Research Article/Review Article

Journal of Computing, Science &Technology.Vol.1, Issue 2 xx, 2024

Journal of Computing, Science &Technology

https://focjournal.unidel.edu.ng/

editorjcst@unidel.edu.ng

ISSN:......Print

ISSN: .... Online

Unsupervised Deep Learning for Dropout Prediction in Higher Institutions of Learning

Ayokunle Olalekan Ige, Temitope Damilola Ariyo[2], B.O. Akingbesote[3]

[1,3]Department of Computer Science, Faculty of Science, Adekunle Ajasin University ,Akungba-Akoko, Ondo State, Nigeria.[2]Department of Educational Management, Faculty of Education, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria.

ayo.ige@aaua.edu.ng[1], ariyotemitope001@gmail.com[2], bakingbesote@gmail.com[3]

Corresponding Author's Email: ayo.ige@aaua.edu.ng

ABSTRACT

Article Info
Date Received: 02-02-2024
Date Accepted: 21-05-2024

Dropout occurs when students enrolled at educational institutions voluntarily abandon their studies, and it is a great challenge in higher institutions. Recent advancements in Artificial Intelligence have enabled researchers to develop traditional machine and deep learning models to predict students' tendency to drop out of higher institutions. However, the dataset of student dropouts is tedious to label, as it involves analyzing a large amount of information from various sources such as school records, surveys, and interviews, leading to high dimensionality data. Furthermore, the labeling process requires high accuracy to ensure that the data is reliable and valid. For this reason, it is crucial to propose unsupervised methods that can learn inherent patterns in student dropout data without labels to predict the possibility of students' dropout. Recently, one dimensional convolutional neural network (1D-CNN) has been applied to tabular data because they often contain relationships and patterns between neighboring features or columns. Based on this, we propose a deep learning autoencoder model designed using 1D-CNN to extract features for dropout prediction in higher institutions from unlabeled data. Extensive experiments and ablation studies on the publicly available SATDAP dataset showed that the proposed model achieved a prediction accuracy of 92.25% and an F1-score of 92.13%, outperforming the traditional machine learning (ML) methods, which used the supervised approach.

Keywords:
dropout, student, autoencoder, unsupervised, higher institution

1.0 INTRODUCTION

The education community is quite concerned about students' dropout rates. Dropout is among the most complicated and challenging problems worldwide faced by students and global institutions [1]. According to a report by the National Center for Education Statistics, NCES (2020), Universities lose billions of dollars yearly because over 40% of students pursuing bachelor's degrees do not finish their degrees within six (6) years. Also, on the students' part, school leavers are more likely to earn less than those who graduated [3]. Therefore, higher education administrators must propose strategies for forecasting, identifying, and retaining students who are most likely to drop out [4]. In the past years, Universities have focused on the design of retention campaigns to prevent student withdrawal. However, because dropout arises from different context-specific academic and nonacademic factors, retention campaigns might not suffice across all contexts [5].

Consequently, the effectiveness of any retention strategy in higher education institutions depends on the early identification of susceptible students who are prone to dropping out [6]. This would allow educational institutions to take prompt and proactive action. The "at- risk" students can be identified and given academic and administrative help to boost their likelihood of finishing the course [7]. In recent times, advancements in AI has enabled researchers to develop various predictive models capable of revealing several hidden patterns that can explain students' strengths and weaknesses [8]. Recently, several machine learning (ML) models such as Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Logistic Regression (LR), among many others, have been used to predict student dropout rates [9]. For example, Gray and Perkins[10] used the K-Nearest Neighbor (KNN) method to predict student dropout in the third week of the student's first semester and had a 97% success rate. However, the technique is challenging to apply to every institution, as the third week of the first semester is insufficient to justify based on non-demographic factors.

Likewise, using RF, Adnan et al.[8] developed a prediction model for at-risk students to drop out at various percentages of the course length. They obtained a prediction accuracy of 91%. Abdul Bujang[11] used SVM, RF, DT, and LR to develop a predictive analytic model based on previous academic performance of studies. However, their study was designed to focus on a single course as a dropout factor. Similarly, Amare and Simonova[12] considered DT, Naïve Bayes (NB), RF and Logistic RF algorithms in predicting dropout in higher institutions. The result showed that LR outperformed other algorithms by achieving a prediction accuracy of 94%. Four (4) ML algorithms were used in the work of Costa et al.[13] to identify students with a high likelihood of failing early on. The results showed that the

SVM was the most effective algorithm, with 83% accuracy. The study also showed that improving the performance of ML algorithms requires significant data pre-processing and manual feature extraction.

Even though student dropout data are often tabular, several researchers have proposed deep learning models as seen [14][15][16] and [17], among many others. This is because tabular data of student dropout often involves a high number of features, and deep learning models with multiple hidden layers are capable of handling high-dimensional data effectively. Generally, before quality features of student dropout data can be extracted using deep learning, the target class needs to be labelled. However, the dataset of student dropouts is tedious to label, as it involves analyzing a large amount of information from various sources such as school records, surveys, and interviews, leading to high dimensionality data [18]. Furthermore, the labelling process requires high accuracy to ensure that the data is reliable and valid. This means that the labeler must have a deep understanding of the context and be able to identify subtle details and patterns in the data. For this reason, researchers have proposed models that can leverage unlabeled data to predict student dropout from higher institutions.

Based on this, several researchers have proposed semi-supervised and unsupervised approaches for feature learning to predict dropout rates from partly unlabeled and fully unlabeled data. In unsupervised learning problems, the dependent variable is not present or considered [19]. The primary goal is to discover connections between the dropout data's instances [20]. For instance, using data-driven unsupervised clustering and network partitioning methods, Peach et al.,[21] identified students who were likely to drop out based on their academic performance. Also, to demonstrate the potential for building a dropout prediction model with few unlabeled data, [22] employed visualization and clustering methods. Likewise, Ding et al.,[18] used a modified autoencoder model to improve low-performing students' prediction accuracy in higher institutions. However, quality student dropout features have not been extracted from unlabeled dropout datasets using these models. This can be attributed to the fully connected layers often used in the existing autoencoder models designed for dropout predictions, since fully connected layers lack spatial awareness. Generally, the purpose of an autoencoder for unlabeled data is to learn compressed representation or encoding of the input data without requiring explicit labels by finding the common features from unlabeled data. Different from previous studies, our research incorporates 1D-CNN in a denoisingautoencoder model and to boost the representation of discriminative student dropout features. To the best of our knowledge, this work is the first to propose such architecture to extract features of student dropout in higher institutions. By doing this, higher institutions can leverage unlabeled data to predict students' dropout, saving the cost and time of data annotation. Specifically, our contribution can be summarized in three folds:

1. Firstly, an autoencoder model which uses 1D-CNN is proposed to extract features from unlabeled higher institution student dropout data.

Secondly, the unlabeled data is concatenated with noise to improve feature learning in the encoder block.

Lastly, extensive experiments and visualizations show that the proposed model outperformed traditional ML algorithms on the student dropout dataset.

The remainder of this paper is organized as follows: Section 2 presents the literature review on student dropout prediction, Section 3 presents the methodology of the proposed autoencoder model, Section 4 discusses the results of experiments, and Section 5 concludes.

2.0 RELATED WORKS

There has been extensive research done on student dropout rates in higher education. Dropout occurs when an individual enrolled at an educational institution voluntarily abandons studies [5]. Due to its detrimental effects on the overall welfare of the community, higher institutions, and students, dropout has been considered a challenge in private and public higher education institutions [23]. Hence, several works have been done to understand the factors affecting students in higher institutions, leading to dropout. Educational Data Mining (EDM) researchers have supported institutional interventions through student dropout prediction [24]. For instance, some researchers, as seen in [25], [26] and [27], among others, have looked at how demographic and other factors could affect students' dropout or good learning performance.

As discussed in the previous section, several traditional ML algorithms have been leveraged to predict student dropout from higher institutions. An example is in Gupta & Sabitha,[28], where the researchers used DT to identify the factors contributing to student dropout. The algorithm was used to determine the key aspects that would aid students and course creators in enhancing the course content, structure, and delivery. RF, KNN and NB were also applied to analyze students' in-course behavior. In Martins et al. [29], the authors focused on understanding the best approach to predict a student's dropout risk at the earliest stage of the student's academic path. They used RF, SVM, RusBoost algorithms and an easy ensemble model. Results showed that the best performance was achieved when SMOTE resampling technique was used and the model was trained using RF.

In [30], stacking ensemble technique was proposed to improve the prediction of student dropout in higher institutions using fewer features, and the results showed improved performance over single machine learning models. Even though ML algorithms have allowed the prediction of student dropout from higher institutions, a general limitation of traditional ML algorithms is the need for manual feature engineering, which requires domain knowledge. For this reason, deep learning (DL) models, which are capable of processing raw data directly and automatically learning features, have been proposed by researchers for dropout prediction in higher institutions. For example, Korösi and Farkas,[31] used a Recurrent Neural Network (RNN) trained on raw log student records to predict students' learning performance at the end of the course. The outcomes demonstrated that RNN dominated in offering higher performance compared to traditional approaches. Adnan et al.[8] used a Feed Forward DNN to predict the risk of students dropping out at different percentages of course length, and the result showed that the deep learning model outperformed the traditional ML models used for comparison.

Agrusti et al.,[32] employed actual data from about 6000 students to train a CNN model and predict whether the student would drop out. However, the model's performance was limited due to the architectural design of the CNN model. Mubarak et al.[33] proposed a hybrid model combining CNN and Long Short-Term Memory (ConvLSTM) to extract features from dropout data to predict whether a student will drop out. Likewise, [34] leveraged 1D-CNNLSTM with variational autoencoder oversampling to predict student dropout in higher institutions. However, these existing models only considered the supervised approach with labelled dropout data. Similarly, Zheng et al. [35], predicted students' dropout and leveraged CNN in obtaining local feature representations based on learners' behaviour data, and the self-attention mechanism was used to learn correlations between different feature representations. Also, a Bi-LSTM layer was used to obtain a time-series feature vector representation. Also, Fu et al. [36] proposed CLSA, a deep learning model which uses CNN to extract local features and builds feature relations using a kernel strategy, before feeding the high-dimensional vector generated by the CNN to a LSTM network to obtain a time-series incorporated vector representation. However, this approach did not consider the cases whereby students' dropout data is not labelled.

Generally, A practical prediction algorithm results in a high prediction accuracy of the students' dropout potential, allowing such students to be identified early, and proper assistance can be rendered to avert dropout [12][29]. In order to achieve these objectives, a large volume of student data must be analyzed and labelled. However, the dataset of student dropouts is tedious to label, as it involves analyzing a large amount of information from various sources such as school records, surveys, and interviews, leading to high dimensionality data [18]. Hence, recent works have focused on proposing unsupervised learning to capture the features from unlabeled datasets. For instance, Valles-Coral et al. [37] used density-based spatial clustering with noise (DBSCAN), Hierarchical DBSCAN and K-Means to group students based on their tendency to drop out of higher institutions. The result showed that HDBSCAN achieved the highest performance. Kuo et al[38] proposed a stacked denoising autoencoder to extract features from an unlabelled student dropout dataset. After using the features extracted to the latent space, the classification results showed that the model could extract features from the unlabelled dropout dataset. However, quality feature representation was not achieved. This can be attributed to the fully connected layers used in the autoencoder model, since they lack spatial awareness. To address the limitations of the existing dropout prediction models, our research incorporates 1D-CNN in a denoising autoencoder model for student dropout prediction in higher institutions. A more detailed discussion is presented in the following section.

3.0 METHODOLOGY

The proposed autoencoder model used 1D-CNN in the encoder and the decoder block. Generally, an autoencoder consists of the encoder and the decoder parts. The encoder maps the input data to a lower-dimensional representation, typically called the "Latent" space [39]. It compresses the input data into its smaller representation, such that the dimensionality of the latent space is typically much smaller than the dimensionality of the input data, forcing the autoencoder to capture the data's most important features and patterns. In the decoder, the encoded feature representation in the latent space is reconstructed to the original input data. The objective of the autoencoder is to minimize the difference between the original input and the reconstructed output. In the proposed autoencoder, the feature learning capability of the encoder block is improved

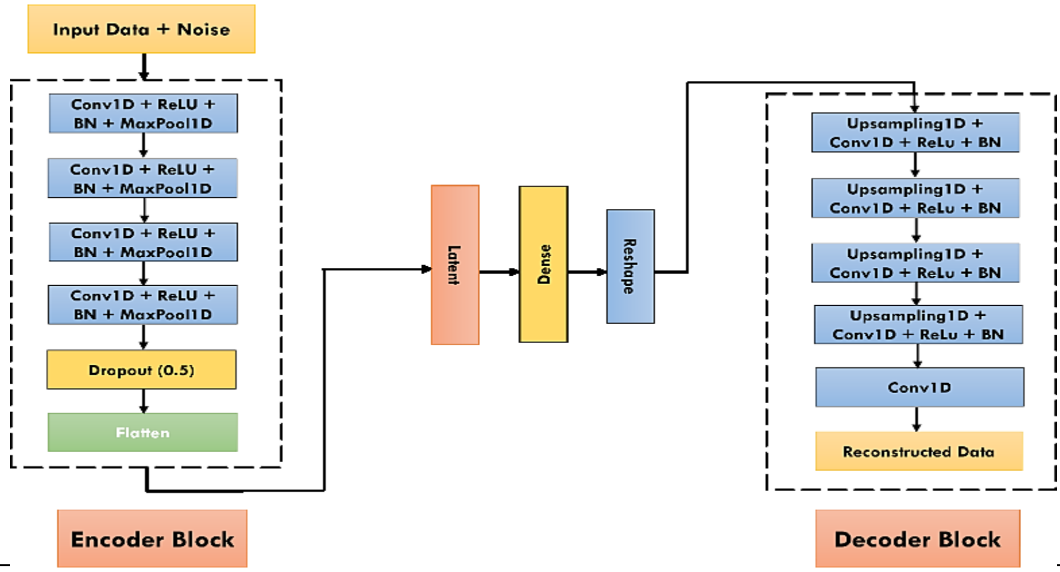

Figure 2. Architecture of the Proposed Autoencoder Model

by corrupting the input data with uniform noise (which is a general concept of denoising in autoencoders). Generally, to evaluate the performance of autoencoder models in extracting quality representative features from unlabelled datasets, a classification layer is added after the latent space after model training, and the features in the latent layer are used to classify the label hidden during training. The system design of the proposed model is presented in Figure 1.

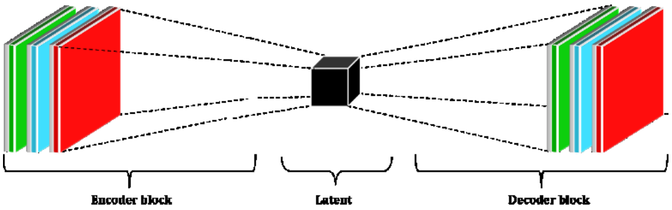

Figure 1. System design of the proposed model

3.1 Proposed Architecture

The architecture of the proposed model is presented in Figure 2. As shown, the encoder block takes in the data without labels as input, and uniform random noise ranging from 0 to 0.1 is concatenated with the input data, such that:

$$X\_train_{noisy} = X_{train} + noise \quad (1)$$

3.2 Encoder Block

Using the 1D-CNN, the channel dimensions were set to 128, 64, 32 and 16, with a kernel size of 3 used on all the convolutional layers in the encoder block. Each of the convolutional layers used ReLU activation function. Then, batch normalization (BN) was used to speed up the training process and ensure that gradient flow is optimized with faster convergence. A 1D-Maxpooling layer is then added after each convolutional layer in the encoder to downsample the input sequence. In the encoder, an FC layer with 128 neurons is used in the excite operation with a reduction ratio of 8. Before passing to the second FC layer with sigmoid AF. This allowed the encoder block to increase responsiveness to important features of the dropout data, based on learned features. The architectural summary of the encoder block is presented in Table 1.

Table 1. Architectural Summary of the Encoder Block

Layer	Configuration	Output
Input	-	32 × 1
1D Conv	Activation = ReLU, Kernel = 3	32 × 128
BN	-	32 × 128
1D MaxPool	Size = 2	16 × 128
1D Conv	Activation = ReLU, Kernel = 3	16 ×64
Batch Normalization	-	16 × 64
1D Maxpool	Size = 2	8 × 64
1D Conv	Activation = ReLU, Kernel = 3	8 × 32
Batch Normalization	-	8 × 32
1D Maxpool	Size = 2	4 × 32
1D Conv	Activation = ReLU, Kernel = 3	4 × 16
Batch Normalization	-	4 × 16
1D Maxpool	Size = 2	2 × 16
Dropout	Rate = 50%	2 × 16
Flatten	-	32
Latent	Activation = Linear	64
Dense	Activation = ReLU	32
Reshape	-	2 × 16

3.3 Decoder Block

In the decoder block, the channel dimensions of the encoder block were inverted to begin with 16, 32, 64 and 128, with a fixed kernel of size 3. The architectural summary of the decoder block is presented in Table 2.

Table 2. Architectural Summary of the Decoder Block

Layer	Settings	Output
1D Upsampling	Size = 2	4 × 16
1D Conv	Activation = ReLU, Kernel = 3	4 × 16
BN	-	4 × 16
1D Upsampling	Size = 2	8 × 16
1D Conv	Activation = ReLU, Kernel = 3	8 × 32
BN	-	8 × 32
1D Upsampling	Size = 2	16 × 32
1D Conv	Activation = ReLU, Kernel = 3	16 × 64
BN	-	16 × 64
1D Upsampling	Size = 2	32 × 64
1D Conv	Activation = ReLU, Kernel = 3	32 × 128
BN	-	32 × 128
1D Conv	-	32 × 1

3.4 Classification

At the top of the encoder network, after the latent layer, a fully connected layer with SoftMax activation function is included to classify the labels in the dataset, as shown in Fig. 3. The labels of the dataset were introduced to evaluate the quality of the extracted features stored in the latent layer.

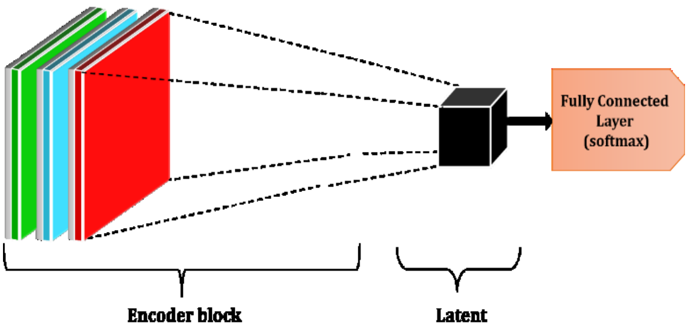

Figure 3: Prediction of Dropout with Features in the Latent space

These labels are denoted by L_n, where n is the number of labels in the dataset [40], and can be expressed as;

$$P(y = L_n|x_i; \theta) = \frac{e^{\theta_n^T x_i}}{\sum_{k=1}^{n} e^{\theta_k^T x_i}} \quad (2)$$

Where θ is the vector parameter of the network.

4.0 RESULTS

This section presents the results of the experiments conducted using the proposed autoencoder model. The various hyperparameters used in model training are presented, and the dataset description is presented.

4.1 Dataset Description

The benchmark dataset used for model evaluation is the student retention dataset, a publicly available dataset. The dataset has been used as the standard benchmark for higher education school dropout classification, as seen in [41] and [42], among many studies. The dataset consists of 4424 instances with 35 attributes related to undergraduate students, including demographics, social-economic factors, and academic performance. Attributes such as the student's marital status, course, age at enrollment, and others. The problem is formulated as a classification task consisting of three (3) classes (dropout, enrolled, and graduate) at the end of the normal duration of the course. The distribution of the classes in the dataset is presented in Table 3.

able 3: Distribution of categories in the proposed dataset.

SN	Category	Number of samples
1.	Graduate	2209
2.	Dropout	1421
3.	Enrolled	794

In this work, we experimented with the three classes (dropout, enrolled and graduate) and also on two classes (dropout and graduate) in order to limit the predictions to only students that graduated or dropped out.
The dataset is available online at https://archive- beta.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success.

4.2 Performance Evaluation Metrics

Since a classification layer is usually placed after the latent layer in an autoencoder model, accuracy, F1 score, and visualizations were considered in evaluating the performance of the proposed model. The visualization evaluation will show the pattern of the original input data against the data reconstructed by the proposed model.

4.3 Implementation Details

The proposed model is built using Keras with Python 3.9 on a workstation equipped with Core i7, 16GB RAM, and GeForce RTX 3050Ti with 4GB GPU. The details of the hyperparameters are presented in Table 4.

Table 4. Model Hyperparameters

Hyperparameter	Details
Optimizer	Adam
Epoch	100
Batch Size	16
Latent dimension	64
Learning rate	Initial Learning rate = $1e^{-4}$ Minimum Learning rate = $1e^{-7}$ Factor = 0.1 Patience = 5
Early stopping	Patience = 40 Monitor = Validation loss
Model loss	Mean Squared Error
Kernel size (Encoder and Decoder)	3

4.3 RESULT OF EXPERIMENTS

In order to evaluate the proposed model against Traditional ML algorithms, extensive experiments were done using the supervised method and baseline autoencoder. Also, the dataset consists of the "enrolled" class, which was dropped in another experiment. The performance of the proposed model was evaluated against SVM, Gaussian NB, DT, RF, and KNN on the two dataset cases. The results of the performance of these models are presented in Table 5.

ble 5. Comparison of Prediction Performance

Class	Model	Accuracy (%)	F1 Score (%)
Dropout, Graduate, Enrolled	SVM	72.88	70.97
	Gaussian NB	70.16	69.21
	DT	69.60	69.44
	RF	76.04	74.56
	KNN	67.57	65.61
	Proposed model	**77.40**	**76.51**
Dropout and Graduate	SVM	89.66	89.48
	Gaussian NB	85.53	85.30
	DT	84.84	84.82
	RF	89.25	89.11
	KNN	83.88	83.21
	Proposed model	**92.25**	**92.13**

When three classes of the dataset were considered (graduate, dropout and enrolled), the result presented in Table 5 shows that SVM had a prediction accuracy of 72.88% and an F1-score of 70.97%, which outperformed the NB, which achieved a 70.16% prediction accuracy and 69.21% F1-score. On the other hand, DT shows a prediction accuracy of 69.60% and 69.44% F1-score. The best performance among the traditional ML algorithms used RF, achieving prediction accuracy of 76.04% and an F1-score of 74.56%. However, KNN had the lowest prediction performance at 67.57% accuracy and 65.61% F1-score. However, the highest prediction performance was achieved by the proposed model, which recorded 77.40% accuracy and 76.51% F1-Score. Also, experiments on two classes (graduate and dropout) showed that SVM had 89.66% accuracy with

89.48% F1-score, NB had 85.53% accuracy, DT returned 84.64% accuracy, and RF had 89.25% accuracy. The lowest performance was achieved by KNN, with 83.88% accuracy and 83.21% F1-score. The 92.25% accuracy and 92.13% F1-score achieved by the proposed autoencoder model outperformed the traditional ML algorithms, even though they used the supervised learning method. The confusion matrix of the proposed model on the two dataset cases are presented in Figure 4 (a) and (b).

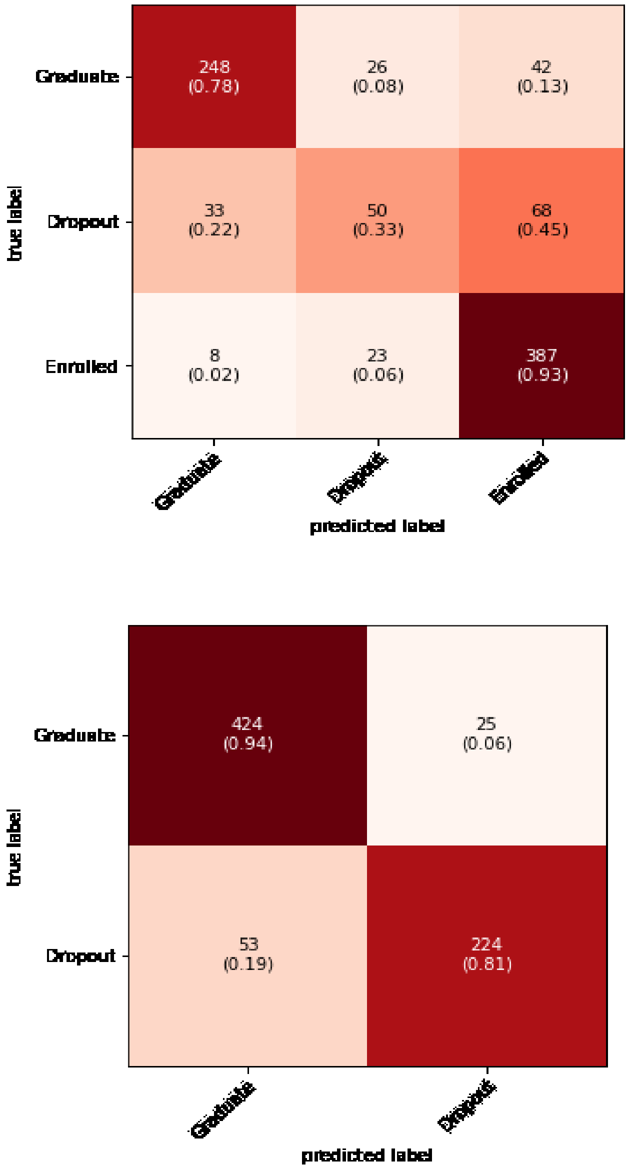

Figure 4. Confusion Matrix of the proposed model on (a) three classes (graduate, dropout, enrolled) (b) two classes (graduate and dropout)

4.3.1 VISUALIZATION RESULTS

The original training data and the data reconstructed by the proposed autoencoder model on the two dataset cases are presented in Figure 5 (a) and (b), (c) and (d), respectively.

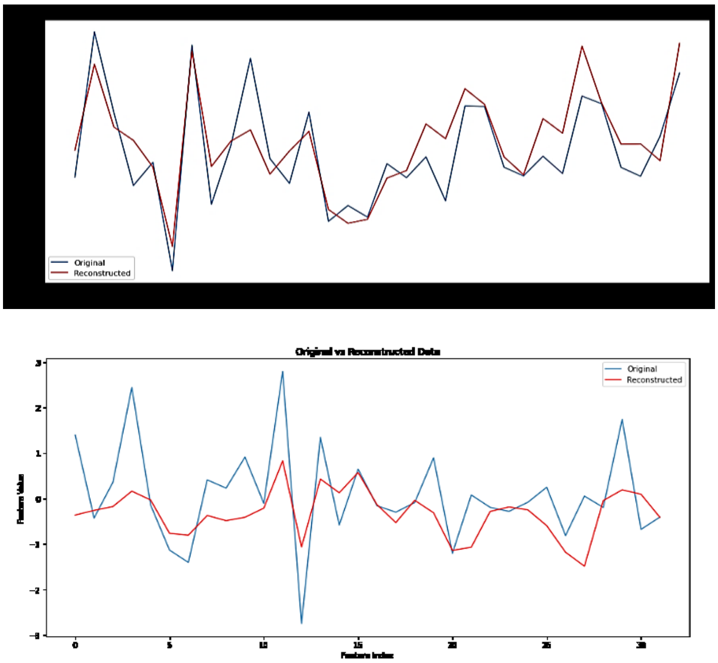

Figure 5. Visualization of reconstructed data using the Proposed Model

As shown in Figure 5, random samples were reconstructed, and the visualization result showed that the magnitude of the data reconstructed by the proposed model is closer to the original input data. Hence, the improved prediction performance.

4.4 Ablation Study

Experiments were conducted using varying latent dimensions of 16, 32, 64 and 128 on the dataset. Also, SMOTE oversampling of the dataset was done to address the imbalance class in the dataset. The result of the experiments is presented in Figure 6.

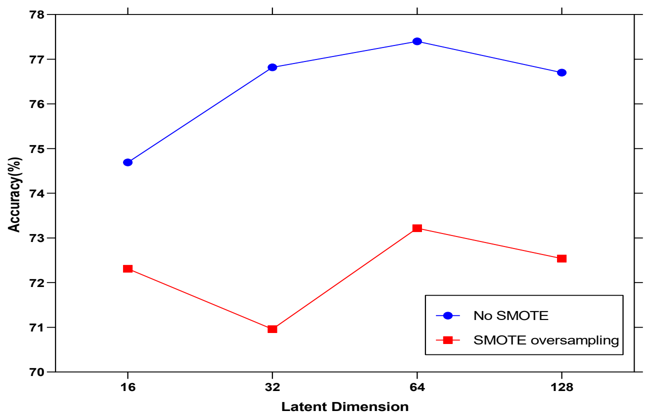

Figure 6. Comparison of Latent Dimension and SMOTE oversampling

As shown in Figure 6, the model's performance reduced when SMOTE oversampling was used to balance the dataset. Using 16 latent dimension, the proposed model achieved a prediction accuracy of 74.69%, while with

SMOTE, the performance was reduced to 72.31%. 32 latent size recorded a prediction accuracy of 75.82%, while with SMOTE, the performance degraded to 70.96%. The highest accuracy was recorded when a latent dimension of 64 was used in the encoder, achieving a prediction accuracy of 77.40%, compared to the 73.22% accuracy recorded when SMOTE was used on the same 64 latent dimension. Finally, 128 latent dimension had a prediction accuracy of 75.70%, compared to 72.54% achieved when SMOTE was used to oversample the dataset.

Experiments were carried out to investigate the effect of batch size on the proposed model on the two dataset cases. Batch sizes 8, 16, 32, 64, 128 and 256 were considered, and the result is presented in Figure 7.

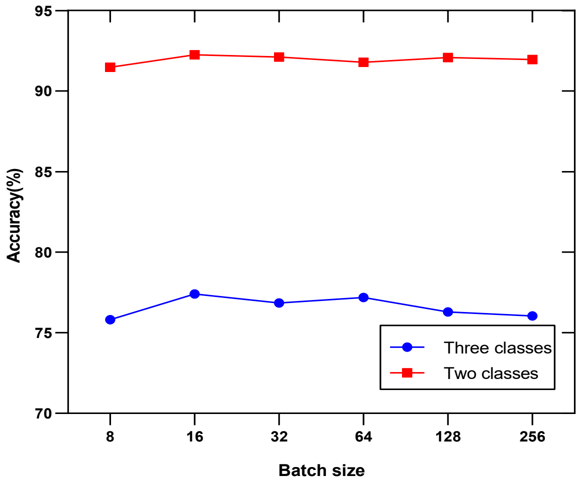

Figure 7. Effect of batch size

As shown in fig. 7, 8 batch size had an accuracy of 75.81% on three classes while 91.48% was recorded on two classes. The best performance was achieved when 16 batch sizes were used which achieved 77.40% accuracy on the three classes experiment and 92.25% on two classes experiments.

5.0 CONCLUSION

Several works have been done to understand the factors leading to dropout in higher institutions, and early prediction of dropout-prone students can help institutions to channel their intervention programs to such students. Several traditional ML and DL models have been used to predict student dropout from higher institutions' data. However, the dataset of student dropouts is tedious to label, as it involves analyzing a large amount of information from various sources such as school records, surveys, and interviews, leading to high dimensionality data. Furthermore, the labelling process requires high accuracy to ensure that the data is reliable and valid. This research addressed the challenges of labelling student dropout data by proposing an autoencoder model which used 1D-CNN to learn features from student dropout data. Experiments showed that the proposed model outperformed traditional ML models using the supervised approach. Also, the “enrolled” class in the benchmark dataset was dropped, to limit the prediction to graduate or dropout, and the result shows that the proposed model also outperformed traditional ML methods. By doing this, students with dropout tendencies can be identified quickly, and the institutions can administer the necessary support. Also, higher institutions will benefit by saving the cost and time needed to annotate datasets to detect the students that are likely to drop out. For future work, the proposed model will be applied to Massive Online Open Courses (MOOCs), and experiments will be conducted on datasets with sufficient features.

Statements and Declarations

Competing interests

There is neither conflict of interest nor any competing interest.

REFERENCES

[1] B. R. Cuji Chacha, W. L. Gavilanes López, V. X. Vicente Guerrero, and W. G. Villacis Villacis, “Student Dropout Model Based on Logistic Regression,” in *Applied Technologies*, 2020, pp. 321–333.

[2] NCES, “Fast Facts. Technical report, National Center for Education Statistics,” 2020. https://nces.ed.gov/fastfacts/display.asp?id=40 (accessed Apr. 25, 2023).

[3] M. N. Mustafa, L. Chowdhury, and M. S. Kamal, “Students dropout prediction for intelligent system from tertiary level in developing country,” *2012 Int. Conf. Informatics, Electron. Vision, ICIEV 2012*, no. October, pp. 113–118, 2012, doi: 10.1109/ICIEV.2012.6317441.

[4] T. N. M. Tran, “Skills and educational aspirations as predictors of secondary school dropout in Vietnam: A dynamic approach,” *Int. J. Educ. Dev.*, vol. 95, p. 102682, 2022, doi: https://doi.org/10.1016/j.ijedudev.2022.102682.

[5] D. Olaya, J. Vásquez, S. Maldonado, J. Miranda, and W. Verbeke, “Uplift Modeling for preventing student dropout in higher education,” *Decis. Support Syst.*, vol. 134, no. May, p. 113320, 2020, doi: 10.1016/j.dss.2020.113320.

[6] E. Murphy-Graham, D. Pacheco Montoya, A. K. Cohen, and E. Valencia Lopez, “Examining school dropout among rural youth in Honduras: Evidence from a mixed-methods longitudinal study,” *Int. J. Educ. Dev.*, vol. 82, p. 102329, 2021, doi: https://doi.org/10.1016/j.ijedudev.2020.102329.

[7] E. Masabo *et al.*, “Early detection of students at risk of poor performance in Rwanda higher education using machine learning techniques,” *Int. J. Inf. Technol.*, vol. 15, no. 6, pp. 3201–3210, 2023, doi: 10.1007/s41870-023-01334-3.

[8] M. Adnan *et al.*, “Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models,”

IEEE Access, vol. 9, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.

[9] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Electr. Eng.*, vol. 66, pp. 541–556, 2018, doi: https://doi.org/10.1016/j.compeleceng.2017.03.005.

[10] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Comput. Educ.*, vol. 131, pp. 22–32, 2019, doi: https://doi.org/10.1016/j.compedu.2018.12.006.

[11] S. D. Abdul Bujang, A. Selamat, and O. Krejcar, "A Predictive Analytics Model for Students Grade Prediction by Supervised Machine Learning," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1051, no. 1, p. 012005, 2021, doi: 10.1088/1757-899x/1051/1/012005.

[12] M. Y. Amare and S. Simonova, "Global challenges of students dropout: A prediction model development using machine learning algorithms on higher education datasets," *SHS Web Conf.*, vol. 129, p. 09001, 2021, doi: 10.1051/shsconf/202112909001.

[13] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Comput. Human Behav.*, vol. 73, pp. 247–256, 2017, doi: https://doi.org/10.1016/j.chb.2017.01.047.

[14] R. B. Basnet, C. Johnson, and T. Doleck, "Dropout prediction in Moocs using deep learning and machine learning," *Educ. Inf. Technol.*, vol. 27, no. 8, pp. 11499–11513, 2022, doi: 10.1007/s10639-022-11068-7.

[15] A. S. Imran, F. Dalipi, and Z. Kastrati, "Predicting Student Dropout in a MOOC: An Evaluation of a Deep Neural Network Model," in *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, 2019, pp. 190–195. doi: 10.1145/3330482.3330514.

[16] W. Xing and D. Du, "Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention," *J. Educ. Comput. Res.*, vol. 57, no. 3, pp. 547–570, 2019, doi: 10.1177/0735633118757015.

[17] M. Baranyi, M. Nagy, and R. Molontay, "Interpretable Deep Learning for University Dropout Prediction," in *Proceedings of the 21st Annual Conference on Information Technology Education*, 2020, pp. 13–19. doi: 10.1145/3368308.3415382.

[18] M. Ding, D. Y. Yeung, K. Yang, and T. C. Pong, "Effective feature learning with unsupervised learning for improving the predictive models in massive open online courses," in *ACM International Conference Proceeding Series*, 2019, pp. 135–144. doi: 10.1145/3303772.3303795.

[19] A. O. Ige and M. H. Mohd Noor, "A survey on unsupervised learning for wearable sensor-based activity recognition," *Appl. Soft Comput.*, p. 109363, 2022, doi: https://doi.org/10.1016/j.asoc.2022.109363.

[20] C. F. de Oliveira, S. R. Sobral, M. J. Ferreira, and F. Moreira, "How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review," *Big Data Cogn. Comput.*, vol. 5, no. 4, 2021, doi: 10.3390/bdcc5040064.

[21] R. L. Peach, S. N. Yaliraki, D. Lefevre, and M. Barahona, "Data-driven unsupervised clustering of online learner behaviour," *npj Sci. Learn.*, vol. 4, no. 1, 2019, doi: 10.1038/s41539-019-0054-0.

[22] L. M. Abu Zohair, "Prediction of Student's performance by modelling small dataset size," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, 2019, doi: 10.1186/s41239-019-0160-3.

[23] J. Pande, "Opportunities and challenges in the adoption of open educational resources for course development: a case study of Uttarakhand Open University," *Int. J. Inf. Technol.*, vol. 10, no. 3, pp. 339–347, 2018, doi: 10.1007/s41870-018-0126-z.

[24] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 905–971, 2023, doi: 10.1007/s10639-022-11152-y.

[25] L. Cen, D. Ruta, L. Powell, B. Hirsch, and J. Ng, "Quantitative approach to collaborative learning: Performance prediction, individual assessment, and group composition," *Int. J. Comput. Collab. Learn.*, vol. 11, pp. 187–225, 2016.

[26] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," *Comput. Human Behav.*, vol. 47, pp. 157–167, 2015, doi: https://doi.org/10.1016/j.chb.2014.05.038.

[27] D. A. Gutierrez-Pachas, G. Garcia-Zanabria, E. Cuadros-Vargas, G. Camara-Chavez, and E. Gomez-Nieto, "Supporting Decision-Making Process on Higher Education Dropout by Analyzing Academic, Socioeconomic, and Equity Factors through Machine Learning and Survival Analysis Methods in the Latin American Context," *Educ. Sci.*, vol. 13, no. 2, 2023, doi: 10.3390/educsci13020154.

[28] S. Gupta and A. S. Sabitha, "Deciphering the attributes of student retention in massive open online courses using data mining techniques," *Educ. Inf. Technol.*, vol. 24, no. 3, pp. 1973–1994, 2019, doi: 10.1007/s10639-018-9829-9.

[29] M. V Martins, L. Baptista, J. Machado, and V. Realinho, "Multi-Class Phased Prediction of Academic Performance and Dropout in Higher

Education," *Appl. Sci.*, vol. 13, no. 8, 2023, doi: 10.3390/app13084702.

[30] J. A. Talamás-Carvajal and H. G. Ceballos, "A stacking ensemble machine learning method for early identification of students at risk of dropout," *Educ. Inf. Technol.*, 2023, doi: 10.1007/s10639-023-11682-z.

[31] G. Korösi and R. Farkas, "MOOC Performance Prediction by Deep Learning from Raw Clickstream Data," in *Advances in Computing and Data Sciences*, 2020, pp. 474–485.

[32] F. Agrusti, M. Mezzini, and G. Bonavolontà, "Deep learning approach for predicting university dropout: A case study at Roma Tre University," *J. e-Learning Knowl. Soc.*, vol. 16, no. 1, pp. 44–54, 2020.

[33] A. A. Mubarak, H. Cao, and I. M. Hezam, "Deep analytic model for student dropout prediction in massive open online courses," *Comput. Electr. Eng.*, vol. 93, p. 107271, 2021, doi: https://doi.org/10.1016/j.compeleceng.2021.107271.

[34] E. C. Coppo, R. S. Caetano, L. M. de Lima, and R. A. Krohling, "Student Dropout Prediction using 1D CNN-LSTM with Variational Autoencoder Oversampling," in *2022 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2022, pp. 1–6.

[35] Y. Zheng, Z. Shao, M. Deng, Z. Gao, and Q. Fu, "MOOC dropout prediction using a fusion deep model based on behaviour features," *Comput. Electr. Eng.*, vol. 104, p. 108409, 2022, doi: https://doi.org/10.1016/j.compeleceng.2022.108409.

[36] Q. Fu, Z. Gao, J. Zhou, and Y. Zheng, "CLSA: A novel deep learning model for MOOC dropout prediction," *Comput. Electr. Eng.*, vol. 94, p. 107315, 2021, doi: https://doi.org/10.1016/j.compeleceng.2021.107315.

[37] M. A. Valles-Coral *et al.*, "Density-Based Unsupervised Learning Algorithm to Categorize College Students into Dropout Risk Levels," *Data*, vol. 7, no. 11, 2022, doi: 10.3390/data7110165.

[38] J. Y. Kuo, C. W. Pan, and B. Lei, "Using Stacked Denoising Autoencoder for the Student Dropout Prediction," *Proc. - 2017 IEEE Int. Symp. Multimedia, ISM 2017*, vol. 2017-Janua, pp. 483–488, 2017, doi: 10.1109/ISM.2017.96.

[39] A. O. Ige and M. H. Mohd Noor, "Unsupervised Feature Learning in Activity Recognition using Convolutional Denoising Autoencoders with Squeeze and Excitation Networks," in *ICOIACT 2022 - 5th International Conference on Information and Communications Technology: A New Way to Make AI Useful for Everyone in the New Normal Era, Proceeding*, 2022, pp. 435–440. doi: 10.1109/ICOIACT55506.2022.9972095.

[40] Fuqiang Gu *et al.*, "Locomotion Activity Recognition Using Stacked Denoising Autoencoders," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2085–2093, 2018, doi: 10.1109/JIOT.2018.2823084.

[41] M. V Martins, D. Tolledo, J. Machado, L. M. T. Baptista, and V. Realinho, "Early Prediction of student's Performance in Higher Education: A Case Study," in *Trends and Applications in Information Systems and Technologies*, 2021, pp. 166–175.

[42] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, "Predicting Student Dropout and Academic Success," *Data*, vol. 7, no. 11, 2022, doi: 0.3390/data7110146.