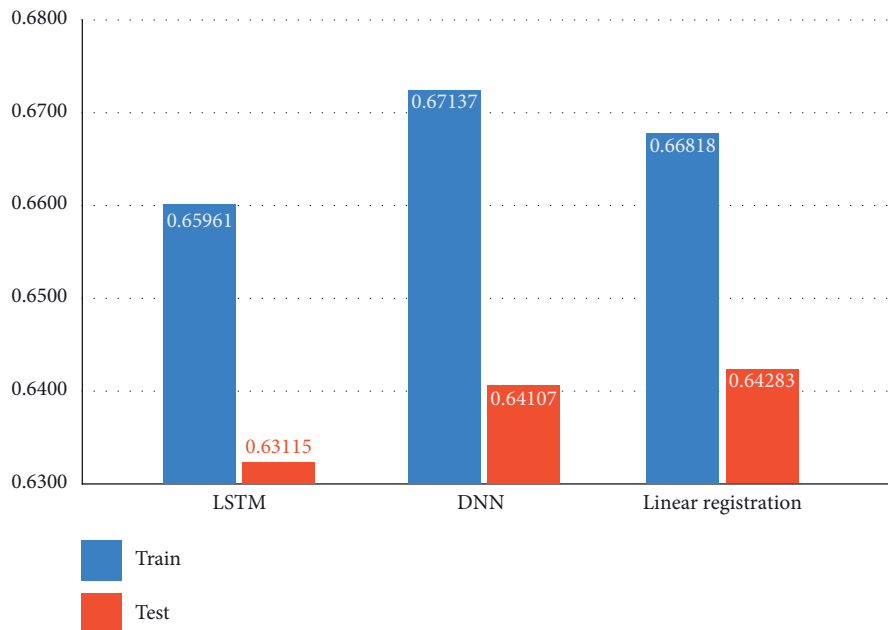| Dataset	MAE	RMSE
Education	0.562	0.782
Environment and natural rec.	0.593	0.765
Economic	0.542	0.711
Foreign languages	0.536	0.732
Law	0.583	0.743
Political science	0.534	0.765
Agriculture	0.591	0.785

FIGURE 9: Test and train modes of different methods at the same dataset.

CNN and for all LSTM datasets is a clear improvement of one layer (1D) of CNN is necessary in this issue.

8.3. Optimizer of RMS Prop with the Deep Learning Regression.
Adam optimizer function and RMS prop comparison of results are shown in Table 2. By using RMS, the achieved improved performance is fourteen out of sixteen of the used datasets. 3.3% was the estimation of average improvement in all the used datasets.

Within testing mode and training mode, at the proposed algorithm, the percentage will differ, but, for other methods, it should be in the same condition for better performance have to compare achieved results with the same existing dataset, so Figure 9 illustrates the performance of the proposed method with a standard of one at the same platform.

With the proposed system, we have two modes firstly training mode that learn the system in advance with labeled data and running the system with standard datasets mentioned in Table 2 due to have known results in advance then secondly testing mode on the required dataset in real testing.

9. Conclusion

In this study, the deep neural network in higher education was proposed and the identification and prediction of higher education students and their scientific behavior by comparing their levels and grades were obtained. Several steps have been proposed for the deep neural network algorithm, including data initialization and preprocessing. The process of building hidden layers in a neural network requires extracting useful features and weights for each one. For increase in prediction accuracy, we used two models such as Adams and RMS prop that help the system perform better . The proposed method proves its worth from the achieved results and can used in practical. Through these results, helping educational institutions in terms of staff and students is easy, predicting future data reduces education difficulties and helps to develop future plans for education policy.

In the future, update features that are extracted may be needed and their weight is chosen carefully; by updating hidden layers in neural network, the system can be made more reliable.

Data Availability

All data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] W. Zhu, L. Xie, J. Han, and X. Guo, "The application of deep learning in cancer prognosis prediction," *Cancers*, vol. 12, no. 3, p. 603, 2020.

[2] B. Bakhshinategh, O. R. Zaiane, S. El-Atia, and D. Ipperciel, "Educational data mining applications and tasks: a survey of the last 10 years," *Education and Information Technologies*, vol. 23, no. 1, pp. 537–553, 2018.

[3] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "KnowEdu: a system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31553–31563, 2018.

[4] Q. E. Aissaoui, Y. E. A. E. Madani, L. Oughdir, A. Dakkak, and Y. E. Allioui, "A multiple linear regression-based approach to predict student performance," in *Proceedings of the International Conference on Advanced Intelligent Systems for Sustainable Development*, Springer, Cham, Switzerland, January 2020.

[5] R. Cuevas, N. Ntoumanis, J. G. Fernandez-Bustos, and K. Bartholomew, "Does teacher evaluation based on student performance predict motivation, well-being, and ill-being?" *Journal of School Psychology*, vol. 68, pp. 154–162, 2018.

[6] M. Xu, "Image processing system based on FPGA and convolutional neural network," *Microprocessors and Microsystems*, vol. 2020, Article ID 103379, 2020.

[7] J. Zong, C. Cui, Y. Ma, L. Yao, M. Chen, and Y. Yin, "Behavior-driven student performance prediction with tribranch convolutional neural network," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, October 2020.

[8] Y. He, R. Chen, X. Li et al., "Online at-risk student identification using RNN-GRU joint neural networks," *Information*, vol. 11, no. 10, p. 474, 2020.

[9] H. Liu, Y. Zhu, T. Zang, Y. Xu, J. Yu, and F. Tang, "Jointly modeling heterogeneous student behaviors and interactions among multiple prediction tasks," 2021, https://arxiv.org/abs/2103.13565.

[10] T. Sämann, P. Schlicht, and F. Hüger, "Strategy to increase the safety of a DNN-based perception for HAD systems," 2020, https://arxiv.org/abs/2002.08935.

[11] Y. Lin, H. Zhao, Y. Tu, S. Mao, Z. Dou et al., "Threats of adversarial attacks in DNN-based modulation recognition," in *Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, IEEE, Toronto, Canada, July 2020.

[12] T. Yamashita, "Grit and second language acquisition: can passion and perseverance predict performance in Japanese language learning?" Master Thesis, University of Massachusetts Amherst, Amherst, MA, USA, 2018.

[13] T. J. Dunn and M. Kennedy, "Technology enhanced learning in higher education; motivations, engagement and academic achievement," *Computers & Education*, vol. 137, pp. 104–113, 2019.

[14] N. Hug, "Surprise: a python library for recommender systems," *Journal of Open Source Software*, vol. 5, no. 52, p. 2174, 2020.

[15] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Computers & Education*, vol. 131, pp. 22–32, 2019.

[16] Z. Iqbal, J. Qadir, A. Mian, and F. Kamiran, "Machine learning based student grade prediction: a case study," 2017, https://arxiv.org/abs/1708.08744.

[17] L. Zhang, T. Luo, F. Zhang, and Y. Wu, "A recommendation model based on deep neural network," *IEEE Access*, vol. 6, pp. 9454–9463, 2018.

[18] M. Fu, H. Qu, Z. Yi, L. Lu, and Y. Liu, "A novel deep learning-based collaborative filtering model for recommendation system," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1084–1096, 2019.

[19] J. Rastrollo-Guerrero, J. A. Gomez-Pulido, and A. Domínguez, "Analyzing and predicting students' performance by means of machine learning: a review," *Applied Sciences*, vol. 10, p. 1042, 2020.

[20] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student performance prediction and classification using machine learning algorithms," in *Proceedings of the 2019 8th International Conference on Educational and Information Technology*, pp. 7–11, March 2019.

[21] D. Prado and M. Lopez, *Advances in Financial Machine Learning*, John Wiley & Sons, Hoboken, NJ, USA, 2018.

[22] G. Sulong and A. Mohammedali, "Human activities recognition via features extraction from skeleton," *Journal of Theoretical & Applied Information Technology*, vol. 68, p. 3, 2014.

[23] E. Hassan, Y. Khalil, and I. Ahmad, "Learning feature fusion in deep learning-based object detector," *Journal of Engineering*, vol. 2020, Article ID 7286187, 11 pages, 2020.

[24] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Systems with Applications*, vol. 129, pp. 273–285, 2019.

[25] G. Sulong and A. Mohammedali, "Recognition of human activities from still image using novel classifier," *Journal of Theoretical & Applied Information Technology*, vol. 71, no. 1, pp. 115–121, 2015.

[26] B. T. Atiyha, S. Aljabbar, A. Ali, and A. Jaber, "An improved cost estimation for unit commitment using back propagation algorithm," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 15, no. 2, pp. 243–248, 2019.