

Research Article

Modelling, prediction and classification of student academic performance using artificial neural networks

E. T. Lau¹ · L. Sun^{1,2} · Q. Yang¹© The Author(s) 2019 **OPEN**

Abstract

The conventional statistical evaluations are limited in providing good predictions of the university educational quality. This paper presents an approach with both conventional statistical analysis and neural network modelling/prediction of students' performance. Conventional statistical evaluations are used to identify the factors that likely affect the students' performance. The neural network is modelled with 11 input variables, two layers of hidden neurons, and one output layer. Levenberg–Marquardt algorithm is employed as the backpropagation training rule. The performance of neural network model is evaluated through the error performance, regression, error histogram, confusion matrix and area under the receiver operating characteristics curve. Overall, the neural network model has achieved a good prediction accuracy of 84.8%, along with limitations.

Keywords Academic performance · Statistical analysis · Artificial neural network · Machine learning

1 Introduction

The academic achievement of university students is the most important benchmark to compare the quality of university students. It serves as the basic criteria for university to monitor the quality of teaching and learning and for university to evaluate and select students. Nowadays, most universities are facing tough challenges to attract prospective students due to increasingly highly competitive educational markets. Therefore, the study of students' academic achievements is of great significance in promoting the student developments and the improvement of higher education quality. However, the student performance is influenced by many factors in a complicated manner, and the student socio-economic background and their historical academic performance may potentially affect their academic performance. Unsurprisingly, most existing research works have been limited to analysing and predicting students' performance in a relatively simple problem formulation using statistical techniques.

To cope with such limitation, machine learning has been increasingly used in the data science applications to analyse complex relationships. It is capable of learning automatically without being programmed explicitly. An Artificial Neural Network (ANN) model, even though has long established history in computing and data science, is gaining growing attention and wide applications. ANN extends the capability of analysing complicated amount of data sets that are not easily to be simplified through the conventional statistical techniques. It has also the ability to implicitly detect non-linear relationships between dependent and independent variables [1]. ANN has been gaining wider attention and has proven a great success in the application of pattern recognitions, classifications, forecasting and prediction in the areas of healthcare, climate and weather, stock markets, etc.

However, the use of ANN is limited in the educational research area. This can be due to the arising complexity of modelled network, the difficulty for a modelled ANN system to provide a suitable explanation (the black-box

✉ E. T. Lau, engtseng.lau@brunel.ac.uk; L. Sun, sunlizhi@qfnu.edu.cn; Q. Yang, qingping.yang@brunel.ac.uk | ¹Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK. ²Qufu Normal University, Qufu 273165, Shandong, People's Republic of China.



nature), proneness to over-fitting, and the time needed for neural network training [1]. In order to mitigate the shortcomings through the application of ANN, a domain of competence is compromised for both conventional statistical and ANN analysis in this paper. The educational data will be studied initially with conventional statistical analysis, and the confirmation of statistical outputs will be used to perform ANN training, validation and testing in order to develop the ANN model with suitable configuration settings to accurately predict and classify the students' performance.

Overall, this paper presents an exploratory modelling and analyses of students' performance through the data collected from one Chinese university. The ANN model serves as a dominant educational quality tool that evaluates the students' performance throughout universities, addressing those disparities and thus continually improving the education quality.

The organisation of the paper is as follows: Sect. 2 presents the reviews of ANN and its summary, Sect. 3 presents the methodology of statistical testing, ANN modelling and verification, Sect. 4 presents results of statistical evaluations, ANN configuration settings and the performance, Sect. 5 concludes the findings.

2 Artificial neural network

Artificial Neural network (ANN) is a powerful and complex modelling tool for modelling nonlinear functions that often describes the real world systems [2–4]. ANN is formed through a collection of artificial neurons that resemble the connection geometry of neurons in human brains in order to execute a task with improved performance through 'learning, training and continuous improvement [2, 5, 6].

ANNs are formed with three layer neuron structures, namely the input, hidden (middle) and output layers. The input layer gathers numerical information data with feature sets and activation values. Input values are propagated through the interconnected neurons to the hidden layer. In the hidden layer, the input neurons are summed in order to compute weighted sum of the input neurons; and summed neurons are further combined to produce results in the output layer using an activation (or transfer) function [5, 7]. Both neurons and connection contain adjustable weights during the learning process. The summed neurons will transform mathematically in the output layer if the activation function threshold is exceeded.

A number of times the training functions are used to update the connection weights in the process of feeding the input values and terminating with output values in ANN is called an Epoch [7]. This is where the inputs

of artificial neurons are multiplied by weights, and the resultant of these summation are fed to the output layer through an activation function [6]. The frequently used of activation functions include linear, sigmoid and hyperbolic tangent functions. The training terminates when the maximum epoch value and/or the validation checks are reached. The resultant trained data is fed into the test data in order to examine the ANN's performance.

The most common learning rule of ANNs is back-propagation (BP), which is a supervised learning approach and can be used for training the deep neural networks. BP adjusts the weights of neurons through the calculated errors and enables the network to learn from the training process. Typical problem solving of ANNs include three archetypes of learning, i.e. supervised learning, unsupervised learning and reinforcement learning [6].

Remarkably, ANN approach has been receiving wide attention for educational research purposes. [2] applied ANN to model and perform data training based on students' course selection behaviour, and further identified the best strategy and configuration to meet students' demands for every courses for optimal course scheduling within a university. ANN was used in classifying students within the musical faculty to predict the perceptions of students' in music education [4]. ANN was combined with particle swarm optimisation to assess performance of lecturers and also to enhance the accuracy of the recognition in the university's lecturer assessment system [7]. An artificial intelligence (AI) procedure based on self-organising neural network model was proposed by [8] that automatically characterised bibliometric profiles of academic researchers and further identify institutions that had similar pattern of academic performance among researchers. Additionally, different prediction models other than ANN technique such as discriminant analysis, random forests, and decision trees were applied by [5] in predicting and classifying the academic performance of students in three different universities.

ANN was also used to model and simulate diversity of learning style among students through two learning paradigms namely the supervised (learning with teacher) and unsupervised learning (learning without teacher and through students' self-study) [9]. Similarly, [10] proposed an Artificial Intelligence-based tool that took into account ANN as one of the Artificial Intelligence methods in minimising disorientation of learning behaviour and overloaded cognitive problems among students. In the tool, the ANN performed better compared with other learning models applied. Additionally, a BP-based ANN was applied in evaluating the quality of the teaching system and the ANN performance was able to meet the requirements of the system's feasibility and precisions [3]. Although the models accurately classified the students throughout

the prediction process with different levels of academic grades, low prediction accuracies were obtained as the result. The paper by [11] successfully applied ANN to predict the student's mood during self-assessment online test, with prediction accuracy of over 80%. A similar approach was adopted by [12] but focused in predicting academic success in mathematic courses using BP-based ANN at three different universities, with prediction accuracy of 93.02%.

Overall, ANN has the ability in performing the neural fitting and prediction, and the ability to classify any data with arbitrary accuracy theoretically. However, there are only limited studies and tools to predict the academic performance of students, especially in smart education context. This paper has therefore proposed the use of ANN as an application modelling tool for predicting the academic performance among students. Unlike the modelling concepts as outlined in the literature, the paper focuses on the students' socio-economic background and their entrance examination results that likely would affect their overall academic performance in a university. This is important not only to examine the significant factors that affect the performance among students, but also to classify students' performance correctly based on the predicted pattern obtained from ANN. As the prediction performance may be greatly reduced by huge discrepancies of samples from different universities [5], this paper would only focus on data samples from one single university in order to ensure the prediction accuracy of academic performance. The modelled ANN model would serve as a framework and tool to predict the future students' academic performance, and to further address those issues that hinder the success of student learning and thus continually improving the educational quality.

3 Methodology

This section presents the methodology of acquiring educational data, statistical hypothesis testing for neural network modelling, and the methodology of evaluating the performance of modelled ANN.

3.1 Educational data collection

The sample data were collected about a total of 1,000 students consisting of 275 female and 810 male students selected from the undergraduate students from University Q. The examination results comprised three undergraduate programs within three departments of the university for the student intakes from year 2011 to 2013, with four year in total courses for each year of intake. In addition, the sample data of students' socio-economic background

and national university entrance examination results were also collected. The entrance examination results included five core subjects: Chinese, English, Math, Comprehensive Science and Proficiency Test. The student performance was determined using the standardised Cumulative Grade Point Average (CGPA) for the entire four year duration of their studies. It is calculated as the weighted average of the grade point gained:

$$CGPA = \frac{\sum c_i g_i}{c_i}, \quad (1)$$

where c_i denotes the credit hours per course i and g_i is the grade point received per course i .

3.2 Statistical hypothesis testing

Prior to the ANN modelling, it is necessary to examine the relationship between the socio-economic, family and education background of students and their academic performance. Statistical evaluations are performed initially in this study, including correlations, two-sample t -test and ANOVA (single factor). Pearson correlation coefficients are calculated to measure the linear the relationships of five core entrance examination subjects and the resultant CGPA. Those correlated variables are used further as input neurons for the ANN modelling. For t -test and ANOVA, a significance level of $p = 0.05$ is chosen to test the significant difference, notably the mean CGPA difference in two or more sample groups. Hypothesis testing in this case is performed to test whether male and female students, geographical locations, roles of parents, types of students have significant influence on the resultant CGPA.

3.3 Neural network modelling

ANN in this case is used for neural prediction of students' CGPA and the data classification of data through input observations. Both schemes are performed based on supervised machine learning. The ANN in this paper is modelled following the earlier research works [5, 7, 13]. It is not the scope of this paper to compare the effectiveness of ANN performance with other machine learning techniques.

Typically, the ANN model can simply be expressed as a mathematical function:

$$\tilde{Y} = f(\tilde{X}, \tilde{W}), \quad (2)$$

where \tilde{Y} and \tilde{X} are the output and input vectors. \tilde{W} is a vector of weight parameters representing the connections within the ANN.

The input layer gathers data with feature sets and the input values are fed to the hidden layer. The output values

of j th neuron y_j of vector \tilde{Y} are computed by means of the weighted sum of input elements x and w :

$$y_j = \theta \left(\sum_{i=1}^{N_i} w_{ij} x_i \right). \tag{3}$$

The θ is the activation function (transfer function), N_i is the total number of i th connection lines to the j th neuron and x_i is the output value from the previous layer of i th neuron. The activation function (θ) of hyperbolic tangent is used to transfer the value of weighted sum of inputs to the output layer. The resultant activated node for the next input layer is therefore:

$$x_j = \theta(y_j). \tag{4}$$

In order to reduce the dimensionality of the predictor space and to prevent the possibility of over-fitting, principal component analysis (PCA) is employed [14]. PCA is a data reduction technique that transforms predictors linearly, removes any redundant dimensions, and generates new sets of variables called as principal components [14].

The BP-based supervised learning approach is applied where both inputs and outputs parameters are supplied to the ANN model. BP is used as the learning rule for the ANN model that adjusts the weights of neurons w_{ij} through the errors computed that further produces desired outputs. The error function (E) of computed BP-based ANN is calculated as the sum of square difference difference between the and the target values and the desired outputs:

$$E = \frac{1}{2} \sum_j^{N_j} (y_j - t_j)^2, \tag{5}$$

where t_j is the target value for neuron i in the output layer and N_j is the total number of output neurons.

In ANN weights are updated recursively. The BP-based Levenberg–Marquardt optimisation algorithm is applied in the ANN training. The Levenberg–Marquardt is a hybrid-based training method using the steepest descent (gradient descent) and Gauss-Newton method. It speeds up the convergence to an optimal solution and therefore is effective in solving non-linear problems over the other training algorithms [15, 16]. The algorithm introduces another approach of approximation to Hessian Matrix, which is similar to the Gauss-Newton method [15, 17]:

$$w_{ij+1} = w_{ij} - [J'J + \zeta I]^{-1} J' e_k, \tag{6}$$

where J denotes the Jacobian matrix, e_k is the error in the network [17], w_{ij} is the current weight and w_{ij} is the updated weight, ζ is the damping factor.

When ζ is small, the Levenberg–Marquardt training algorithm in Eq. (6) is based on Gauss-Newton method,

and in contrast, becomes gradient descent algorithm when ζ is large. Therefore, ζ is adjusted at every iteration in order to guide the optimisation process and switched between those two algorithms.

The number of neuron in the output layer is the resultant decision prediction of the problem [4]. The output layer consists collections of vector \tilde{Y} , which is the collection of predicted CGPA.

3.4 ANN performance evaluation criteria

For the later data analysis, in order to evaluate the performance of ANN, this paper introduces several new perspectives that mitigates the arising of over-fitting issues to ensure the appropriateness of ANN performance. The evaluation includes the computation of Mean Square Error (MSE), regression analysis, error histogram and confusion matrix. A well-trained ANN model should have low MSE value (close to zero), which means that the predicted outputs converge closely to the target outputs (t_{ij}). MSE is calculated as:

$$MSE = \frac{1}{N_i N_j} \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} (y_{ij} - t_{ij})^2. \tag{7}$$

As mentioned, an over-fitting of the trained network is possible if obtained MSE value is low. This further indicates that ANN only works well in the training stage, but not in validation and testing phase. To mitigate this a regression is performed along with computed R-value which demonstrates the goodness of fitting between the predicted and the desired outputs [18]. The plot is useful in examining the fitting performance. If poor fitting (low R-value) is obtained, further trainings are required with modification of hidden layers and neurons.

As mentioned, another way of measuring ANN performance is to tabulate the error histogram. The error histogram demonstrates how the errors are distributed with most errors are occurred near zero. The error is simply the difference between the targeted outputs t_{ij} and the predicted outputs y_{ij} .

In order to verify ANN's performance in terms of classifications, a confusion matrix (also known as error matrix) is used. A confusion matrix of a binary classification, is a two-by-two table showing values of True Negatives (TN), False Negatives (FN), True Positives (TP) and False Positives (FP) resulting from predicted classes of data [19, 20]. The confusion matrix allows the measures of rates such as prediction accuracy, error rate, sensitivity, specificity and precision [19], which are included in this paper (Refer [21] for calculations of the rates derived from confusion matrix). Furthermore, a Receiver Operating Characteristic

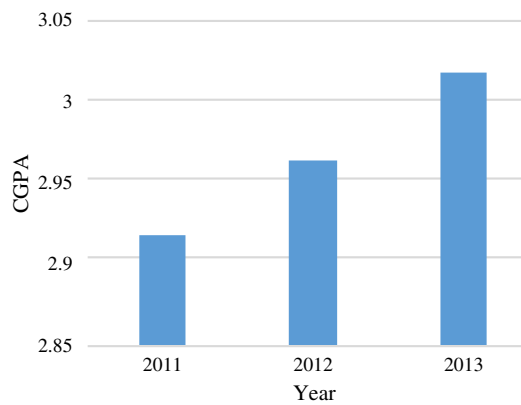


Fig. 1 Yearly average students' CGPA performance for the intake year 2011–2013

Curve (ROC) is also employed that detects the trade-off between the TP rate and FP rate, and also the area under the ROC curve (AUC) [20]. ROC is widely used in machine learning in verifying and evaluating performance of classifications [19].

4 Results

This section presents the statistical findings, the ANN configuration and the performance based on the obtained yearly academic performance educational data of 1,000 students (175 female and 810 male students) for the intake year 2011–2013 from the University Q.

4.1 Statistical evaluations

Figure 1 presents the performance of yearly students' average CGPA for the intake year of 2011–2013. Additionally, the ANOVA for three consecutive academic performance years with $p < 0.05$ indicates that there is significant difference of students' yearly CGPA performance, with notable improvements in their performance for next two academic years.

Two-sample t -tests are performed in three types of sample data to test the null hypotheses concerning the students' gender, location, and repeating students.

Table 1 shows the t -test result of CGPA performance based on students' gender. Since $p < 0.05$, there is a significant differences of CGPA performance for male and female students, where in this case female students score better than male students. Table 2 presents another t -test of CGPA performance of students based on urban and rural settlements. The t -test results with $p > 0.05$ concludes that there is no significant difference of CGPA performance based on different settlement types. Finally, the

Table 1 T -test result of CGPA based on students' gender

Gender	Male	Female
Frequency	249	751
Mean CGPA	2.54	3.11
Variance	0.22	0.18
t -cal	– 17.33	
t -crit	1.65	
df	391	
p (one-tailed)	0.01	
p (two-tailed)	0.01	
Decision	Reject	

Table 2 T -test result of CGPA based on students' settlement type

Settlement-based	Urban	Rural
Frequency	245	755
Mean CGPA	2.93	2.98
Variance	0.28	0.24
t -cal	– 1.32	
t -crit	1.64	
df	390	
p (one-tailed)	0.09	
p (two-tailed)	0.18	
Decision	Accept	

Table 3 T -test result of CGPA based repeating students

Repeating students	No	Yes
Frequency	711	289
Mean CGPA	3.04	2.82
Variance	0.23	0.25
t -cal	6.04	
t -crit	1.64	
df	513	
p	0.01	
p (two-tailed)	0.01	
Decision	Reject	

t -test result with $p < 0.05$ in Table 3 indicates that mean CGPA performance of current and repeating students is statistically significant, with current students score better than repeating students.

Correlation test is performed to examine the degree of closest relationship among the entrance subject examination results and the CGPA result. The correlation coefficients between five entrance examination subjects and the CGPA performance are shown in Table 4, with descending order: English, Maths, Chinese, Comprehensive Science and Proficiency test.

ANOVA tests are performed to examine the effects of parent’s occupation on the resultant students’ CGPA. The parents’ occupation is classified into five categories and the effects of mother and father’s occupation in affecting the students’ performance are separately evaluated, with corresponding results as shown in Tables 5 and 6. The $p < 0.05$ in Table 5 indicates that the students’ CGPA are significantly affected by mothers’ occupation. It can be inferred that one of students’ great achievements are based on motivations from their mother. Additionally, it can be explained by that mothers are typically more concerned about academic performance among their children, hence influence a lot on students’ performance. In contrast, there is no significant difference ($p > 0.05$) of students’ mean CGPA based on fathers’ job occupation as shown in Table 6, where fathers’ motivation are not sufficiently strong enough to improve academic performance among students. It is not the scope in this paper to further evaluate the mean difference of individual job units within parents and the resultant CGPA.

4.2 ANN configuration settings

The ANN modelling and evaluations are performed using MathWorks MATLAB software. The input layer consists of 11 variables about students’ background and their entrance exam results, with the background information

including gender, location, whether or not repeating students, previous school area, parent’s occupation, and the entrance exam results including Chinese, English, Maths, Comprehensive Science and Proficiency test. The modelled ANN has two hidden layers, where each hidden layer consists of 30 neurons as such configuration provides the best outcome throughout several simulations with different hidden layer and neuron settings. Each hidden layer with 30 neurons is fed into a single output neuron that carries the decision of the variable, which is the prediction of students’ CGPA.

The activation function of hyperbolic tangent is used. Decisions must be taken to divide the dataset into training, validation and test ratio. Data samples of 1,000 students are randomly mixed and 0.7 of the mixed samples are used for training, 0.15 used for validation and the remaining 0.15 used for the testing. By following [14], PCA is kept only enough components to explain 95% of the variance.

During the training and learning phase, Levenberg–Marquardt algorithm in Eq. (6) is used to determine the optimal weights that are fed to the next input layer. The damping factor ζ is set to 0.001. The training epoch is set to 1500. The ANN training performs continuously and terminates when the validation error failed to decrease for six iterations during the validation process. Typically, the validation protects over-training of ANN.

Table 4 Correlation coefficient of five subject areas and CGPA

	Chinese	Maths	English	Comprehensive science	Proficiency test	CGPA
Chinese	1					
Maths	0.10	1				
English	0.30	0.21	1			
Comprehensive Science	0.33	0.45	0.54	1		
Proficiency test	0.25	0.22	0.46	0.32	1	
CGPA	0.49	0.58	0.60	0.39	0.32	1

Table 5 One-way ANOVA against mothers’ occupations

Source	Sum of square	df	Mean square	F-cal	p	F-crit
Between groups	2.93	5	0.59	2.39	0.04	2.22
Within groups	243.89	995	0.25			
Total	246.82	1000				

Table 6 One-way ANOVA against fathers’ occupations

Source	Sum of square	df	Mean square	F-cal	p	F-crit
Between groups	1.38	5	0.28	1.12	0.35	2.22
Within groups	245.58	995	0.25			
Total	246.96	1000				

4.3 ANN simulation results

After the ANN training, PCA is able to keep five components, and the explained variance per component (in order) is: 62.9%, 16.6%, 8.8%, 5.1%, 2.8%, 2.2%, 0.9%, 0.6%, 0.1%, 0% (where variances of least important components are hidden from display). The ANN performance with $MSE \approx 0.27$ or 6.9% ($< 10\%$) indicates the sufficiently good performance of ANN in completing training training and validation runs.

Regression plots of ANN performance are shown in Fig. 2. For a good data fitting, the data should have the predicted outputs y_{ij} lying closely with target outputs t_{ij} . In

this case, the resultant regression plots achieve marginally good fits with an overall R -value of 0.64.

The error histogram evaluates the error distributions based on the resultant ANN predictions. The distribution of errors are shown in Fig. 3. Most errors occurred near zeroth point (horizontal axis), and errors are gradually decreasing when moving away from zeroth point. This proves that ANN performs the prediction successfully with acceptable error distributions as shown in Fig. 3.

Figure 4 presents the results of the confusion matrix for a binary classifier resulting from two classes (students' gender). As can be seen, 136 out of 249 male students and 712 out of 751 female students have been classified correctly.

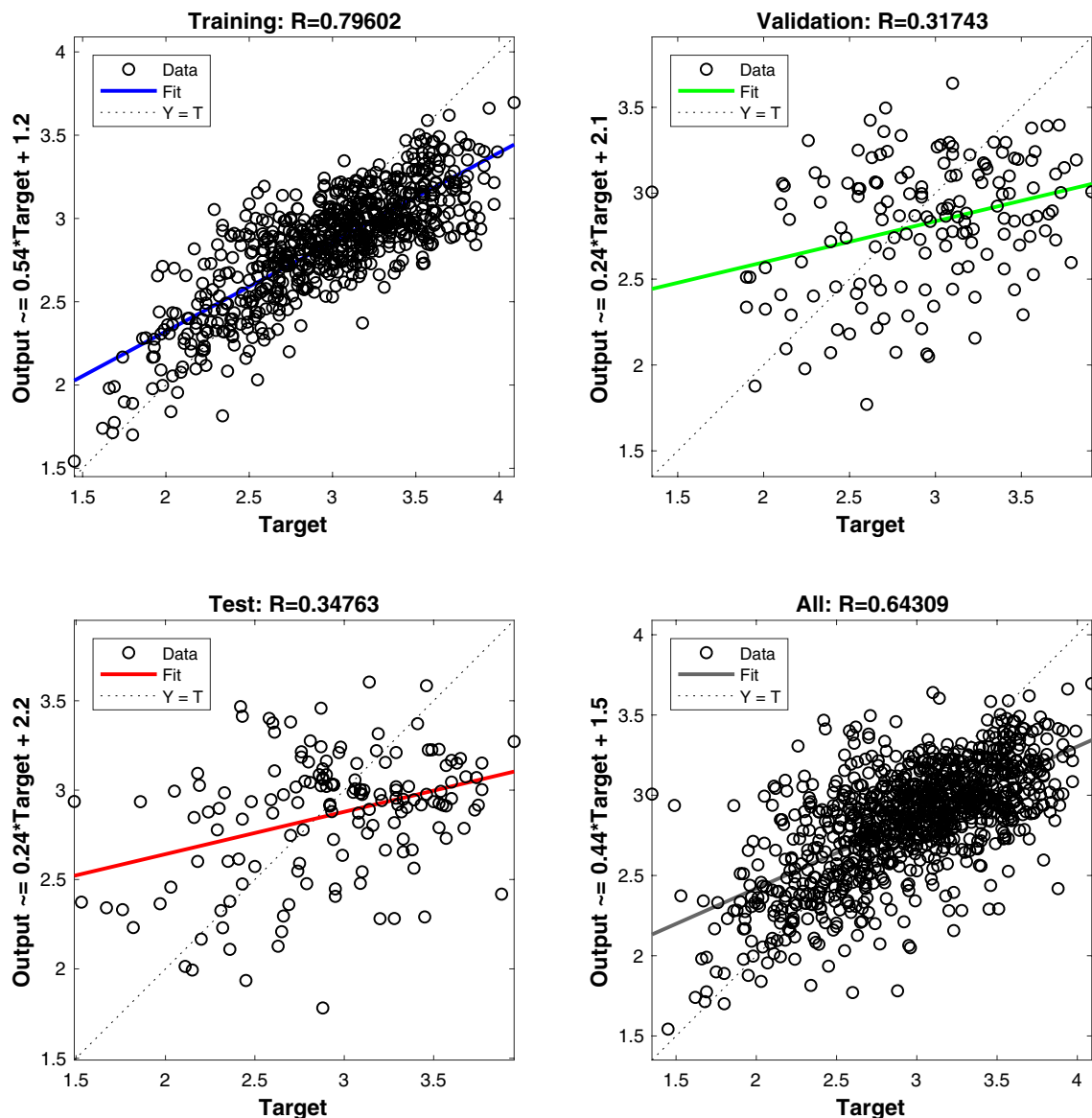
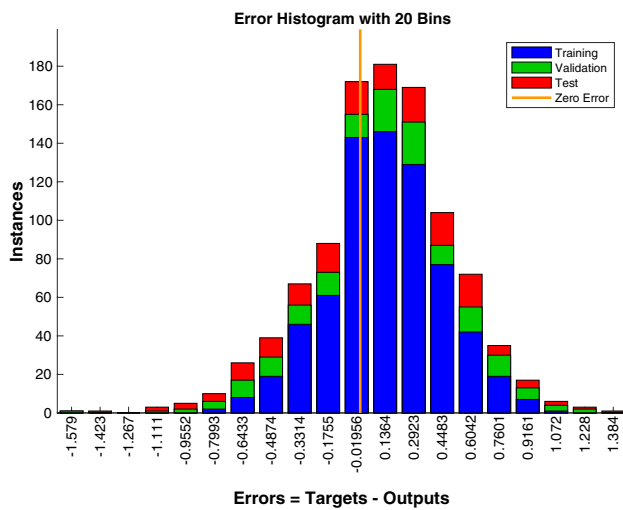


Fig. 2 Regression plot for ANN model: top left: training data; bottom left: test data; Top right: validation data; bottom right; resultant regressions



Accuracy = 84.8%
 Error = 15.2%
 Sensitivity = 94.8%
 Specificity = 54.6%
 Precision = 86.3%

The ROC curve of ANN is shown in Fig. 6. The area under the ROC curve (AUC) is commonly used to evaluate the effectiveness of ANN’s accuracy in prediction and classification, where an AUC of 1 represents a perfect test. Based on the graph, the AUC value of 0.86 is achieved, with TP and FP tradeoff rate as 0.05 and 0.55. This shows the sufficient success of the ANN study of in predicting and classifying students’ performance.

Fig. 3 Error histogram plot

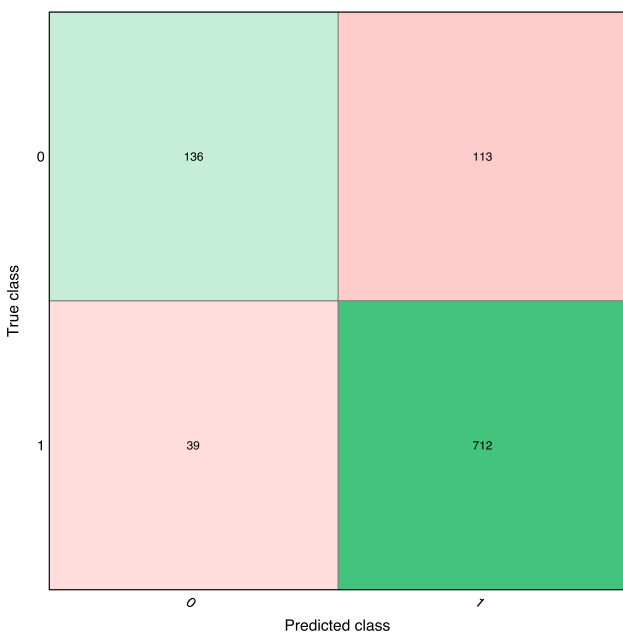


Fig. 4 Binary classifier of confusion matrix based on classification results by ANN

Similarly, Fig. 5 shows the percentage distributions of the confusion matrix from Fig. 4. The True Positive (TP) rates are computed as 55% (male) and 95% (female) and in turn, the False Negative (FN) rates are computed as 45% (male) and 5% (female). Such high FN rates for classification of male students may be due to the high imbalanced ratio of students’ gender (249 male versus 751 female students).

The use of confusion matrix allows the determination of rates such as accuracy, error rate, sensitivity, specificity and precision, and the results are as follows (in percentage):

5 Conclusion

In this paper ANN is used to evaluate and predict the students’ CGPA using the data about their socio-economic background and entrance examination results of the undergraduate students from a Chinese university. The initial stage involves statistical analyses of the students’ background information have indicated significant improvement in the case of yearly students’ CGPA ; female students scoring significantly better than male students; there is no significant difference of students’ performance from rural and urban areas; and non-repeating students performing significantly better than repeating students. The correlation analysis indicates that the English exam result in the entrance examination has the strongest correlation with the student’s CGPA. Students’ mother play a significantly more important role in the students’ academic performance than father.

In order to evaluate the performance of ANN, computations of Mean Square Error (MSE), regression analysis, error histogram and confusion matrix are introduced to ensure the appropriateness of ANN’s performance in mitigating the arising of overfitting issues. Overall, the ANN has achieved a good prediction accuracy of 84.8%, and with good AUC value of 0.86. The good ANN performance makes ANN such a useful technique for education evaluation purposes. However, ANN performs poorly in classifications of students according to their gender, as high False Negative (FN) rates are obtained as results. This is likely due to high imbalance ratio of two different types of sample data (students’ gender) applied. Such shortcomings can be mitigated by introducing new samples with more balanced type of sample populations.

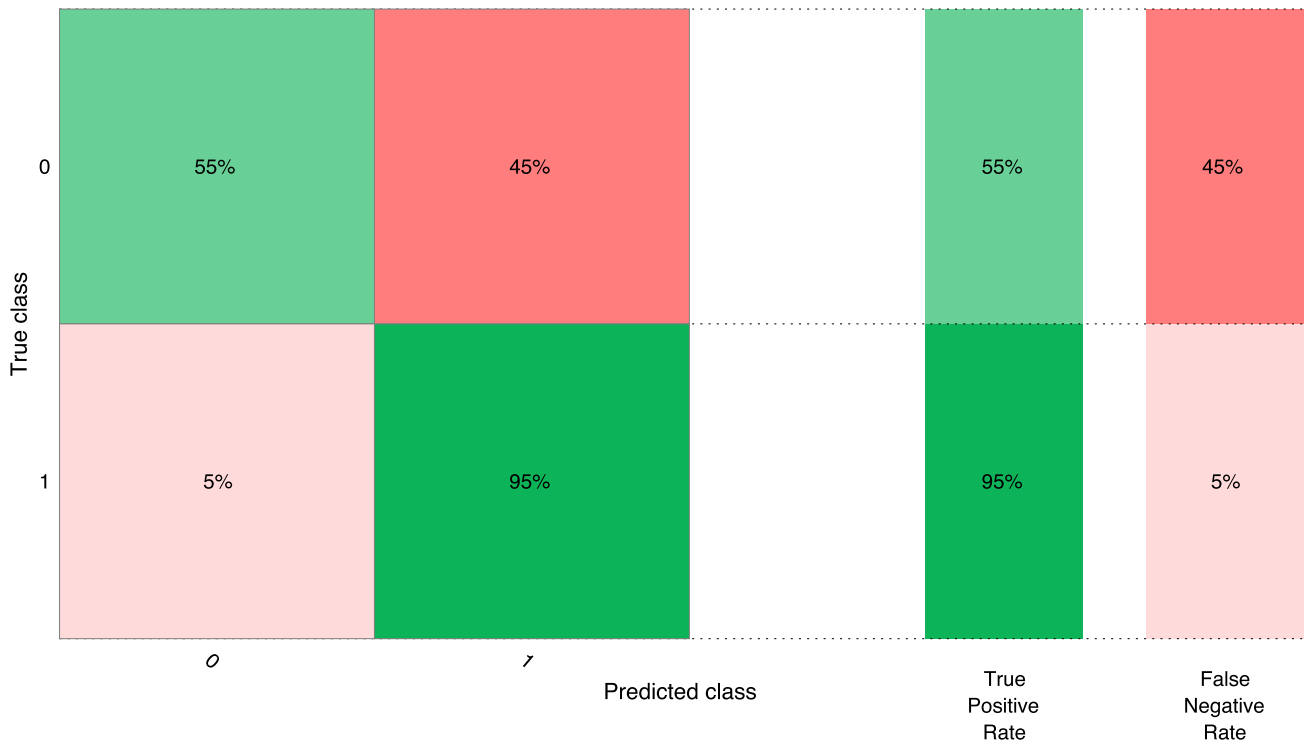


Fig. 5 Binary classifier of confusion matrix based on classification results by ANN in terms of percentage

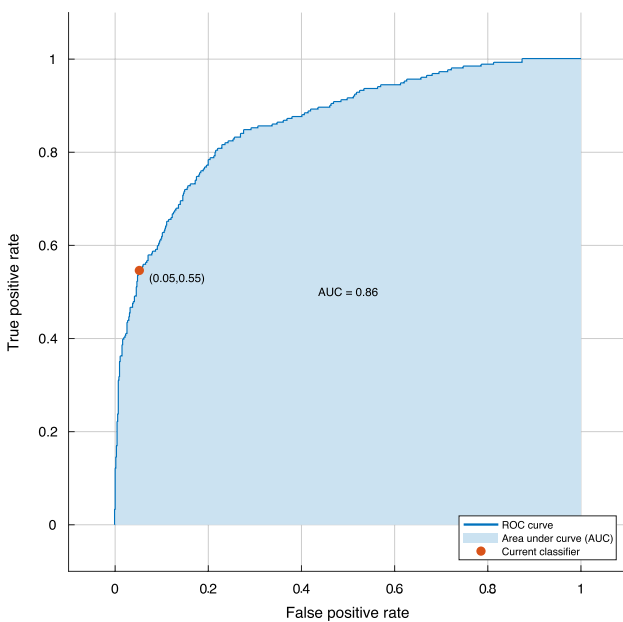


Fig. 6 ROC curve of ANN. AUC-area under the ROC Curve

In the light of the results, it is strongly believed that education modelling settings using ANN can help to evaluate students’ performance. The future models may include other attributes such as the lecturers’ role,

additional curricular classes, course feedbacks and e-learning system are immediate tasks in order to complement a complete ANN modelling, while focusing on ANN’s prediction accuracy. We strongly believe that the future education with Artificial Intelligence-based analytics setting can further enhance the capability of students in achieving high academic performance within academic institutions.

Funding This research is partially supported by Project DIA170362 (2017–2020) under the 13th Five-Year Plan for National Education Science 2017 Projects.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 49:1225–1231
2. Kardan AA, Sadeghi H, Ghidary SS, Sani MR (2013) Prediction of student course selection in online higher education institutes using neural network. *Comput Educ* 65:1–11
3. Hu B (2017) Teaching quality evaluation research based on neural network for university physical education. In: *IEEE*, pp 290–293. <https://doi.org/10.1109/ICSGEA.2017.155>
4. Özçelik S, Hardalaç N (2011) The statistical measurements and neural network analysis of the effect of musical education to musical hearing and sensing. *Expert Syst Appl* 38:9517–9521
5. Vandamme JP, Meskens N, Superby JF (2007) Predicting academic performance by data mining methods. *Educ Econ* 15(4):405–419
6. Kose U, Arslan A (2017) Optimization of self-learning in computer science engineering course: an intelligent software system supported by artificial neural network and vortex optimization algorithm. *Comput Appl Eng Educ* 25:142–156
7. Rashid TA, Ahmad HA (2016) Using neural network with particle swarm optimization. *Comput Appl Eng Educ* 24:629–638
8. Villasenor EA, Arencibia-Jorge R, Carrillo-Calvet H (2017) Multiparametric characterization of scientometric performance profiles assisted by neural networks: a study of mexican higher education institutions. *Scientometrics* 110:77–104
9. Mustafa HMH, Al-Hamadi A, Al-Ghamdi SA, Hassan MM, Khedr AA (2013) On assessment of students' academic achievement considering categorized individual differences at engineering education (neural networks approach). In: *IEEE*, pp 1–10. <https://doi.org/10.1109/ITHET.2013.6671003>
10. Bajaj R, Sharma V (2018) Smart education with artificial intelligence based determination of learning styles. *Proc Comput Sci* 132:834–842
11. Moridis CN, Economides AA (2009) Prediction of student's mood during an online test using formula-based and neural network-based method. *Comput Educ* 53:644–652
12. Bahadır E (2016) Using neural network and logistic regression analysis to predict prospective mathematics teachers' academic success upon entering graduate education. *Educ Sci Theory Pract* 16(3):943–964
13. Zhang Q, Kuldip C, Devabhaktuni VK (2003) Artificial neural network for rf and microwave design—from theory to practice. *IEEE Trans Microw Theory Tech* 51(4):1339–1350
14. MathWorks (2013) Feature selection and feature transformation using classification learner app. <https://uk.mathworks.com/help/stats/feature-selection-and-feature-transformation.html#buwh6hc-1>. Accessed 17 June 2019
15. Yu H, Wilamowski BM (2011) *Leverberg Marquardt training industrial electronic handbook, intelligent systems*, vol 5, 2nd edn. CRC Press, Boca Raton
16. Wilson P, Mantooth HA (2013) *Model-based engineering for complex electronic systems*. Newnes, Oxford
17. MathWorks (2019a) trainlm—Leverberg—Marquardt backpropagation. <http://uk.mathworks.com/help/nnet/ref/trainlm.html>. Accessed 17 June 2019
18. MathWorks (2019b) Body fat estimation. <https://uk.mathworks.com/help/nnet/examples/body-fat-estimation.html>. Accessed 17 July 2019
19. Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* 10(3):1–21
20. Grossman MR, Cormack GV (2013) Edrm page & the grossman-cormack glossary of technology-assisted review, with foreword by john m. facciola, u.s. magistrate judge. *Federal Courts Law Rev* 7(1):1–33
21. Data School (2014) Simple guide to confusion matrix terminology. <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>. Accessed 7 July 2018

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.