2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH) | 978-1-6654-0973-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMARTTECH54121.2022.00022

Predicting At-Risk Students Using the Deep Learning BLSTM Approach

Wiem Souai
Laboratory Hatem Bettaher (IRESCOMATH), University of Gabes, Gabes, Tunisia
wiem.souai@isimg.tn

Alaeddine Mihoub
Department of Management Information Systems and Production Management College of Business and Economics, Qassim University
P.O. Box: 6640, Buraidah: 51452, Saudi Arabia
a.mihoub@qu.edu.sa

Mounira Tarhouni
Laboratory Hatem Bettaher (IRESCOMATH), University of Gabes, Gabes, Tunisia
mounira.tarhouni@isimg.tn

Salah Zidi
Laboratory Hatem Bettaher (IRESCOMATH), University of Gabes, Gabes, Tunisia
salah_zidi@yahoo.fr

Moez Krichen
FCSIT, Albaha University, Albaha, Saudi Arabia
ReDCAD Laboratory, University of Sfax, Tunisia
moez.krichen@redcad.org

Sami Mahfoudhi
Department of Management Information Systems and Production Management College of Business and Economics, Qassim University
P.O. Box: 6640, Buraidah: 51452, Saudi Arabia
s.mahfoudhi@qu.edu.sa

ABSTRACT
Recently, the high usage of online learning platforms by schools and universities has been correlated with an increasing incompletion rate of online courses. Predicting students' academic performance helps the lecturer provide timely intervention and prevent dropping out of classes. This study focuses on applying Deep Learning algorithms to model the learning behaviors of students in a Virtual Learning Environment, predict their performance, and prevent students at-risk from failure. The proposed model is implemented using the Bidirectional Long-Short Term Memory algorithm (BLSTM). Applied to the Open University Learning Analytics Dataset (OULAD), the BLSTM model has achieved relevant results compared to previous approaches namely a cross-validation accuracy rate of 97%.

Keywords
Educational Data Mining, Predicting student performance, Virtual Learning Environment, Deep Learning, Bidirectional Long Short-Term Memory, BLSTM, OULAD.

1. INTRODUCTION
The recent advancements of Artificial Intelligence (AI) have made this technology present in several applications and domains [1]–[4], especially in the educational systems [5]. One of the principal aims of any pedagogical or educational platform is to provide students with the necessary information and skills to move into successful paths. Today, the effectiveness of global education systems in achieving this goal is a determining factor in economic and social change. In addition, a large number of students enter universities every year and it is becoming more and more challenging to provide all of them with good quality education and guidance. As a result, many students failed to graduate within the allotted time. To solve educational problems and challenges, researchers are investing more time to propose effective and powerful AI models that can be implemented in educational systems. In particular, the area of Educational Data Mining (EDM) [6], [7] has gained considerable attention from researchers interested in education technologies. One objective of Educational Data Mining is to allow teachers to better understand the barriers that face students in the learning process. In addition, it allows to model student data and predicts their future performance [8]. In fact, academic performance prediction for students has always been an important research topic in EDM. It relies essentially on Data Mining and Machine Learning techniques to explore and model rich data extracted from educational environments.

Exploring the EDM field, predicting students' academic outcomes represents a challenging task because of the multiple factors that impact their performance. However, the accelerated development of technology and digital resources has led to drastic shifts in education and especially, the widespread use of the so-called Virtual Learning Environments (VLE) [9]. VLE and related technologies have enabled researchers to collect vast amounts of data describing in detail the behavior of students. This large amount of available data has given the opportunity to better model the students' behaviors and attitudes and to develop more robust prediction algorithms.

In the special context of the COVID-19 outbreak [10], Virtual Learning Environments such as Massive Open Online Course (MOOC) [11], are becoming commonplace in student life. However, online schooling faces serious obstacles caused essentially by high dropout rates and extreme academic disappointment. Some studies estimated that the completion rate on MOOCs is typically less than 7% [12]. For instance, the Coursera platform completion rate is around 8% while the EDX's completion rate is near 5% [13]. Similarly, the dropout rate of the Open University in the UK is about 78% [14].

Unfortunately, these high withdrawal percentages cast doubt on the reliability of online educational platforms. Thus, estimating the students' future decisions and results related to the courses they are studying, allows teachers to carry out an instant intervention

978-1-6654-0973-5/22/$31.00 ©2022 IEEE
DOI 10.1109/SMARTTECH54121.2022.00022

that may motivate less performing learners to actively continue their learning paths.

Therefore, our goal in this paper is to help universities to design a learning experience to maximize student success. We aim also, via our modeling, to identify behaviors that maximize the opportunity for success and promote those behaviors with students. Concretely, our main goal is to identify students at risk based on current behaviors and general patterns with the aim of assisting those risky students, conducting proactive intervention, and increasing the opportunity of having better learning outcomes. Indeed, two main contributions are emphasized in this work:

- First, a feature engineering phase to optimize the machine learning inputs.
- Second, the application and the fine-tuning of a BLSTM (Bidirectional Long Short Term Memory) [15] model to get the best predictions for student performance.

The rest of the paper is organized as follows: Section 2 discusses related work on existing models for performance prediction. In Section 3, we present our methodology to model at-risk students and how to estimate their future performance. Section 4 describes the dataset used in our research and experiments' details. Results and discussion are presented in Section 5. The last section concludes the paper and discusses future work.

2. RELATED WORK

In the field of Educational Data Mining, substantial research has been undertaken to try to model a learner's comprehension of target skills. Generally, this is effectuated by the continuous assessment of students via their interaction and repetitive quizzes. Authors in [16], established new student assessment models based on behavioral characteristics extracted from the student interaction with the learning platform. They dealt basically with an online learning framework called Kalboard 360 using the Experience API Web Service (xAPI). Many Machine Learning techniques were applied such as Artificial Neural Networks (ANN), Decision Trees, Bagging, and Random Forests to measure the relevance of extracted behavioral features on student academic success. The outcome has demonstrated a significant relationship between the student's behavior on the platform and academic achievement. Some models such as ANN have achieved up to 29% improvement over classic models that do not include the extracted behavioral features [17].

Authors in [18] have studied success prediction models by investigating special features and their impact on academic outcomes such as students' absence and parents' interventions in the learning process. Three classic classifiers namely Naïve Bayes (NB), ANN, and Decision Trees (DT) were tested to analyze the effect of these characteristics on the students learning. As a result, the accuracy of the proposed model has been substantially increased by 10% to 15% compared to the same models without these investigated features i.e students' punctuality and parents' involvement.

With the same objective of predicting student performance, Ashfaq et al. [19] developed several models by applying many Data mining and Knowledge Discovery in Databases (KDD) techniques. Since imbalanced data were investigated, which may cause bias in predictions, they used several Data-balancing techniques, namely SMOTE (Synthetic Minority Over-sampling Technique) [20] and ADASYN (Adaptive Synthetic Sampling) [21], [22]. In addition, two features selection algorithms were tested to pick the most influencing features. These algorithms are the Fast Correlation Based Features (FCBF) [23] and the Recursive Feature Elimination (RFE) algorithm [24]. For the machine learning part, many classic models were applied and the best overall results were obtained by the Random Forest (RF) approach that has given an accuracy rate of 86%.

Deep learning methods were also applied in the field of learning outputs prediction. Aljohani et al. [25] have developed an LSTM (Long Short Term Memory) based model to estimate student performance over a certain period. Applied to the OULA dataset their best model has achieved nearly 95% for the accuracy rate and has improved the baseline model by nearly 18%.

In this paper, the main objective is to contribute to the limited amount of existing research on the application of deep learning to predict the learning outputs of students by implementing a model using the BLSTM approach (Bidirectional Long Short Term Memory). The choice of this model is motivated by its numerous advantages compared to other classic machine learning or deep learning approaches [15], [26]. More details on our methodology are given in the next section.

3. METHODOLOGY

3.1 Overview

Figure 1 illustrates our overall methodology. The original dataset was retrieved from the Open University Learning Analytics Dataset (OULAD) [27]. The OULAD dataset, certified by the Open Data Institute, contains data about students and their interactions with Virtual Learning Environment for seven selected courses. Briefly, the used dataset consists of seven tables connected using unique identifiers. From raw data to final results, the whole process is made of the following main phases:

- Data Cleaning: consists of correcting or deleting incorrect, corrupted, improperly formatted, duplicate, or incomplete data.
- Features Extraction: is the process of decreasing the amount of resources needed to describe a huge amount of data by choosing appropriate representative entities called features.
- Neural Network Model Building: consists of selecting the neural network's topology, which includes the number of hidden layers, nodes that make up each layer as well as the other neural network initial parameters.
- Hyperparameters Tuning: entails seeking the best set of hyperparameters to use in order to get the best model performance.
- Model Training: consists of reviewing a large number of samples to find the best values for all the weights and trainable hyperparameters of the network.
- Model Testing: consists of comparing the results generated by the obtained neural network against results from reference sets.

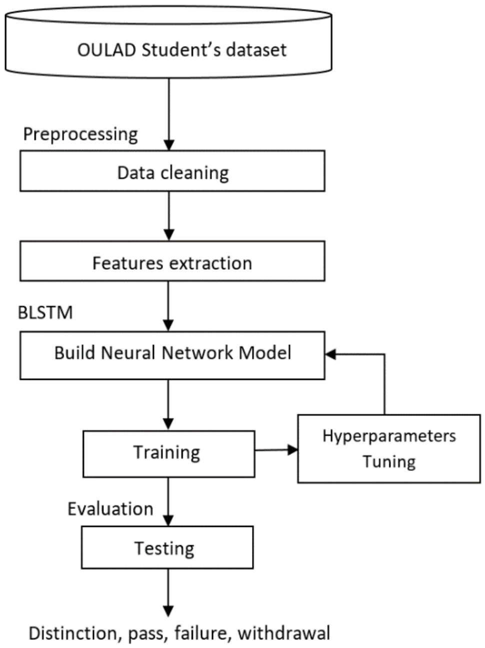

Figure 1: General approach

3.2 Bidirectional Long Short Term Memory

BLSTMs (Bidirectional Long Short Term Memory) [15], [26], [28] are a type of LSTM (Long Short Term Memory) that can increase model performance on sequences classification issues by extending traditional LSTMs. More precisely, a BLSTM is a sequence-processing model made up of two LSTMs, one of which takes the input in one way and the other in the opposite direction. BLSTMs effectively improve the quantity of data available to the network, providing the algorithm with more context and making the learning process faster.

The architecture of BLSTM is shown in Figure 2 where x_t is the input sequence at time t, y_t is the output sequence at time t, h_t is the hidden state at time t, and $\boldsymbol{\sigma}$ represents the activation function. For more details on the Bidirectional Long Short Term Memory please refer to [29].

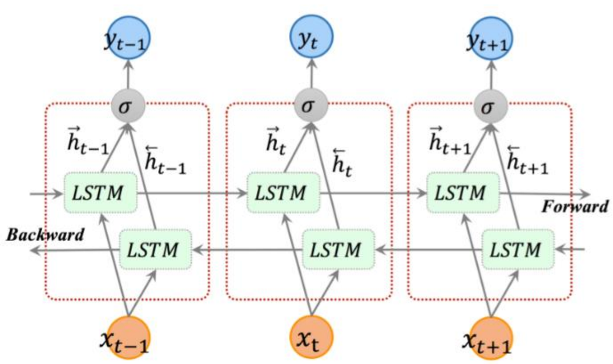

Figure 2. BLSTM Architecture (extracted from [29]).

3.3 Our model

The goal of our study is to see how effective BLSTM is in predicting students who are likely to fail and which factors have the most impact on their grades. More precisely, our aims are as follows:

- Gaining a better understanding of students' activities that put them at risk of failure. This way, we enrich decision-making processes that apply early intervention approaches to improve students' behaviors.
- Assessing the efficiency of the deployed deep BLSTM model with respect to existing models.

For the systematic assessment of students' academic performance, we target the prediction of course performance. In other words, our model strives to predict the final result of each student's course.

Unlike previous approaches [25], which use only two types of features (demographic features and behavioral features), we included in our modeling a third type of features which is the assessment features. For the assessment features, every assessment has a weight, each student has an assessment score in each period and datasets records show also whether the student submitted all assessments of the course or not. Therefore, these assessment features help to check not only the student's mid-term level but also procrastination and low motivation. Figure 3 represents the overall model. As shown in the figure, the model is composed of an input layer, the BLSTM layers, a dense layer to fully connect the previous outputs and, a softmax layer that is usually used in the multi-class classification problem.

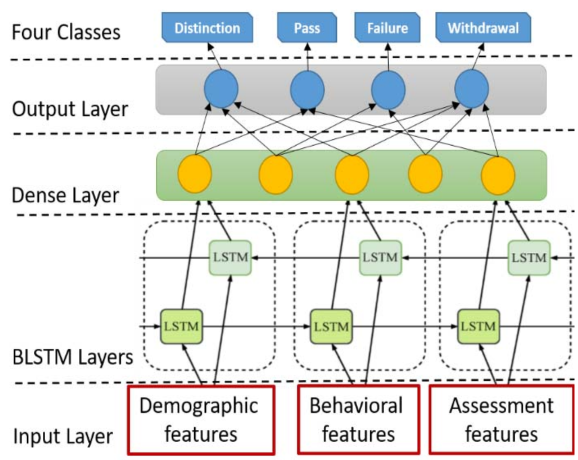

Figure 3. Proposed BLSTM model.

4. APPLICATION

4.1 Dataset

The Procured Open University Learning Analytics Dataset (OULAD) consists of seven tables [27]:

- Student Info Table: includes details about the students' demographics as well as their grades.
- StudentAssessment Table: includes the scores of the assessments completed by the students. There will be no result if the student does not complete the assessment.
- Assessments Table: includes details regarding assessments.

- StudentVLE Table: includes information about how each student interacted with the VLE's resources.
- Courses Table: includes the list of offered courses.
- StudentRegistration Table: includes details regarding the student's registration for the course presentation. The date of unregistration is also saved for students who dropped out.
- VLE Table: includes information on the VLE's available materials.

The seven tables are divided into three groups based on the information they provide, namely:

- Student demographics group: includes only the table StudentInfo.
- Student activities group: includes the tables StudentRegistration, StudentAssessment, and StudentVLE. This group provides details about 32,593 students. It covers the students' activities spread over a period of 9 months, between 2014 and 2015.
- Module presentation group: includes the tables Assessments, Courses, and VLE. It provides details about 22 module presentations freely available at the university. The data consists of many courses, and each course is taught at various intervals throughout the year.

Four different classifications levels are defined:

- Distinction: means that the student excelled in this course.
- Pass: means that the student has passed this course.
- Failure: means that the student failed this course.
- Withdrawal: means that the student has withdrawn this course.

Module presentation material is typically accessible in the Virtual Learning Environment few weeks before the official module begins. During the lecture, the student's expertise is assessed through a sequence of tests that identify the milestones in the module. Typically, there is a final test at the end. OULAD provides information on student contact with VLE, the storage of course lectures, also materials and appraisal information. Students interact with the Virtual Learning Environment to watch lectures, complete homework, read materials, and engage with each other. All their interaction with the VLE was registered and stored in log files [30].

4.2 Experiments

This section describes the experimental work where we evaluate the deployed deep BLSTM model. First, the considered datasets were merged using their unique IDs. The missing values of the dataset were ignored. In addition, one-hot encoder [31] was applied to the features. Moreover, since the dataset is imbalanced, we applied SMOTE oversampling in order to make the instances of classes equal.

Our BLSTM network is composed of 2 hidden layers containing each 50 units, trained on 100 epochs. The dropout of 0.5 was implemented to reduce overfitting. This also reduces interdependency between neurons, allowing the model to learn more effectively and rigorously [32].

The Softmax function [33] was applied as an activation function for the output layer. The ADAM optimizer [34], which is an optimization algorithm with an adaptive learning rate designed specifically for deep learning models, was applied in our experimentation too. In order to assess the efficiency of our model, in terms of difference between real and predicted values, we applied the categorical cross-entropy [35] as a loss function since it is well suited for classification with a number of classes bigger than two. The categorical cross-entropy is defined as follows:

$$Loss = -\sum_{i=1}^{output\_size} y_i * \log \hat{y}_i$$

where $\hat{y}_i$ is the i^{th} value of the model output, y_i is the corresponding actual value and the *output_size* is the number of output values of the studied model.

All experimentations were implemented using python and related data preprocessing, machine learning and deep learning packages (Pandas, Scikit-learn and Keras). Our result was initially based on the dataset splitting: 80% for training and 20% for testing. Moreover, we first trained our model without including the assessments scores and weights in order to estimate their impact on accuracy. In this case, our model reached an accuracy of 92.80%, which is relatively low. For that, we decided to train the model with the inclusion of assessments scores and weights. The obtained results are presented in the next section.

5. RESULTS AND DISCUSSION

Table 1 presents the evaluation of the BLSTM results compared to the results of a reference study [25]. The authors in that study applied the LSTM method using mainly the demographic and behavioral features, which we will call "set1" of features. In our study, we used what we call "set2" of features which adds the assessment features to "set1".

Table 1. Comparison of accuracy results

Approach	Algorithm	Features	Accuracy
Reference approach [25]	LSTM	Set1	95.23%
Initial approach	BLSTM	Set1	92.80%
Our approach	BLSTM	Set2	**96.90%**

As reported in the table, there is a clear improvement in the accuracy compared to the reference approach (96.90% vs. 95.23%). For the BLSTM model, the improved accuracy rate using "set2" of features (96.90% vs. 92.80%) confirms that assessments scores and their weights give a good indicator of students' future performance and their expected learning outputs. This may help instructors to deliver immediate actions to concerned students via the VLE platform. Figure 4 represents our approach accuracy over epochs. The accuracy tends to improve over epochs reaching 96.90% at epoch number 100. Similarly, Figure 5 represents the loss value over epochs. As shown in the plot, the loss starts at 1.33 and ends at nearly 0.14. Moreover, to ensure the efficiency of our model a second protocol is tested for our approach (see Table 2). The second protocol consists of using 5 cross-validation (5-CV) instead of the initial protocol that applied an 80%-20% splitting of the dataset. As shown in the table, the obtained accuracy of 97 % for the 5-CV protocol confirms the relevance of our proposed approach.

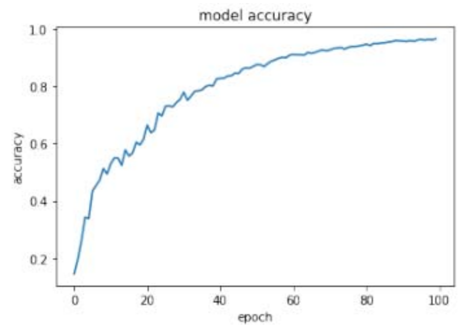

Figure 4. Accuracy rate over epochs

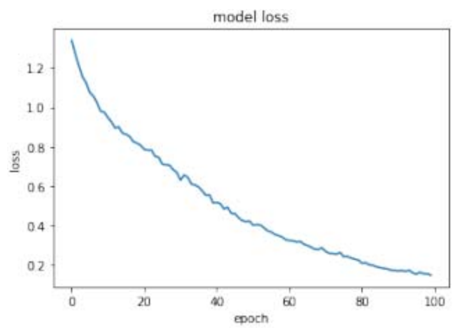

Figure 5. Loss value over epochs

Table 2. Comparison of accuracy results between two training/testing protocols

Approach	Protocol	Accuracy
Our approach	80%-20%	**96.90%**
Our approach	5-CV	**97.00%**

After predicting the students' performance and especially those who are at risk to fail, it is important to analyze the habits of the successful students. In this study, our analysis of student behavior shows that students that get high scores in short assessments during the module, interact frequently with the VLE, and show resilience in studying courses material, have a higher chance to pass their courses with "pass" or "distinction" academic output. On the opposite, the students who fail, tend to have lower scores in assessments and show poor interaction with the VLE such as a very low number of clicks. Those alarming habits should alert lecturers and motivate them to take timely preventive actions.

6. CONCLUSION

In a large-scale open education environment, timely prediction of students' academic performance is a necessary way to help lecturers provide effective intervention. Online learning platform collects fine-grained learning behavioral data, which provides an opportunity to study prediction models. However, the diversity of students and the sparseness of schools data brings challenges to improve forecasting performance. To ensure effectiveness through intervention, predictive models need to detect at-risk students in a precise manner and at early stages.

The demographic characteristics of students and the time-series logs extracted from VLE platforms are both useful sources of information used to know the high-risk learner. In this study, we proposed an innovative application of recent deep learning models by implementing a BLSTM model that allows predicting four classes of academic outputs: distinction, pass, failure, and withdrawal. Three types of inputs were used namely, demographic, behavioral, and assessments features. The proposed model has succeeded to enhance the accuracy rate (96.90%) compared to previous baseline models.

Moreover, our model shows high accuracy when including assessment weights and scores in the input features. This finding demonstrates the importance of the learning path of the student during the module in obtaining great academic results. In future work, we intent to study more deeply the factors impacting students' performance and we plan to investigate the best intelligent approaches to give automatic and adaptive feedback to risky students even without instructors' interventions.

REFERENCES

[1] A. Mihoub, 'A Deep Learning-Based Framework for Human Activity Recognition in Smart Homes', *Mobile Information Systems*, vol. 2021, Sep. 2021, doi: 10.1155/2021/6961343.

[2] A. Mihoub, O. B. Fredj, O. Cheikhrouhou, A. Derhab, and M. Krichen, 'Denial of service attack detection and mitigation for internet of things using looking-back-enabled machine learning techniques', *Computers & Electrical Engineering*, vol. 98, p. 107716, Mar. 2022, doi: 10.1016/j.compeleceng.2022.107716.

[3] O. Ben Fredj, A. Mihoub, M. Krichen, O. Cheikhrouhou, and A. Derhab, 'CyberSecurity Attack Prediction: A Deep Learning Approach', in *13th International Conference on Security of Information and Networks*, New York, NY, USA, Nov. 2020, pp. 1–6. doi: 10.1145/3433174.3433614.

[4] A. Mihoub, H. Snoun, M. Krichen, R. B. H. Salah, and M. Kahia, 'Predicting COVID-19 Spread Level using Socio-Economic Indicators and Machine Learning Techniques', in *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, Nov. 2020, pp. 128–133. doi: 10.1109/SMART-TECH49988.2020.00041.

[5] L. Chen, P. Chen, and Z. Lin, 'Artificial Intelligence in Education: A Review', *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[6] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, 'Educational data mining and learning analytics for 21st century higher education: A review and synthesis', *Telematics and Informatics*, vol. 37, pp. 13–49, Apr. 2019, doi: 10.1016/j.tele.2019.01.007.

[7] A. Dutt, M. A. Ismail, and T. Herawan, 'A Systematic Review on Educational Data Mining', *IEEE Access*, vol. 5, pp. 15991–16005, 2017, doi: 10.1109/ACCESS.2017.2654247.

[8] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, 'Analyzing undergraduate students' performance using educational data mining', *Computers & Education*, vol. 113, pp. 177–194, Oct. 2017, doi: 10.1016/j.compedu.2017.05.007.

[9] J. Barker and P. Gossman, 'The learning impact of a virtual learning environment: students' views', *Teacher Education Advancement Network Journal (TEAN)*, vol. 5, no. 2, Art. no. 2, Jul. 2013.

[10] C. Torres Martín, C. Acal, M. El Homrani, and Á. C. Mingorance Estrada, 'Impact on the Virtual Learning Environment Due to COVID-19', *Sustainability*, vol. 13, no. 2, Art. no. 2, Jan. 2021, doi: 10.3390/su13020582.

[11] B. Sathanarugsawait and C. Samat, 'Designing of Learning Environment Model to Enhance Student' Self Regulation by Using Massive Open Online Course', in *Innovative*

Technologies and Learning, Cham, 2021, pp. 542–547. doi: 10.1007/978-3-030-91540-7_55.
[12] K. Jordan, 'Initial trends in enrolment and completion of massive open online courses', *IRRODL*, vol. 15, no. 1, Jan. 2014, doi: 10.19173/irrodl.v15i1.1651.
[13] W. Feng, J. Tang, and T. X. Liu, 'Understanding Dropouts in MOOCs', *AAAI*, vol. 33, pp. 517–524, Jul. 2019, doi: 10.1609/aaai.v33i01.3301517.
[14] Mingjie Tan and Peiji Shao, 'Prediction of Student Dropout in E-Learning Program Through the Use of Machine Learning Method', *International Journal of Emerging Technologies in Learning*, vol. 10, no. 1, pp. 11–17, Mar. 2015, doi: 10.3991/ijet.v10i1.4189.
[15] D. Uliyan, A. S. Aljaloud, A. Alkhalil, H. S. A. Amer, M. A. E. A. Mohamed, and A. F. M. Alogali, 'Deep Learning Model to Predict Students Retention Using BLSTM and CRF', *IEEE Access*, vol. 9, pp. 135550–135558, 2021, doi: 10.1109/ACCESS.2021.3117117.
[16] E. A. Amrieh, T. Hamtini, and I. Aljarah, 'Mining Educational Data to Predict Student's academic Performance using Ensemble Methods', *IJDTA*, vol. 9, no. 8, pp. 119–136, Aug. 2016, doi: 10.14257/ijdta.2016.9.8.13.
[17] E. A. Amrieh, T. Hamtini, and I. Aljarah, 'Preprocessing and analyzing educational data set using X-API for improving student's performance', in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Nov. 2015, pp. 1–5. doi: 10.1109/AEECT.2015.7360581.
[18] S. B, I. F. Siddiqui, and Q. A. Arain, 'Analyzing Students' Academic Performance through Educational Data Mining', *3C Tecnología. Glosas de innovación aplicadas a la pyme*, pp. 402–421, May 2019.
[19] Usman Ashfaq, Booma P. M., and Raheem Mafas, 'Managing Student Performance: A Predictive Analytics using Imbalanced Data', *IJRTE*, vol. 8, no. 6, pp. 2277–2283, Mar. 2020, doi: 10.35940/ijrte.E7008.038620.
[20] M. G. Karthik and M. B. M. Krishnan, 'Hybrid random forest and synthetic minority over sampling technique for detecting internet of things attacks', *J Ambient Intell Human Comput*, Mar. 2021, doi: 10.1007/s12652-021-03082-3.
[21] A. Balaram and S. Vasundra, 'Prediction of software fault-prone classes using ensemble random forest with adaptive synthetic sampling algorithm', *Autom Softw Eng*, vol. 29, no. 1, p. 6, Nov. 2021, doi: 10.1007/s10515-021-00311-z.
[22] J.-R. Jiang and C.-L. Li, 'Binary- and Multi-class Network Intrusion Detection with Adaptive Synthetic Sampling and Deep Learning', in *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, Sep. 2021, pp. 1–2. doi: 10.1109/ICCE-TW52618.2021.9603206.
[23] C. Ying, G. Wang, and H. Li, 'Design of Feature Selection Algorithm Based on Improved FCBF', in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Apr. 2021, pp. 323–327. doi: 10.1109/ICSP51882.2021.9408729.
[24] Y. Lao, F. Qi, T. Yu, and X. Fang, 'An LSTM Model for Large-Scale Industrial Electricity Demand Forecast Based on RFE Feature Selection', in *Advances in Intelligent Automation and Soft Computing*, Cham, 2022, pp. 1139–1148. doi: 10.1007/978-3-030-81007-8_131.
[25] N. R. Aljohani, A. Fayoumi, and S.-U. Hassan, 'Predicting At-Risk Students Using Clickstream Data in the Virtual Learning Environment', *Sustainability*, vol. 11, no. 24, Art. no. 24, Jan. 2019, doi: 10.3390/su11247238.
[26] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, 'A Systematic Review of Deep Learning Approaches to Educational Data Mining', *Complexity*, vol. 2019, p. e1306039, May 2019, doi: 10.1155/2019/1306039.
[27] J. Kuzilek, M. Hlosta, and Z. Zdrahal, 'Open University Learning Analytics dataset', *Sci Data*, vol. 4, no. 1, p. 170171, Nov. 2017, doi: 10.1038/sdata.2017.171.
[28] N. Minh-Tuan and Y.-H. Kim, 'Bidirectional Long Short-Term Memory Neural Networks for Linear Sum Assignment Problems', *Applied Sciences*, vol. 9, no. 17, Art. no. 17, Jan. 2019, doi: 10.3390/app9173470.
[29] Z. Cui, R. Ke, Z. Pu, and Y. Wang, 'Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values', *Transportation Research Part C: Emerging Technologies*, vol. 118, p. 102674, Sep. 2020, doi: 10.1016/j.trc.2020.102674.
[30] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, 'Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores', *Computational Intelligence and Neuroscience*, vol. 2018, p. e6347186, Oct. 2018, doi: 10.1155/2018/6347186.
[31] Z. Lv, H. Ding, L. Wang, and Q. Zou, 'A Convolutional Neural Network Using Dinucleotide One-hot Encoder for identifying DNA N6-Methyladenine Sites in the Rice Genome', *Neurocomputing*, vol. 422, pp. 214–221, Jan. 2021, doi: 10.1016/j.neucom.2020.09.056.
[32] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, 'Student Academic Performance Prediction using Supervised Learning Techniques', *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 14, Art. no. 14, Jul. 2019, doi: 10.3991/ijet.v14i14.10310.
[33] D. Zhu, S. Lu, M. Wang, J. Lin, and Z. Wang, 'Efficient Precision-Adjustable Architecture for Softmax Function in Deep Learning', *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 12, pp. 3382–3386, Dec. 2020, doi: 10.1109/TCSII.2020.3002564.
[34] Z. Zhang, 'Improved Adam Optimizer for Deep Neural Networks', in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, Jun. 2018, pp. 1–2. doi: 10.1109/IWQoS.2018.8624183.
[35] P. Li *et al.*, 'Improved Categorical Cross-Entropy Loss for Training Deep Neural Networks with Noisy Labels', in *Pattern Recognition and Computer Vision*, Cham, 2021, pp. 78–89. doi: 10.1007/978-3-030-88013-2_7.