CrossMark

Predicting students' final degree classification using an extended profile

Sahar Al-Sudani [1] · Ramaswamy Palaniappan [1]

Abstract

The students' progression and attainment gap are considered as key performance indicators of many universities worldwide. Therefore, universities invest significantly in resources to reduce the attainment gap between good and poor performing students. In this regard, various mathematical models have been utilised to predict students' performances in the hope of informing the support team to intervene at an early stage of the *at risk* student's at the university. In this work, we used a combination of institutional, academic, demographic, psychological and economic factors to predict students' performances using a multi-layered neural network (NN) to classify students' degrees into either a good or basic degree class. To our knowledge, the usage of such an extended profile is novel. A feed-forward network with 100 nodes in the hidden layer trained using Levenberg-Marquardt learning algorithm was able to achieve the best performance with an average classification accuracy of 83.7%, sensitivity of 77.37%, specificity of 85.16%, Positive Predictive Value of 94.04%, and Negative Predictive Value of 50.93%. The NN model was also compared against other classifiers specifically k-Nearest Neighbour, Decision Tree and Support Vector Machine on the same dataset using the same features. The results indicate that the NN outperforms all other classifiers in terms of overall classification accuracy and shows promise for the method to be used in Student Success ventures in the universities in an automatic manner.

Keywords Attainment gap · Degree classification · Neural network · Student's performance · Student success

---

✉ Sahar Al-Sudani
  S.Al-Sudani@kent.ac.uk

  Ramaswamy Palaniappan
  R.Palani@kent.ac.uk

[1] Data Science Research Group, School of Computing, University of Kent, Chatham ME4 3JE, UK

1 Introduction

The student success statistics such as progression and retention are commonly regarded as the main indicators of institutional performance. Additionally, attainment gap is considered as one of the factors that contribute in the institutional performance (HESA 2018). Many universities and more specifically UK based universities are concerned with the attainment gaps and students' performances. Early identification of students at risk of low performance is of great importance to students and the university, which will give the opportunity to put in place effective interventions to reduce the attainment gap. In the UK universities, the Equality Challenge Unit (ECU) is responsible for analysing the demographic (ethnicity, gender, age and disability) profiles of students and staff across the UK universities with the aim of focusing on areas that universities requires to address to provide an inclusive environment for all students and staff (ECU 2018).

The phrase *degree attainment gap* refers to the difference in *good degrees* i.e. a First or an Upper Second degree classification awarded to different groups of students. For example, in UK, *ethnicity attainment gap* refers to the difference between the proportions of White British students who are awarded a First or Upper Second degree classification and the proportion of UK-domiciled Black and Minority Ethnic (BME) students who are awarded the same degree classifications (ECU 2018).

Statistics for the past five years shows that the biggest attainment gap is found within students from ethnic backgrounds. In 2015/16, the attainment gap between white students and BME was largest in England, where 78.8% of white qualifiers received a First or an Upper Second class degree while it was 63.2% for BME qualifiers i.e. a 15.6% attainment gap (ECU 2018). This has led UK universities to research further and invest in projects such as *Student Success Initiatives* to reduce the attainment gap. But one of the main requirements in such projects would be to first identify students at risk of low performance at early stage.

Several research studies have explored various approaches to predict students' performances at various stages of their study. This included using different types of prediction variables in addition to using various mathematical prediction models such as Bayesian, Decision Tree (DT), Random Forest (RF), k-Nearest Neighbours (kNN), Naïve-Bayes (NB), Repeated Incremental Pruning to Produce Error Reduction, and Support Vector Machine (SVM) (Bekele and McPherson 2011; Yadav et al. 2012; Koutina and Kermanidis 2011; Shahiri and Husain 2015; Aulck et al. 2016).

Historically, demographic and cognitive features were mainly used to predict students' performances. According to Kotsiantis and Pintelas (2005), demographic characteristics and students′ marks were used as training set for a regression model in order to predict the students' performances; however psychological and economic related factors were less incorporated in the model. Other authors such as Ikbal et al. 2015, Hoe et al. 2013, and Oladokun et al. 2008 used similar demographic factors, for example Hoe et al. 2013 used data mining tool that identified the most significant correlation of variables associated with academic success based on 10 years of demographic and students' performance data.

In this research, we utilised a feed forward multi-layer neural network (NN) with nine selected input features to model and classify students' performances. This prediction model can be used to *automatically* inform the support team to intervene at an early stage of a student's life at the university. We used a novel extended combination of academic, demographic, institutional, psychological and economic factors to predict

students' final year performances, which is the main contribution of this work. We believed that such a combination of more than one type of feature domain can be utilised to produce good results regardless of the dataset size. Consequently, combining few features from various domains will capture the required information about students' profile that will be useful for the NN model to predict their degree classification.

Academic factors included pre-entry features such as entry qualifications and tariff points (which are pre-university grades converted into points), while demographic features selected for this model were ethnicity, age, and disability. Psychological motivation factors included date of the application for a course (during the normal or the Clearing period i.e. late application process), and whether student maintained the same level of activity (i.e. engagement with course) in the second year.

Other factors were the student's financial status, which was captured according to the area they live in and whether they commuted to university or not.A set of additional institutional factors such as program of study and fee type and other factors such as address and gender were used at the early stage of the analysis, but then removed due to redundancy or weak correlation to the prediction model.

The selected input factors were used to classify students as either at no risk of low performance which means they could obtain a "Good Degree" i.e. First class or Upper Second degree or at risk of low performance which means they could obtain either Lower Second or Third class degree i.e. a "Basic Degree".

Three-layered NN architecture was used as the main classifier and performance measures such as averages of classification accuracy, sensitivity, specificity, Positive Predictive and Negative Predictive values were obtained. To evaluate the effectiveness of NN, we also used three other classifiers on the same dataset using the same features and compared their performances. These classifiers were kNN, DT, and SVM. Unlike other research, we considered using a small dataset rather than large datasets that have been used in previous research. In addition, including a combination of features that captures various relevant domains, rather than selecting large number of features of the same domain, is another main contribution of this work.

2 Related research

NN models have been proven as excellent prediction methods for various applications such as image processing, speech recognition, and other pattern recognition applications. They have also been used for predicting students' performances especially using multi-layer NN (MLNN) and features selection plays a vital role in the network's performance (Siri 2015; Cerny and Proximity 2001; Kumar 2012; Wang and Mitrovic 2002). For example, the work presented by Siri 2015 considers personal and academic careers data related to 810 students enrolled in first year of a health care professional course. MLNN was used with 49 variable input factors and 34 nodes in the hidden layer and with output layer predicting students' dropout rate. The prediction model classified students as either 1, 2 or 3 denoting regular student, irregular student, students at risk of abandonment, respectively. The selected features were mainly captured from students' demographics and academic performances but did not include other important features such as economical, physiological etc. that would have a considerable impact on overall students' performances.

Lesinski et al. 2016 used a MLNN with 39 variables in input layer, one hidden layer, and three output nodes to predict performance of large dataset that consisted of 5100 students. The number of hidden nodes was varied from 10 to 70 while examining the impact on model's performance. No other classifiers were considered in this work.

Recent research at University of Washington by Mason et al. 2018 used deep learning NN to predict students' performances and compare its classification accuracy with MLNN and other classification models such as logistic regression. The model utilised 58 student variables related to demographics and academic background, which have historically been connected with engineering student attrition.

Students' learning activities were also used as candidate features that worked well with the prediction models. Okubo et al. 2017 used Learning Management System logs of 108 students attending a specific course (Information Science) with Recurrent and Convolutional NNs to identify students at risk of low performance.

Another study by Gerritsen 2017 compared MLNN classification performance against six other classifiers on the same dataset. These classifiers were NB, kNN, DT, RF, SVM and logistic regression. Results showed that MLNN with three hidden layers with 16 nodes each outperformed all six classifiers in terms of accuracy and was on par with the best classifiers in terms of recall. Learning activities such as course ID, number of sessions, total time online, average length of login session, total number of clicks etc. have been used with these prediction models.

Adejo and Connolly 2017 studied a conceptual multi-dimensional framework that considered six interconnected variable-domains: demographic (age, ethnicity, gender), cognitive (examination marks, presentation skills), economic (income, parents' financial status), personality (learning style, motivation), institutional (program of study, learning environment, support), and psychological (self-efficacy, achievement, interest). The authors' stress on using a combination of factors from all these domains to complement each other to predict students' performances.

3 Methodology

3.1 Data gathering and pre-processing

The data used in this work is a historical dataset of 481 students at a case study university. The population consists of 83% males and 17% females, with most of the students i.e. about 86% of them living closer to the University campus while 14% are commuting from elsewhere. Eighty five percentage of these students are young students (i.e age between 18 and 22 years old) while mature students constitutes the rest 15%. Sixty eight percentage of the students are White British and 32% are BME. Only 22% of the students' entry qualifications are Level-3 Business and Technology Education Council (BTEC) Higher National Diploma while the rest 78% obtained Advanced Level qualifications.

As explained earlier, the aim of designing the NN model was to predict students at risk of low performance at early stage of their studies so that staff can intervene in good time. The dataset provided for this research was initially consisted of 13 attributes related to academic, institutional, demographic, psychological and financial

background: *entry qualification, tariff points, program of study, fees type, address, age, gender, ethnicity, disability, late application, being active in second year, living area and commuting*. These types of features are commonly used in Student Success projects in UK universities to manually analyse students' performances (Equality Challenge Unit (ECU), 2018).

The dataset required pre-processing of its content. It has been noticed that a small percentage (2% of the population) has missing information, for example some students did not declare their ethnicity. Therefore, we excluded these students' records from the analysis. For any new data, the normal approach that can be used when there is a missing data is to replace the missing value by the average or median value, which can be applied on continuous variables such as missing tariff points. Despite the existence of various algorithmic techniques that assist mainly to fill the missing data such as regression, we think that using methods like boosting to deal with missing data can be used more successfully with our model and this is something that we will explore in the future.

Excluding missing data from the analysis resulting with only 470 students that have complete record. The next step was to analyse attributes with high similarity, for example the student *address* attribute provided the postcode information which was also provided in the *living area* attribute that classify students' economic status based on postcode. Therefore, address attribute was considered as a redundant variable and was removed. Pearson's correlation and significance level were calculated for all attributes and in relation to the output class. The features with low Pearson correlation factor or if its significance level was >0.05 were removed, for example *gender, program of study*, and *fees type*.

It is standard practice to normalise the inputs before using them with the NN or any classifier. Here, the normalisation step was required for both the input vectors and the output vectors in the data set. In this way, the network always provides a normalised output range. Therefore, all categorical attributes such as *age* (either young (18–22) or mature $>= 23$), were converted to a vector representation of [0 1] and [1 0]. While continuous values such as *tariff point*s were scaled using the equation:

$$tariff_{normalised} = \frac{tariff - min}{max - min} \tag{1}$$

where *min* and *max* represent the minimum and maximum values of tariff points.

After the elimination of the redundant variables, the final set of nine attributes were used as 17 input vectors to the NN as listed in Table 1. The output variable value was either that student will obtain a good degree or student will obtain a basic degree.

This was represented in the format of [0 1] and [1 0].

3.2 Model architecture

In this work, we utilised a NN with one hidden layer to classify students' performances. After data pre-processing, the general network model was created with 17 input nodes and two output nodes but with three different architectures as described next.

Table 1 Attributes with description and type along with Pearson correlation and significance level

Attribute	Description	Type	Pearson correlation	Significance level
Entry qualification	Pre-entry qualification is either Advanced-Level or Level-3 Higher National Diploma	Categorical	0.242	1.87e-82
Late application	Date of the application for a course (either during the normal or the Clearing period)	Binary	0.025	4.49e-90
Commuting	Students who lives away from university i.e. commuter or not	Binary	0.143	1.09e-92
Disability	Student is disabled or not	Binary	0.003	2.68e-68
Ethnicity	Student is either white British or Black or Minority Ethnic (BME)	Categorical	−0.295	6.33e-35
Age	Student is either young (18–22) or mature $>= 23$	Categorical	0.084	1.50e-88
Motivation	Student is either active or not active in the second year	Binary	0.288	9.71e-101
Living area	Students economic status by living area	Binary	−0.129	3.45e-45
Tariff points	Converted qualification and grades into points	Continuous	0.147	1.30e-78

The dataset related to the 470 students was divided randomly into three sets: training (70%), testing (25%) and validation (5%). Out of the 470 students, 330 students were awarded good degrees while 140 students were awarded a basic degree each.

The 17 nodes input vector to the NN represents the normalised feature vector that captured students' profile, for example the vector [0 1, 1 0, 0 1, 1 0, 0 1, 0 1,1 0,1 0, 0.0717] represents a student with level-3 HND entry qualification, applied late to university (i.e. during the clearing period), living close to the university campus, having no disability, White British, young, active in the second year, from widening participation area, and had 479 tariff points on entry (479 has been normalised to 0.0717).

The neural network classifies students' award as either "Good Degree" or "Basic Degree". In the example of the student above, the two nodes output vector for the neural network during training could be [0 1] to indicate a "Basic Degree" or [1 0] to indicate a "Good Degree". When used to predict a new record, the maximum value of the two outputs of the neural network are used to decide whether it is a "Basic Degree" or a "Good Degree". For example, a value of [0.6 0.1] would indicate that the neural network has predicted the class "Good Degree"; this will be matched with actual label to decide if the classification is correct or incorrect.

3.2.1 Feed-forward network

As shown in Fig. 1, generated using MATLAB software (MATLAB 2018), the feed-forward (FF) architecture consisted of 100 nodes in the hidden layer with sigmoid function and pure linear function for output layer. The training algorithm used was Levenberg-Marquardt with Mean-Squared Error (MSE) used for error computation to be backpropagated.
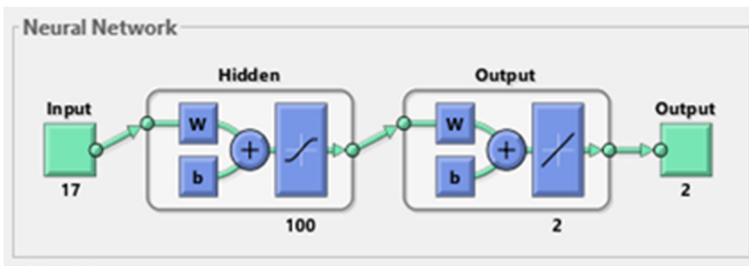
Fig. 1 Feed-forward network

3.2.2 Cascade-forward network

Cascade-forward (CS) networks are similar to feed-forward networks but include a connection from the input and every previous layer to following layers. As with feed-forward networks, a two-or more layer cascade-network can learn any finite input-output relationship arbitrarily well given enough hidden nodes. As depicted in Fig. 2 (also generated from MATLAB software (MATLAB 2018)), it consisted of 100 nodes in the hidden layer with sigmoid function and pure linear function for output layer. The training algorithm was Levenberg-Marquardt with MSE used for error computation.

3.2.3 Feed-forward network variation

This is depicted in Fig. 3 and is a variation of standard feed-forward. The training algorithm used here was Scaled Conjugate Gradient with Cross Entropy for performance calculation. The various learning and error computation methods were chosen after some preliminary simulations.

4 Results and discussion

In this section we will discuss the performance of the networks. Before training the NN, as stated above, the dataset was divided randomly into three sets: training (70%), testing (25%) and validation (5%). These sets were applied to the three NN models
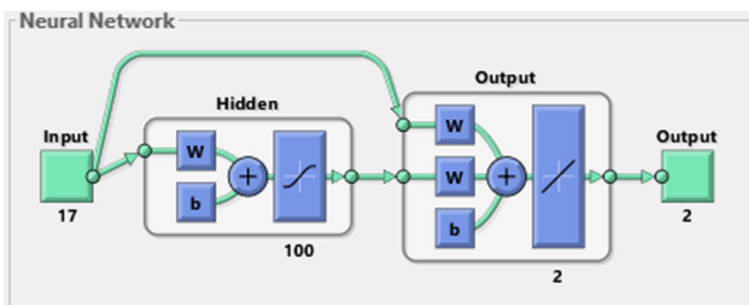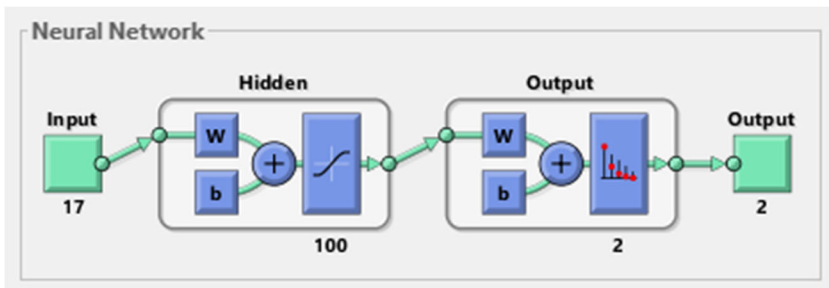

Fig. 2 Cascade-forward network

Fig. 3 FF variation network

described in the previous section. Table 2 shows the average of the performances using ten-fold cross validation implemented on the three NN models where the classification accuracy, standard deviation, sensitivity, specificity, PPV, and NPV are calculated for each model.

The FF network achieved the best results. The overall mean classification accuracy for the model was $83.7 \pm 1.5\%$. This is in par with the second model but better than the third model.

Figure 4 shows the confusion matrices of the average classification results of both FF and CF networks with best validation performance graphs. The rows in the confusion matrices show the predicted class, and the columns show the actual class. The first column shows True Positive, False Positive and PPV. The second column shows False Negative, True Negative and NPV. The last column shows specificity, sensitivity and overall accuracy.

In the sample considered, 330 students were awarded good degrees while 140 students awarded a basic degree. The FF NN correctly predicted 310 students who obtained a good degree and 71 students who obtained a basic degree, the results were similar to CF network.

As the motivation of this work was to identify the students with risk of low performance at early stage, additional experiments have been carried out to improve the NPV results by modifying the network design to increase NPV results. The modified network architecture was 17 input nodes, one output node, with 50 nodes in the hidden layer. Using average of the ten-fold cross validation result, NPV result was improved to 56%. Moreover, we changed the threshold value used to distinguish between the two classes of output to 0.6 (i.e. increased the confidence in the classification, rather than just taking output values closer to 0 or 1 as the predicted label) which further increased average NPV to 60%.

Table 2 Classification results of the three NN models

NN-model	Accuracy	Standard deviation	Sensitivity	Specificity	PPV	NPV
FF	83.7%	1.52	77.37%	85.16%	94.04%	50.93%
CF	83.5%	1.91	77.18%	85.30%	94.74%	50.01%
FF variation	80.8%	2.65	67.09%	82.63%	94.50%	40.02%

(a)

Confusion Matrix of Feed-forward NN

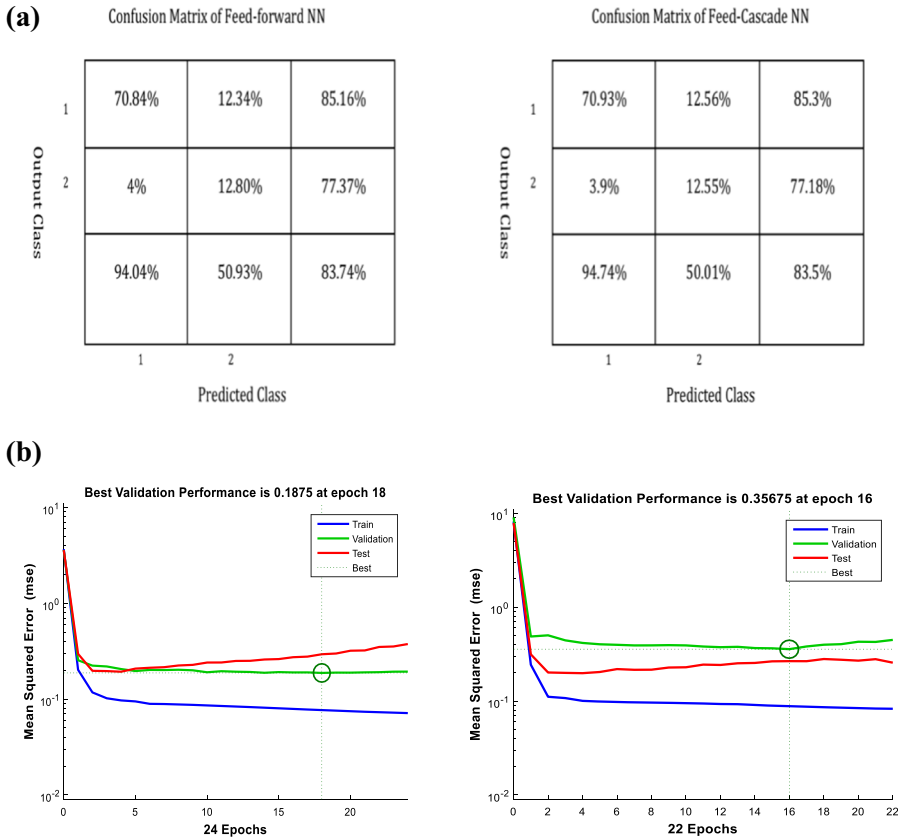Confusion Matrix of Feed-Cascade NN

(b)

Fig. 4 **a** Confusion matrices of the first two classifiers and **b** best validation performance

To evaluate the overall performance of NN in predicting students' performance, we compare the best-performed NN model with other classifiers using the same dataset. These classifiers used here were DT, SVM and kNN.

DT can be used for classification problems in addition to regression. It builds tree like structure using the attributes and partitions the examples using the attribute values. This is recursively repeated until the regressed value or class label is obtained. Usually, information gain is used to select the order of the attributes when building the tree. For this work, we used RF as the usage of many trees would give improved performance over a single tree. The *fitcensemble* built-in function available in MATLAB to train and

Table 3 Mean classification results and standard deviation values of ten-fold cross validation of all classifiers

Model	Classification accuracy
Feed forward NN	$83.7 \pm 1.5\%$
DT	$76.9 \pm 3.2\%$
SVM	$72.9 \pm 4.0\%$
kNN	$71.7 + 4.4\%$

Table 4 Results of ANOVA on accuracy scores between classifier

ANOVA						
Source of variation	SS	df	MS	F	P value	F crit
Between Groups	882.59	3	294.20	21.21	4.7E-08	2.86
Within Groups	499.38	36	13.87			
Total	1381.98	39				

cross validate RF model. As we had two classes for the output vector, *LogitBoost* was used as the ensemble-aggregation algorithm and 100 trees were composed in the ensemble.

SVM is a supervised learning model for binary problems though kernel trick can be used for multi-class problems. The model maps instances in space points with widest gap possible and a new instance is predicted based on the where it falls in the space. In this work, we utilised *fitcsvm* built-in function in MATLAB to train and cross validate the SVM model. In this case, we used a linear function due to the two class problem.

Another classifier used here was kNN, which is a relatively straightforward classifier that does not require model to be built prior to classifying instances. A new test instance is classified using distance measures computed to the instances in the training data set. A set of neighbours is used (this is known as k) and class of the test instance is predicted using majority class labels of the nearest k neighbours (Alpaydin 2009). Usually, k is chosen to be odd number and in this work, a rule of thumb was used, $k = $ sqrt (N) where N is the total number of training instances, giving $k = 21$. The MATLAB function *fitcknn* was utilised. Euclidean was the distance function used with exhaustive searcher for neighbour searching.

When looking at the results depicted in Table 3, the accuracy varies widely between classifiers, ranging from the lowest kNN with accuracy score of 71.7% to the highest NN with score of 83.7%. SVM scored higher than kNN by 1.26% and RF accuracy was lower than NN by 6.8%. Analysis of variance was used to determine if the difference in means between the classifiers was significant. It was performed using the output of the ten-fold cross validation results. Results of analysis are shown in Table 4. A p value was obtained below 0.05, confirming that the difference in results between these classifiers is significant. T-test was also been carried out to identify which classifier was the best with statistical significance. As shown in Table 5, the p value of all classifiers compared to NN were less than the significance level of 0.05 and comparing the pair-wise classifiers (kNN vs DT, kNN vs SVM, DT vs SVM), it is obvious that NN best performance is statistically valid.

Table 5 Results of t-test - two samples assuming equal variances

Classifier group	NN-KNN	NN-DT	NN-SVM	kNN-DT	SVM-DT	kNN-SVM
P(T < =t) two-tail	5.4e-7	2.8e-5	7.8e-7	1.1e-2	3.6e-2	5.3e-1

Table 6 Comparison with previous research

Paper	Input vector size	Dataset size	Features' domain	Accuracy %	Sensitivity	Specificity	PPV	NPV	Compared with other models
Our Work	17	470	Academic Demographic Institutional Psychological Economic	83.7	77.37	85.16	90.04	50.93	Yes
(Gerritsen 2017)	14	4601	Academic	66.1	–	–	–	–	Yes
(Lesinski et al. 2016)	39	5100	Academic Demographic	95.8	89.5	85.5	–	–	No
(Siri 2015)	43	810	Academic Demographic	84	81	76	–	–	No
(Mason et al. 2018)	58	682	Academic Demographic	90.1	71.2	94.5	–	–	Yes

Table 6 presents a summary comparing our work with previous research conducted using NN model to predict students' performance. The comparison is based on input vector size, dataset size, feature domain, prediction accuracy, sensitivity, specificity, Positive Predictive Value, Negative Predictive Value, and whether the method was compared with other models or not.

As shown in the table, all the previous research studies have used large sized datasets with the aim of increasing the performance of the neural network. While in this work, we have used a much smaller dataset and with much fewer input features but with a combination of features from various domains that can capture the required information for the performance prediction.

Large sample is essential to train the NN model. As such, we believe that using our model, one could obtain much improved degree of prediction accuracy if a larger dataset be used with the proposed extended profile of students i.e. a combination of features from various domains.

5 Conclusions

In this research, we utilised several classifiers to predict students' degree classifications. The feed forward neural network model is the best performing model which can be used to inform the student support team to intervene and provide support at good time. The proposed neural network model has been tested using a small dataset in comparison with other work. In addition a combination of features from more than one domain (academic, institutional, physiological, demographic and economic) have been used in the model which achieved good results. The NN gave statistically valid best performance of 83.7% when compared with RF, SVM and kNN. The model was further modified to improve NPV value to 60% from 56% using modified confidence threshold approach.

The findings presented in this work were constrained by a number of limitations. First, relating to the sample size which was restricted to a small dataset of only 470 students and it is possible that NN classification performance can be increased by increasing the sample size. Second, the model presented utilised shallow learning networks only, deep learning networks with larger sample size are left for future consideration.

Overall, the results indicate that using an extended combination of domain information of the students can give a good prediction performance of the degree classification to be obtained. This can be used to intervene and help to minimise the attainment gap and improve the performance of the students.

References

Adejo, O., & Connolly, T. (2017). An integrated system framework for predicting students' performance in higher education institution. *International Journal of Computer Science and Information Technology (IJCSIT), 9*(3), 149–157.

Alpaydin, E. (2009). *Introduction to machine learning*. MIT press.

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. *arXiv preprint arXiv:1606.06364*.

Bekele, R., & McPherson, M. (2011). A Bayesian performance prediction model for mathematics education: A prototypical approach for effective group composition. *British Journal of Educational Technology, 42*(3), 395–416.

Cerny, P. A., & Proximity, M. A. (2001). Data mining and neural networks from a commercial perspective. *In ORSNZ Conference Twenty Naught One, pp.*, 1–10.

Equality Challenge Unit (ECU). (2018) https://www.ecu.ac.uk/guidance-resources/student-recruitment-retention-attainment/student-attainment/degree-attainment-gaps/. Accessed on 20 August 2018.

Gerritsen, L. (2017). *Predicting student performance with Neural Networks doctoral dissertation*. Tilburg University.

Higher Education Statistical Agency. (2018) https://www.hesa.ac.uk/data-and-analysis/performance-indicators/guide. Accessed on 20 August 2018.

Hoe, A. C. K., Ahmad, M. S., Hooi, T. C., Shanmugam, M., Gunasekaran, S. S., Cob, Z. C., & Ramasamy, A. (2013, November). Analyzing students records to identify patterns of students' performance. In *In IEEE International Conference on Research and Innovation in Information Systems (ICRIIS)* (pp. 544–547).

Ikbal, S., Tamhane, A., Sengupta, B., Chetlur, M., Ghosh, S., & Appleton, J. (2015). On early prediction of risks in academic performance for students. *IBM Journal of Research and Development, 59*(6), 5–1.

Kotsiantis, S. B., & Pintelas, P. E. (2005, July). Predicting students marks in hellenic open university. In *In Fifth IEEE International Conference on Advanced Learning Technologies (ICALT)* (pp. 664–668).

Koutina, M., & Kermanidis, K. L. (2011). Predicting postgraduate students' performance using machine learning techniques. In *Artificial intelligence applications and innovations* (pp. 159–168). Berlin, Heidelberg: Springer.

Kumar, D. A. (2012). Appraising the significance of self-regulated learning in higher education using neural networks. *International Journal of Engineering Research and Development, 1*(1), 09–15.

Lesinski, G., Corns, S., & Dagli, C. (2016). Application of an artificial neural network to predict graduation success at the United States military academy. *Procedia Computer Science, 95*, 375–382.

Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. *Research in Higher Education, 59*(3), 382–400.

MATLAB and Statistics Toolbox Release. (2018). The MathWorks, Inc. In *Natick*. Massachusetts: United States.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017, March). A neural network approach for students' performance prediction. In *ACM Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp. 598–599.

Oladokun, V. O., Adebanjo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology, 9*(1), 72–79.

Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72*, 414–422.

Siri, A. (2015). Predicting students' dropout at university using artificial neural networks. *Italian Journal of Sociology of Education, 7*(2).

Wang, T., & Mitrovic, A. (2002). Using neural networks to predict student's performance. *International Conference on Computers in Education*, 969.

Yadav, S.K., Bharadwaj, B., & Pal, S. (2012). Mining education data to predict student's retention: A comparative study. *arXiv preprint arXiv:1203.2987*.