A Deep Neural Network-Based Prediction Model for Students' Academic Performance

[1*]Ghaith Al-Tameemi, [1]James Xue, [1]Suraj Ajit, [1]Triantafyllos Kanakis,
[2]Israa Hadi, [3]Thar Baker, [4]Mohammed Al-Khafajiy, and [5]Rawaa Al-Jumeily
[1]*University of Northampton*, Northampton, United Kingdom
[2]*University of Babylon*, Babil, Iraq
[3]*University of Sharjah*, Sharjah, United Arab Emirates
[4]*University of Reading*, Reading, United Kingdom
[5]*Liverpool John Moores University*, Liverpool, United Kingdom
[1]{ghaith.al-tameemi, James.Xue, Suraj.Ajit, Triantafyllos.Kanakis}@northampton.ac.uk
[2] israa_hadi@itnet.uobabylon.edu.iq
[3] tshamsa@sharjah.ac.ae
[4] m.d.al-khafajiy@reading.ac.uk
[5] R.AlJumeily@2019.ljmu.ac.uk

***Abstract*—Education providers are increasingly using artificial techniques for predicting students' performance based on their interactions in Virtual Learning Environments (VLE). In this paper, the Open University Learning Analytics Dataset (OULAD), which contains student demographic information, assessment scores, number of clicks in the virtual learning environment and final results, etc, has been used to predict student performance. Various techniques such as standardisation and normalisation have been employed in the pre-processing stage. Spearman's correlation coefficient is used to measure the correlation between the activity types and the students' final results to determine the importance of the activities. Deep learning has been utilised to predict students' performance based on their engagement in the VLE. The empirical results show that our model has the ability to accurately predict student academic performance.**

***Index Terms*—deep learning, student engagement, correlation coefficient, student performance.**

I. Introduction

Due to rapid development of technologies and the significant impact of technology on higher education, educational data has grown significantly. Therefore, this has opened opportunities for improving student engagement in VLEs, which benefits the whole educational system [1]. Through the extensive data collected from the VLEs, a variety of research approaches such as predictive learning analytics have been developed to enhance student learning and reduce dropout rates [2].

Different data mining techniques can be used to discover patterns and trends in this variety of data [3]. A number of different statistical methods (e.g., regression analysis, cluster analysis, etc) and AI techniques (e.g, genetic algorithms, and neural networks, etc) have been used to provide more insightful view of student learning activities. Educational database management systems enable easy extraction of useful information that can be fed into data mining models. The adaptation of educational datasets [4] and data mining is known as Educational Data Mining (EDM), which is essential for the exploration of patterns in education and learning behaviour and has attracted plenty of research interests [5].

Deep Learning is a machine learning technique that is categorised by complex computational layers that can be learned from patterns or events rather than by traditional methods [6]. Deep learning has been proven to be useful in resolving many complex issues relating to the extraction of human interpretable patterns [7]. Although various studies have demonstrated that deep learning and learning analytics are popular, there is insufficient proof that deep learning can be used in learning analytics and EDM [8].

Academic performance at many Higher Educational Institutions (HEIs) is primarily monitored by instructors; however, the process is time-consuming [8]. EDM and deep learning can contribute to a better understanding of student behaviours. Through the use of EDM and deep learning, instructors will be able to detect poorly engaged students in VLE so that early interventions can be made to improve the students' academic performance [9].

The purpose of this study is to investigate whether or not student interactions on VLEs have an immediate effect on academic performance. Hence, a deep learning model has been developed to predict student academic performance based on students' interactions in VLE. Since the deep learning model has yielded outstanding results in terms of predicting student academic performance, the findings are useful for the educational sectors.

The remainder of the paper is organised as follows. Section II conducts literature review of predictive learning analytics in higher education institutions by describing the history of this topic as well as the approaches that are primarily used to predict student success. Section III describes the methods that have been adopted to achieve the paper aim. Section IV discusses the experimental results for predictive model. Finally, Section V concludes the paper.

II. Related work

There has been extensive work that examines the significance of predictive learning analytics and EDM in higher education institutions and how they solve related issues in HEIs [10]. *Khobragade et al.* used Naive Bayes and decision tree algorithms to address student failure in courses. They also employed surveys and report data to predict student failure. The data consisted of the students' marks, family information, social features and past performance. Naive Bayes provided better performance than the decision tree with an accuracy of 87% [11]. The work in [12] follows an approach similar to the previous one - Decision tree, rule-based and Naive Bayes algorithms have been used to predict the academic performance of first-year students in computer science courses. The data that was used in this work were demographics, previous academic performance and family information. Based on the experiment results, the rule-based algorithm provided a better understanding than the others, with 71% accuracy [12].

Neural networks have been used widely to predict student performance and identify a student's at-risk status. In 2016, there was a case study at the Bangabandhu Sheikh Mujibur Rahman Science and Technology University (BSMRSTU) to predict student performance using neural networks [13]. The real data of the students was taken to ensure the accuracy of the results. The authors performed a neural network in the form of the Cumulative Grade- Point Average (CGPA) and compared it with the real CGPA. *Tsung-Yen et al.* predicted student grades based on clickstream videos using the Time Series Neural Network [14].

Wei-Xiang et al., proposed a Machine Learning Feature Selection (MLFS) and the Support Vector Machine (SVM) to extract the key features that impact a student's academic performance in a Taiwan elementary school. The data was collected from the same school and included all the student grades as well as fifteen condition features extracted from two different databases (student profile and tutorship records). Using the above techniques, the authors achieved 92% of prediction accuracy [15]. Furthermore, a model has been developed for the Universidad Nacional de Colombia to predict low academic performance based on specific enrolment. Two different data mining approaches – the Naive Bayes and Decision Tree Classifier have been used to predict student performance. The historical data was used in a training model for this task. The decision tree was more consistent than the Naive Bayes classification techniques in terms of test and training datasets [16].

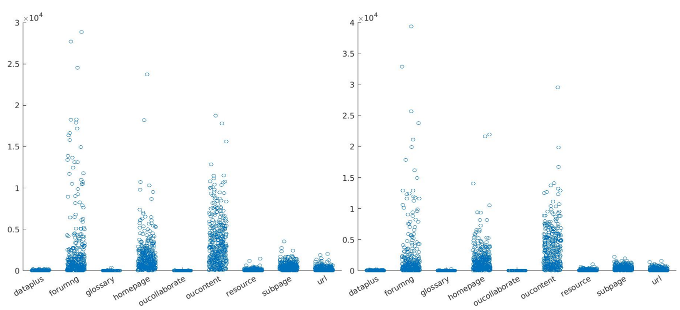
Furthermore, the Open University in the UK has utilised several machine learning algorithms to identify low-engagement students in an e-learning system. The aim is to know how student engagement can be useful in student performance. Different types of data such as results, scores on the assessments, numbers of clicks and higher education levels were collected in this study. Based on the above criteria, the student level of engagement is addressed. The machine learning algorithms that have been used in this study are a decision tree, J48, a Classification and Regression Tree, JRIP Decision Rules, Gradient Boosting Trees and Naive Bayes. The J48, the decision tree, JRIP and the GBT have exhibited better performance in terms of accuracy and kappa values [17].

III. Methodology

This part of the research describes the development process of the deep learning model to predict student performance. Several of the most challenging aspects of predictive learning analytics are discussed, such as data collection and preprocessing. Moreover, deep learning has been illustrated in-depth to predict the students' final results. Figure 3 presents the development process of this work.

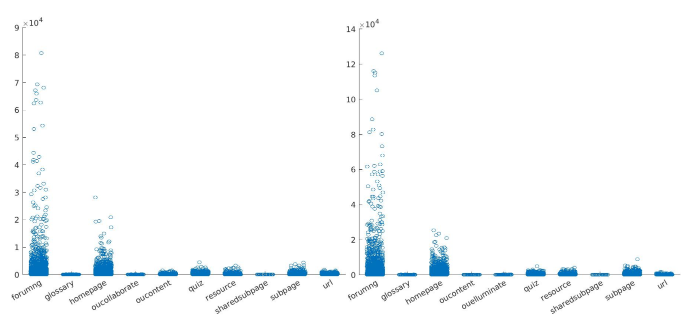
A. Data Collection and Preprocessing

The OULAD has been utilized in this work [18]. The dataset is comprised of student demographic information, logs of student behaviour on the VLE, and student results on each module. Students and course information can be found in seven files in the OULAD: studentInfo, courses, assessments, the VLE, studentRegistration, studentAssessment, and studentVLE. We selected the AAA and BBB modules from the social science course for two presentations: 2013J and 2014J.



(a) 2013J AAA (b) 2014J AAA

Fig. 1. Distribution of student learning activities for the AAA module



(a) 2013J BBB (b) 2014J BBB

Fig. 2. Distribution of student learning activities for the BBB module

1) Feature Extraction: The total number of clicks for each activity type in the VLE has been used as a predictor for the deep learning model. The VLE features are resource (R_e), oucontent (Ou_{cn}), url (U_r), homepage (H_p), subpage (S_{pa}), forumng (F_r), oucollaborate (Ou_{cl}), dataplus (D_p), glossary

TABLE I
THE DESCRIPTION OF THE FEATURES IN OULAD

Feature	Description	Type
id_student	The unique identification of each student in the course	Nominal
F_r	This refers to the discussion forum for the students	Quantitative
R_e	Reference the course materials that available in the VLE such as course lectures slides, books, and lecture notes in PDF and HTML formats	Quantitative
Ou_{cn}	Contains the study materials in HTML format	Quantitative
H_p	Represent the home page of each course	Quantitative
G_l	Reference details about the open university and higher education acronyms.	Quantitative
D_p	Represent the module that developed by the open university to allow the student to view their records	Quantitative
Q_u	Represent the number of quizzes performed by the students in each module	Quantitative
S_r	The final student result at the end of the course which indicates that the student completed the course successfully or not	Categorised (Fail or Pass)

(G_l), ouelluminate (Ou_{el}), sharedsubpage (Sh_p), questionnaire (Q_u), externalquiz (E_q), ouwiki (Ou_w), dualpane (D_{pl}), repeatactivity (R_e), folder (F_l), htmlactivity (H_a), and quiz (Q_u). The output features are the final student result (S_r) (Pass or Fail) that indicate whether the student has successfully completed his course or not. All of these features have fully illustrated in table I.

2) Missing Values: The dataset contains missing values for some activity types. For example, the Q_u activity in the AAA has not been performed by any student. Nevertheless, this activity plays a significant role in predicting the student's final performance in BBB. The predictive model will therefore perform better and faster if these missing values are removed. In addition, some features have also been eliminated from the two modules (AAA and BBB) like Ou_{el}, Sh_p, Q_u, E_q, Ou_w, D_{pl}, and R_e since the student performance was set to zero in each activity type.

3) Standardization and Normalization: Student learning activities in this dataset are not in the same scale as shown in figures 1 and 2, for instance, in AAA modules for the two presentation years, the range of the values in foruming, H_p, as well as the Ou_{cn} are very high compared to the activity types. While, the BBB module has a very high range of values for foruming compared to the other activity types. Thus, we have scaled these data using standardization and normalisation techniques.

Standardisation (x_{stdz}) is one of the standard requirements for machine learning prediction. It has been considered as one of the scaling techniques which centre the values around the mean with the standard deviation unit. By standardizing the values of the features, the machine learning performance can be improved. The standardisation was calculated by taking the difference between x and its mean (μ) value divided by the standard deviation (σ) of x, as it shown in equation 1 [19].

$$x_{stdz} = \frac{x - \mu}{\sigma} \tag{1}$$

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

After implementing the standardisation, the data will be ready to perform the normalisation (x_n) step. Normalisation is the process of scaling the standard data to have a unit norm and to be in the range of 0 to 1. Therefore, the highest value will be 1, and the lowest is 0.

4) Label encoding: The results in this dataset has categorised into four categories which are Withdrawn, Fail, Pass, and distinction. Our experiment in this research aims to identify the pass and fail students based on their interactions on the VLE. Therefore, the distinction and pass students have been merged into one category, namely pass. Hence, two classes were defined for the binary classification namely, class 0: Fail; and Class 1: Pass.

5) Correlation Coefficient Measurement: The Spearman's correlation coefficient is a statistical analysis technique to measure monotonic relationship between paired data. It is denoted by r_s and designed as follow:

$$-1 \leq r_s \leq 1$$

The closer r_s to ± 1 the stronger to the monotonic relationship. Thus, the strength of the correlation has described as follow $.0 - 0.19$ very weak, $0.2 - 0.39$ weak, $0.40 - 0.59$ moderate, $0.60 - 0.79$ strong, and $0.80 - 1$ very strong. This technique has been used to examine the correlation between input features in order to choose the optimum number of features for the predictive model [20]. This technique has been performed using MATLAB for the two modules over two years (AAA 2013J, AAA 2014J, BBB 2013J, and BBB 2014J. In equation 3, d represents the difference of the data and n is the number of pairs of data.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{3}$$

6) Feature Selection: Feature selection is the final step of pre-processing the data before applying the deep learning model. Standardization normalization was applied after removing all the missing values. We then transform the categorised results into a numeric format to make them suitable for deep learning. In addition, to obtain better results for prediction, the correlation coefficient was also used to minimize the number of features and select those that would be more effective.

B. Deep learning for Student Performance Prediction

A deep learning model has been used to predict students' academic performance. The deep learning method is the most popular method of machine learning, and it is widely used in many practical applications [21]. The Feedforward paradigm, one of the most well-known deep learning algorithms, has been used to predict a student's final result in this work. This algorithm divides the network into different layers (input, hidden, and output), and those layers are interconnected. For

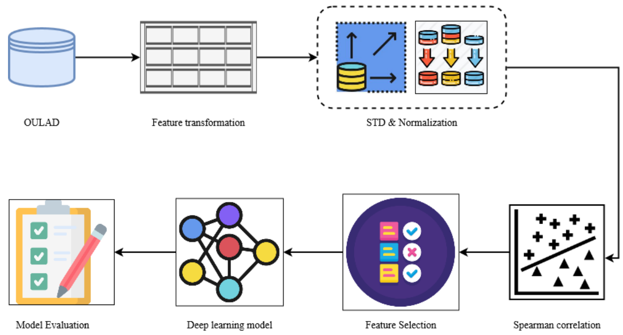


Fig. 3. System overview

instance, the input variables in this work are the number of clicks that the students performed in their course study. While the hidden layers are in the middle of the network as well as the bias which is specific to each neuron. The output layer represents the final student result (pass or fail) of the predictive model.

The problem of over-fitting and under-fitting is common with deep learning, as with many other nonlinear machine learning approaches, particularly because it results in low predictive results when it is far behind the actual range [22]. It is impossible to avoid over-fitting when the number of training samples and testing samples are different. For solving this type of problem, weight decay and early stopping, known as regularization techniques, have been used. A compromise is established with the objective function in regards to the weights parameter, bias parameter, and other parameters [23].

There are two groups of datasets in this model: training datasets and testing datasets. 80% of the dataset is used to train the deep learning model and estimate the neural network weights. The remaining 20% of the dataset were assigned to the testing dataset, which is utilized at the end of the iterative process to ensure consistency and assess the model's performance.

$$y_i = f_i\left(net_i\right) \tag{4}$$

- f_i is the activation function of neurons i.
- net_i is the summation of the weight input nodes.

$$net_i = \sum_{j=1}^{n_j} W_{j,i} Y_{j,i} + W_{i,0} \tag{5}$$

- $Y_{i,j}$ is the input node
- $W_{i,j}$ is the weight of the input nodes
- $W_{i,0}$ is the bias weight

It is imperative to manage the variations in weight by minimizing the error percentage so that the gradient is reduced accordingly [24]. The training algorithm will adjust the weights in the network to reduce the error between the predicted values and the actual target. The choice of the activation function has been considered a crucial step in the neural network prediction process. To generate the output value, the net value will be passed to the activation function after being calculated. The Rectified Linear Unit (ReLU) layer has been chosen as an activation function for the hidden layer to produce the output value.

$$a = f\left(net\right) \tag{6}$$

$$f(net) = \begin{cases} net, net \geq 0 \\ 0, net < 0 \end{cases} \tag{7}$$

The ReLU learns faster than the other activation function such as sigmoid and TanH; also, it solves the vanishing gradient problem. From equation 7, it can be noticed that this activation function provides the output of 0 when the value of the net is less than 0; otherwise, it is the original positive value. Table II shows the network properties.

TABLE II
DESCRIPTION OF THE NEURAL NETWORK PROPERTIES

Parameter name	Function/value
Number of hidden layers in network	2
Number of neurons in hidden layers	10-20
Training method	Bayesian Regularization
Loss function	MSE
Number of epochs	500

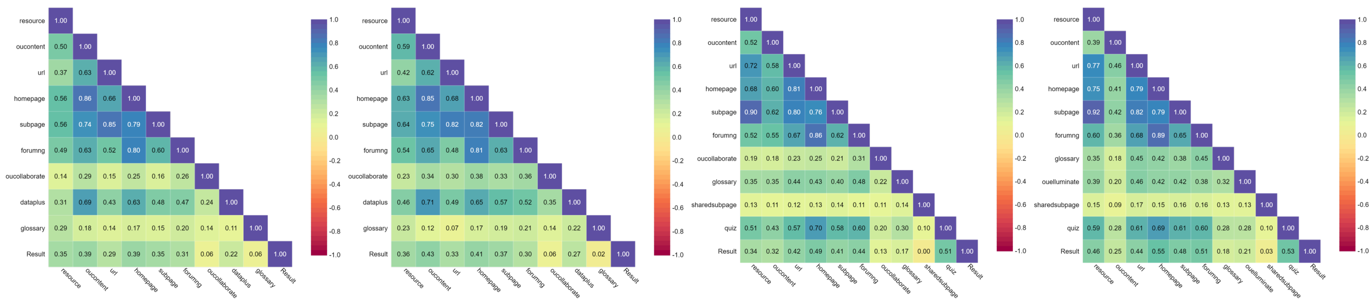
IV. RESULTS

This section presents the obtained results of the deep learning model. As a first step, the statistical analysis of spearman's correlation coefficient is used to determine the extent to which the input features are correlated with the output features. The prediction of student academic performance was also illustrated using deep learning.

A. *Correlation Coefficient*

The spearman's correlation coefficient has been used to measure the strength and direction association between the input features as well as the output feature. In these course modules, however, it is shown that certain types of activities have a positive influence on student final results via this technique. For instance, in the 2013J AAA, the R_e, Ou_{cn}, H_p, S_{pa}, F_r and the U_r are positively correlated to the student final results. Figure 4 represents the correlation coefficient results between the input features as well as the output feature. We can minimize the number of input features in order to improve deep learning performance [24]. In 2013J AAA, we can see that S_{pa} and U_r, F_r and H_p are highly correlated. Consequently, the U_r and F_r have been eliminated from the predictor list. Similarly, in 2014J AAA, the H_p are highly correlated to the Ou_{cn}, S_{pa}, and F_r while, S_{pa} is also correlated to the U_r. H_p and U_r have been eliminated.

The mode of correlation measurement in 2013J BBB is similar to the previous modules. Figure 4c shows that the H_p and Q_u have a good correlation to the final result. S_{pa} is highly correlated to the R_e, H_p and F_r are highly correlated. The low correlated input features to the output features will be eliminated. Therefore, the F_r and R_e was removed from the predictor list. In 2014J BBB, the correlation results are



(a) 2013J AAA (b) 2014J AAA (c) 2013J BBB (d) 2014J BBB

Fig. 4. Figure presents the correlation coefficient results between input features as well as the final results. (a) & (b) shows the correlation results for AAA module and (c) & (d) illustrates the correlation results for BBB module

similar to those achieved in 2013J BBB as S_{pa} and H_p has high correlation to the R_e and F_r respectively.

B. Classification results

Deep Neural Network has been adopted to achieve the objective of this research. Two different modules (AAA and BBB) for 2 presentation years (2013J and 2014J) have been used to estimate the prediction result using deep learning. The number of students in each model is different from the others; for instance, AAA 2013J course has 319 students, AAA 2014J has 292 students, BBB 2013J course has 1492 students, and BBB 2014J has 1154 students. MATLAB has been used to build the deep learning for predicting the final student results. The data has been divided into two groups, training and testing. 80% were assigned for training the deep learning model and 20% were assigned for the testing. Majority of data has been assigned for training the model to avoid the over-fitting issue [24].

Equations 8-14 are used to specify the accuracy (A_c), precision (p_r), recall (r_e), and F_1 score of the predictive model [25]. Equation 8 is to specify the overall accuracy while equations 9 and 10 have been adopted to calculate the sensitivity (S_e) and specificity (S_p). F_1 score has also been used to evaluate the predictive model and the scores are high for all the four models which means that the false positives and false negatives are low. In other words, F_1 score is a harmonic mean of the precision and recall for the predictive model. The data is unbalanced as the number of samples for each class is profoundly different from the other. For instance, in the AAA 2013 course, the Pass class has 277 out of 319 students in the dataset. Kappa ($kappa$) value can be checked to ensure the performance efficiency of the deep learning model.

$$A_c = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$S_e = \frac{TP}{TP + FN} \tag{9}$$

$$S_p = \frac{TN}{TN + FP} \tag{10}$$

$$P_r = \frac{TP}{TP + FP} \tag{11}$$

$$R_e = \frac{TN}{TN + FN} \tag{12}$$

$$F_1 = 2 \times \frac{P_r \times R_e}{P_r + R_e} \tag{13}$$

$$Kappa = \frac{P_o - P_e}{1 - P_e} \tag{14}$$

The obtained results from the deep learning model after applying the spearman's correlation coefficient was significant. The spearman's correlation coefficient was adopted to reduce the number of input features as well as indicating the ones that has a good correlation to the output class (student result).

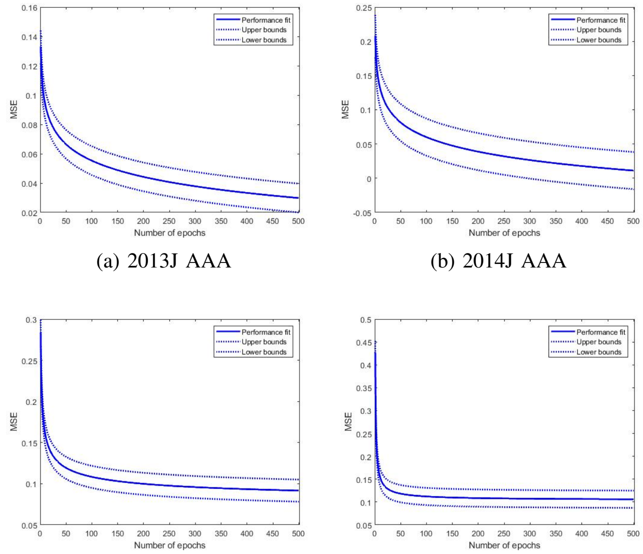
TABLE III
CLASSIFICATION RESULTS

Modules	S_e	S_p	F_1	$Kappa$	A_c
2013J AAA	99.2%	78%	98%	0.834	96.6%
2014J AAA	97.2%	90%	97.8%	0.846	96.2%
2013J BBB	95.3%	88.1%	95.3%	0.835	93.2%
2014J BBB	96.3%	88.8%	96%	0.85	94.2%

Table III presents the results obtained from adopting the deep learning with the spearman correlation. There was a significant positive impact of adopting the spearman correlation with the deep learning model because the A_c, S_e and S_p was high in the two modules. An inspection of the results in table III reveals that the Kappa results for all the modules are close and high in the same time. Figure 5 compares the MSE curves for each module.

V. Conclusion and Future Work

In this study, the effect of student engagement in VLE on student academic performance has been examined. We found that, on average, student engagement in VLE contributed to students' academic achievement. Additionally, the current approach allows teachers to gain a better understanding of their students' behavior so they can make the appropriate intervention when needed. The data used in this study were obtained from the Open University for two specific modules: AAA 2013 and 2014, as well as BBB 2013 and 2014. Our



(a) 2013J AAA (b) 2014J AAA

(c) 2013J BBB (d) 2014J BBB

Fig. 5. MSE curves

deep learning model has provided high results in terms of accuracy, sensitivity, specificity, and kappa analysis.

According to the current methodology, the number of clicks for each type of activity performed by the students is used to predict final student performance. However, there are a variety of other factors that should also be taken into consideration, such as the teaching style, the course design, and the experience of the teacher. Therefore, the next step is to take into consideration the type of student engagement. For instance, when the student is accessing resources, it is important to check what sort of resources the student is accessing and what is the exact file reached by the student. Future directions will involve developing a generic method of hybrid machine learning (supervised and unsupervised) that will be applied to several educational datasets to help develop new algorithm that can be applied to any other educational institution with the same level of efficiency.

References

[1] T. Treasure-Jones, C. Sarigianni, R. Maier, P. Santos, and R. Dewey, "Scaffolded contributions, active meetings and scaled engagement: How technology shapes informal learning practices in healthcare sme networks," *Computers in Human Behavior* **95** (2019) 1 – 13.

[2] F. Bonafini, C. Chae, E. Park, and K. Jablokow, "How much does student engagement with videos and forums in a mooc affect their achievement?" *Online Learning Journal* **21** no. 4, (Dec., 2017) 223–240.

[3] M. Khalaf, A. J. Hussain, D. Al-Jumeily, T. Baker, R. Keight, P. Lisboa, P. Fergus, and A. S. Al Kafri, "A data science methodology based on machine learning algorithms for flood severity prediction," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8. 2018.

[4] S. Ghareeb, A. Hussain, W. Khan, D. Al-Jumeily, T. Baker, and R. Al-Jumeily, "Dataset of student level prediction in uae," *Data in Brief* **35** (2021) 106908.

[5] L. C. Liñán and Ángel Alejandro Juan Pérez, "Educational data mining and learning analytics: differences, similarities, and time evolution," *International Journal of Educational Technology in Higher Education volume* **15** (July, 2015) 98–112.

[6] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomedical Engineering* **2** (Feb., 2018) 158–164.

[7] R. M. de Albuquerque, A. A. Bezerra, D. A. de Souza, L. B. P. do Nascimento, J. J. de Mesquita Sá, and J. C. do Nascimento, "Using neural networks to predict the future performance of students," in *2015 International Symposium on Computers in Education (SIIE)*, pp. 109–113. 2015.

[8] G. Al-Tameemi, J. Xue, S. Ajit, T. Kanakis, and I. Hadi, "Predictive learning analytics in higher education: Factors, methods and challenges," in *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, pp. 1–9. 2020.

[9] G. Akçapınar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *International Journal of Educational Technology in Higher Education volume* **16** (October, 2019) .

[10] G. Al-Tameemi and J. Xue, "Towards an intelligent system to improve student engagement and retention," *Procedia Computer Science* **151** (2019) 1120–1127. The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops.

[11] L. P. Khobragade and P. P. Mahadik, "Students' academic failure prediction using data mining," 2015.

[12] F. Ahmad, N. Ismail, and A. Aziz, "The prediction of students' academic performance using classification data mining techniques,".

[13] M. F. Sikder, M. J. Uddin, and S. Halder, "Predicting students yearly performance using neural network: A case study of bsmrstu," in *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 524–529. 2016.

[14] T. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang, "Behavior-based grade prediction for moocs via time series neural networks," *IEEE Journal of Selected Topics in Signal Processing* **11** no. 5, (2017) 716–728.

[15] W.-X. Liu and C.-H. Cheng, "A hybrid method based on mlfs approach to analyze students' academic achievement," pp. 1625–1630. 08, 2016.

[16] C. López, E. L. Guzman, and F. González, "Data mining model to predict academic performance at the universidad nacional de colombia," 2013.

[17] M. Hussain, W. Zhu, W. Zhang, and R. Abidi, "Student engagement predictions in an e-learning system and their impact on student course assessment scores," *Computational Intelligence and Neuroscience* **2018** (10, 2018) 1–21.

[18] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Scientific data* **4** no. 1, (2017) 1–8.

[19] S. Qiao, H. Wang, C. Liu, W. Shen, and A. Yuille, "Micro-Batch Training with Batch-Channel Normalization and Weight Standardization," *arXiv e-prints* (Mar., 2019) arXiv:1903.10520, `arXiv:1903.10520 [cs.CV]`.

[20] Y. You, D. Liang, R. Wei, M. Li, Y. Li, J. Wang, X. Wang, X. Zheng, W. Jia, and T. Chen, "Evaluation of metabolite-microbe correlation detection methods," *Analytical biochemistry* **567** (2019) 106–111.

[21] F. Okubo, T. Yamashita, A. Shimada, and H. Ogata, "A neural network approach for students' performance prediction," pp. 598–599. 03, 2017.

[22] S. Salman and X. Liu, "Overfitting Mechanism and Avoidance in Deep Neural Networks," *arXiv e-prints* (Jan., 2019) arXiv:1901.06566, `arXiv:1901.06566 [cs.LG]`.

[23] B. O. Ayinde, T. Inanc, and J. M. Zurada, "Regularizing deep neural networks by enhancing diversity in feature extraction," *IEEE Transactions on Neural Networks and Learning Systems* **30** no. 9, (2019) 2650–2661.

[24] A. Ashiquzzaman, A. K. Tushar, M. Islam, J.-M. Kim, *et al.*, "Reduction of overfitting in diabetes prediction using deep learning neural network," *arXiv preprint arXiv:1707.08386* (2017) .

[25] J. L. G. Balboa, M. V. A. Fernández, F. J. Ariza-López, and J. Rodríguez-Avi, "Homogeneity test for confusion matrices: A method and an example," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (2018) 1203–1205.