Effectiveness of data augmentation to predict students at risk using deep learning algorithms

Social Network Analysis and Mining (2023) 13:113
https://doi.org/10.1007/s13278-023-01117-5

ORIGINAL ARTICLE

Effectiveness of data augmentation to predict students at risk using deep learning algorithms

Kiran Fahd[1] · **Shah J. Miah**[1]

Received: 21 May 2022 / Revised: 18 August 2023 / Accepted: 19 August 2023 / Published online: 11 September 2023

Abstract
The academic intervention to predict at-risk higher education (HE) students requires effective data model development. Such data modelling projects in the HE context may have common issues related to (a) adopting small-scale modelling that gives limited options for early intervention and (b) using imbalanced data that hinders capturing effective details of poorly performing students. We address the issues going beyond the distribution-based algorithm, using a multilayer perceptron classifier which shows better on confusion metric, recall, and precision measures for identifying at-risk students. Our proposed deep learning-based model, which uses data augmentation techniques to supplement the data instances and balance the dataset, aims to improve the prediction accuracy of whether the student will fail or not based on their interaction with the learning management systems to prevent struggling students from evasion.

Keywords Deep learning · Data augmentation · Multilayer perceptron (MLP) · Deep forest (DF) · SMOTE · Distribution-based algorithm

1 Introduction

Higher education (HE) services are to enable the effective development of human skills that may contribute to the economy of a country. A 2017 study predicted that more than 60% of future jobs require a post-secondary degree for sustainable economic success (Hoffait and Schyns 2017). With the rapidly growing demands of tertiary education, communities are more seriously recently comprehended individuals for pursuing their tertiary degrees at a higher rate of success. This influx of students brings opportunities as well as challenges to achieve their HE aspirations. Therefore, for educational and economic reasons, student attrition is emerging as a global problem for the tertiary education sector. The attrition rate is the rate at which students discontinue their studies without graduating. It is also known as the student dropout rate. This has become a critical issue worldwide because studies indicated that one out of three students drops out of their HE programme (Heublein 2014; Hippel and Hofflinger 2020).

✉ Shah J. Miah
shah.miah@vu.edu.au; shah.miah@newcastle.edu.au

[1] Newcastle Business School, The University of Newcastle, Newcastle City Campus, Newcastle, NSW, Australia

This student attrition has grave repercussions for the individual student. The entire sector may suffer if we do not intervene appropriately (Berens et al. 2019) to minimize the student dropout rate. The impact of student dropout affects individuals financially and socially—leaving them heavily in debt and without better career opportunities (Allah 2020). The financial consequences for education providers are correspondingly heavy given the resources already invested in students who then drop out.

Educational institutions have employed numerous measures to reduce student attrition. However, it is challenging to address this issue adequately (Ahmad Tarmizi et al. 2019). HE institutions implement strategies focusing on improving internal factors such as the student experience, student engagement, and financial pressure, as well as assessing quality teaching practices to reduce student attrition rates (Canty et al. 2020; George et al. 2021). Several steps were taken to improve students' sense of belonging, providing flexible study modes (part-time, blended learning), enhanced guidance and pastoral care, flexible payment plans, better academic, and social integration to address the attrition issue (Munguia 2020; Shcheglova et al. 2020). Furthermore, different student attributes such as demographic, personal, social, and academic attributes have been analysed to identify the factors contributing to

student attrition. Relevant literature has identified student academic progress as one of the key determinants of student attrition (Beer and Lawson 2016).

An ability to identify students who are struggling or at risk by evaluating a student's academic performance may provide HE institutes with ways to offer various intervention programmes (Iqbal et al. 2019; Mngadi et al. 2020; Wakelam et al. 2020). These intervention programmes can lead to lower attrition rates by predicting student academic performance and taking measures to improve student progress (Imran et al. 2019). For instance, HE institutes can extend academic support to students through quality learning and teaching to enhance their academic performance. One approach to targeting this specialized support is to undertake student data mining. This can evaluate student academic performance; therefore, it may lead to identifying students at risk (Ajoodha et al. 2020).

Many researchers have explored various data mining techniques including statistical analysis and machine learning techniques to support student academic progress in an educational environment (Shingari et al. 2017; Sultana et al. 2019). Traditional statistical analysis approaches such as descriptive statistics, correlation analysis, or regression analysis have focused on analysing and summarizing data to predict student academic performance. However, these traditional statistical approaches have shown their limitations as they require humans to discover patterns or classify massive student datasets.

In the recent literature, machine learning (ML) has been used to analyse and detect patterns that can be transformed from educational data. These studies have mainly concentrated on the use of classification, regression, clustering, and association rules-based models (Aldowah et al. 2019) for developing data solutions. In recent years, the use of ML has increased in education to forecast student performance, improve assessment and feedback, recommend suitable resources, and identify struggling students. Deep learning (DL) models, a subset of ML, utilize neural networks for complex abstractions in datasets. In the majority of the earlier studies (Akour et al. 2020; Fok et al. 2018; Sultana et al. 2019; Sun et al. 2019; Tsiakmaki et al. 2020; Xing and Du 2018; Zhao 2017), DL algorithms were applied to the dataset of traditional student attributes such as semester marks and assessment submissions (academic features), or gender or parent education (personal socio-economical features). While data regarding student interaction with the learning management systems (LMS) during the semester are significantly valuable for capturing huge associative insights, previous studies are limited in applying DL for building data models. Therefore, it is imperative for us to examine the valuable interactions data as a source of providing earlier accurate identification on at-risk students.

The purpose of this study was to apply multiple DL algorithms on a publicly available dataset based on student interaction with LMS to precisely identify students at risk. Furthermore, existing studies (Munappy et al. 2019; Najafabadi et al. 2015; Shin et al. 2020) have mentioned that DL does not perform well on small-scale datasets. The publicly available dataset used in this study is small; therefore, this study also applied different augmentation techniques to achieve improved accuracy, as a part of the intervention. We developed a DL-based model, which uses data augmentation techniques to supplement the data instances and balance the dataset. Our DL model aims to improve the prediction accuracy of whether the student will fail or not based on their interaction with the LMS to prevent struggling students from evasion. The main objective of the study was to achieve the highest classification accuracy of the predictive model to identify students at risk by integrating existing data augmentation and balancing algorithms and DL algorithms.

The objective and major contribution of this study are summarized as follows:

- This study augmented the dataset by using data augmentation and balancing algorithms to increase the scale of the original dataset.
- A discussion of the experimental result of the evaluation comparison metric of the predictive model before and after approaching the small-scale dataset is provided in the paper.
- Access to the augmented dataset based on student interaction with LMS on an online repository for future use by researchers.
- This study aimed to assist educators and educational administrators in mitigating the ongoing and challenging issue of student attrition.

The rest of the paper is arranged as follows. Section 2 briefly explains the DL and implications of the small-scale dataset and introduces a few common methods to handle these implications. Section 3 presents the methodology of this study in detail. Section 4 represents the evaluation and discussion of the result of this study, and the conclusion is the final section of this paper.

2 Study background: deep learning and implications of small-scale dataset

Several studies in the literature have reported the application of DL techniques in HE to predict student academic progress (Doleck et al. 2019; Hernández-Blanco et al. 2019). This section briefly discusses these DL techniques, multiple DL architectures, characteristics of DL, and subsequent implications.

2.1 Deep learning (DL)

DL methods are based on neural networks that are set up with several layers of parameterized differentiable nonlinear nodes or modules. Each DL model consists of an input layer, one or more hidden layers according to the depth of the model, and an output layer as shown in Fig. 1 (Fonseca and Cabral 2019; Hosseini et al. 2020; Wlodarczak 2019).

In a supervised learning context, these layers can be trained by forward propagation or backpropagation (Hosseini et al. 2020; Wlodarczak 2019). In forward propagation, the input is mapped to the output layer in only one direction, whereas in backpropagation, the input layer is mapped to output in forward and backward networks. DL algorithms are employed in this study aimed at predicting student academic performance to mitigate the challenging issue of student attrition. DL algorithms are adopted in this study to process and analyse complex and enormous volume of data. The DL algorithms excel at discovering hidden patterns and interpreting data nonlinear dependencies (Katarya and Arora 2020; Kedia et al. 2021). In addition, DL-based models continuously improve their performance by learning from unseen data. These capabilities enable accurate and generalized predictive models that can be leveraged by educational institutions to identify students at risk of failing and enable timely interventions and support.

There are different neural network architectures. These include convolutional neural networks (CNN), recurrent or recursive neural networks (RNN), multilayer perceptrons (MLPs), deep neural networks (DNN), long short-term memory (LSTM), and pretrained unsupervised networks (PUNs) or autoencoders (Barari 2019; Fonseca and Cabral 2019; Hosseini et al. 2020; Wlodarczak 2019). These DL algorithms could be used for a range of data applications design in the educational sector (Issah et al. 2023; Katarya et al. 2021; Rahul and Katarya 2019; Rahul and Katarya 2023). For instance, CNN is used to train a student dropout rate prediction model (Sun et al. 2019); RNN is used in the task of identifying potential dropout students (Xing and Du 2018); the MLP technique is applied to student data in various studies to predict student performance (Lehr et al. 2016; Salal and Abdullaev 2020; Sultana et al. 2019). This evidence of proven cases makes the DL method a very promising approach (see Table 1 for more detailed cases), and the distribution of DL methods used in the reviewed paper to predict students' academic performance is given in Fig. 2.

As shown in Fig. 2, it has been stated in the literature that the DL techniques predominantly DNN (MLP) are frequently used in predicting the academic progress of the students.

Most of the above-mentioned studies utilized the dataset obtained from traditional classroom settings (face-to-face) or completely online courses (e.g. MOOC). Very few of the studies examine the datasets from blended learning to develop DL-based predictive model to predict academic performance. Furthermore, datasets used in these studies to develop the DL models are based on either socio-demographic features (e.g. socioeconomic status or parental education level) or academic features (e.g. assessment scores, pre-entry grades, and entry test scores,). In the existing literature, student participation and interaction with LMS in blended learning have not been much investigated. Also, not much LMS data are

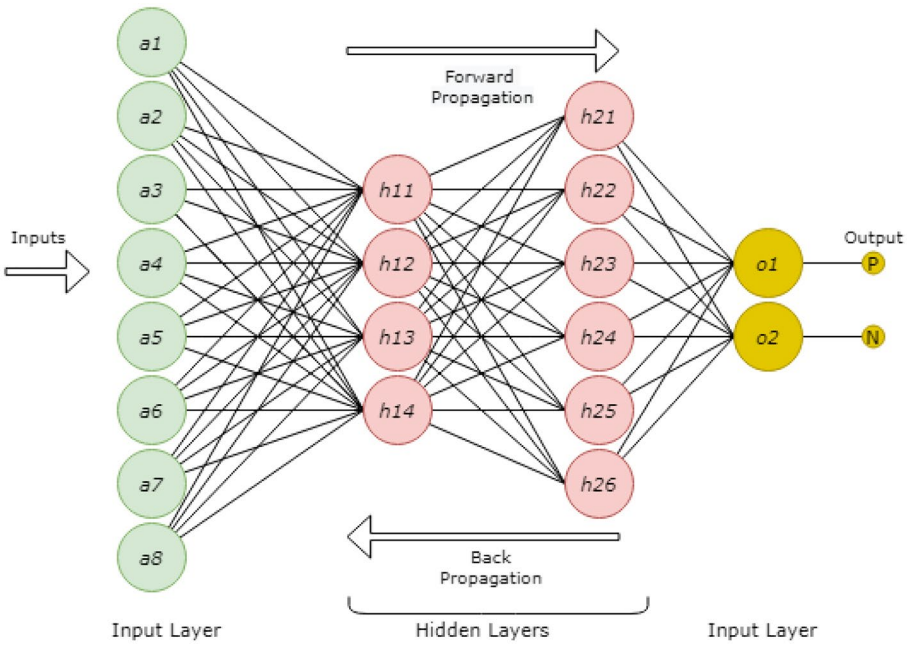


Fig. 1 Layers in deep neural network

Table 1 Existing application of DL in the education sector

Studies	Key topic areas	Key DL algorithms
Learning analytics using deep learning techniques for efficiently managing educational institutes (Veluri et al. 2022)	Developed a DL-based framework to predict student performance, behaviour, attitude, and their major selection. The dataset consists of student data such as gender, SAT scores, expenses, financial aid, and academic emphasis	ANN
DL-based Ensemble Approach to Predict Student Academic Performance: Case Study (Salal and Abdullaev 2020)	Developed a DL model using meta-ensembles and meta-classifier to predict student collective performance. The attributes of the student dataset contain information about their socioeconomic data, grades, and demographic data	MLP
A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC (Monllaó Olivé et al. 2020)	Developed a DL-based predictive model using a neural network to predict students at risk of dropping out of MOOC courses. The data are collected through Moodle such as forum posts, ratings, and details of quizzes attempted or failed	ANN
Student Performance Prediction using DL and Data Mining methods (Sultana et al. 2019)	Developed a DL model using DNN algorithms and data mining techniques to discover the academic performance of students. Student attendance, assessment marks, and LMS generated data like the count of resources visited and announcement viewed data are utilized in the study	MLP
DL for Dropout Prediction in MOOCs (Sun et al. 2019)	Developed a DL-based predictive model to predict students at risk of dropping out of a MOOC course. The study has utilized MOOC platform-generated student data such as time spent online, duration of watching a video on LMS, and count of forums visited	CNN
Dropout Prediction in MOOCs: Using DL for Personalized Intervention (Xing and Du 2018)	Developed a DL-based dropout prediction to accurately identify student dropout probability and offer personalized and prioritized intervention strategies. The dataset generated by the MOOC platform about student behaviour is leveraged to train and test the DL model, e.g. the count of announcement views, the count of quiz or course access, or the count of assessment submissions	RNN
Use Educational Data Mining to Predict Undergraduate Retention (Lehr et al. 2016)	Developed a DL-based model to predict undergraduate retention and provide timely decision recommendations to improve student retention. The dataset contains students' academic data such as their grades, GPA, and major	MLP and KNN

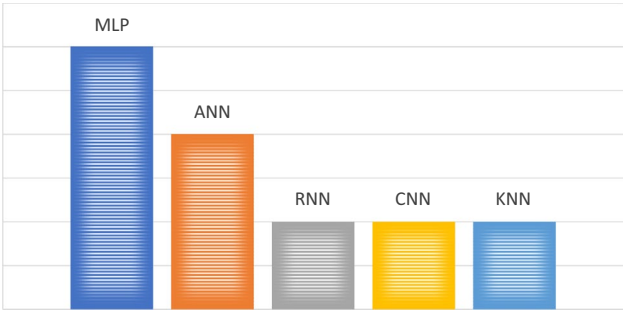


Fig. 2 Distribution of DL algorithms used in the reviewed studies from the existing literature

available to study the academic progress of the student which provides complete information about the student's actions during the workshops.

The application of DL techniques poses challenges such as DL being highly dependent on the dataset. It requires a large labelled dataset and extensive computing power for successful application (Najafabadi et al. 2015; Zhou and Feng 2017). DL algorithms did not demonstrate reasonable prediction accuracy or even failed when applied to a small-scale dataset (Najafabadi et al. 2015). Educational datasets are not always correctly labelled and are therefore imbalanced.

2.2 Dealing with small-scale datasets and imbalance dataset problems

The application of DL techniques to student data has been studied on a broader scale, but only a few studies have focused on the imbalanced data issue. For a new study, it is important to capture the basic core issues of dealing with small-scale datasets and techniques which can be used to address these issues. Data imbalance creates optimization difficulties to reduce errors and loss when fitting an ML model, or poor model generalization to demonstrate how accurately the model works for unseen data. Also, DL is based on a neural network with a large number of nodes with several layers. Thus, when using DL, more parameters must be estimated and require a larger amount of data in comparison with ML.

A DL model trained with a small-scale dataset is more likely to overfit with a lack of generalization—their accuracy is limited by the scant data, and they do not perform accurately against unseen or testing data. The imbalance dataset refers to the imbalance ratio of the classes of categorical variables and the unequal distribution of instances among different classes. Imbalanced datasets are those in which the label distribution of the classes is not the same for all classes, i.e. the number of instances of the majority class is higher than the minority class. Such a biased dataset influences the performance of the models. The minority class is often ignored, and this increases the probability of incorrect prediction of the minority class (Gupta et al. 2021). The imbalance ratio is expressed as the ratio of majority class instances to minority class instances. Furthermore, overlapping of data instances of majority and minority classes also impacts the model performance (Fatima et al. 2021).

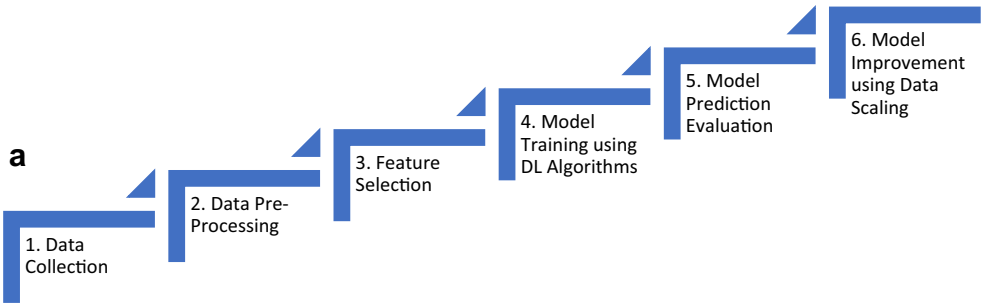
Ensemble methods such as boosting and bagging for the prediction can be used to reduce the influence of data imbalance or small datasets on the model. However, at the data level, modifying the class distribution by using the data augmentation technique is one way to address the dataset imbalance problem. Data augmentation expands the data observations by adding synthesized instances from the existing data. The inclusion of these synthesized data items addresses the small-scale dataset problem as well as the imbalanced dataset problem. In general, sampling methods, at the classifier level, are the most feasible option considered. However, in this study, we considered data augmentation algorithms that address data imbalance and scale problems.

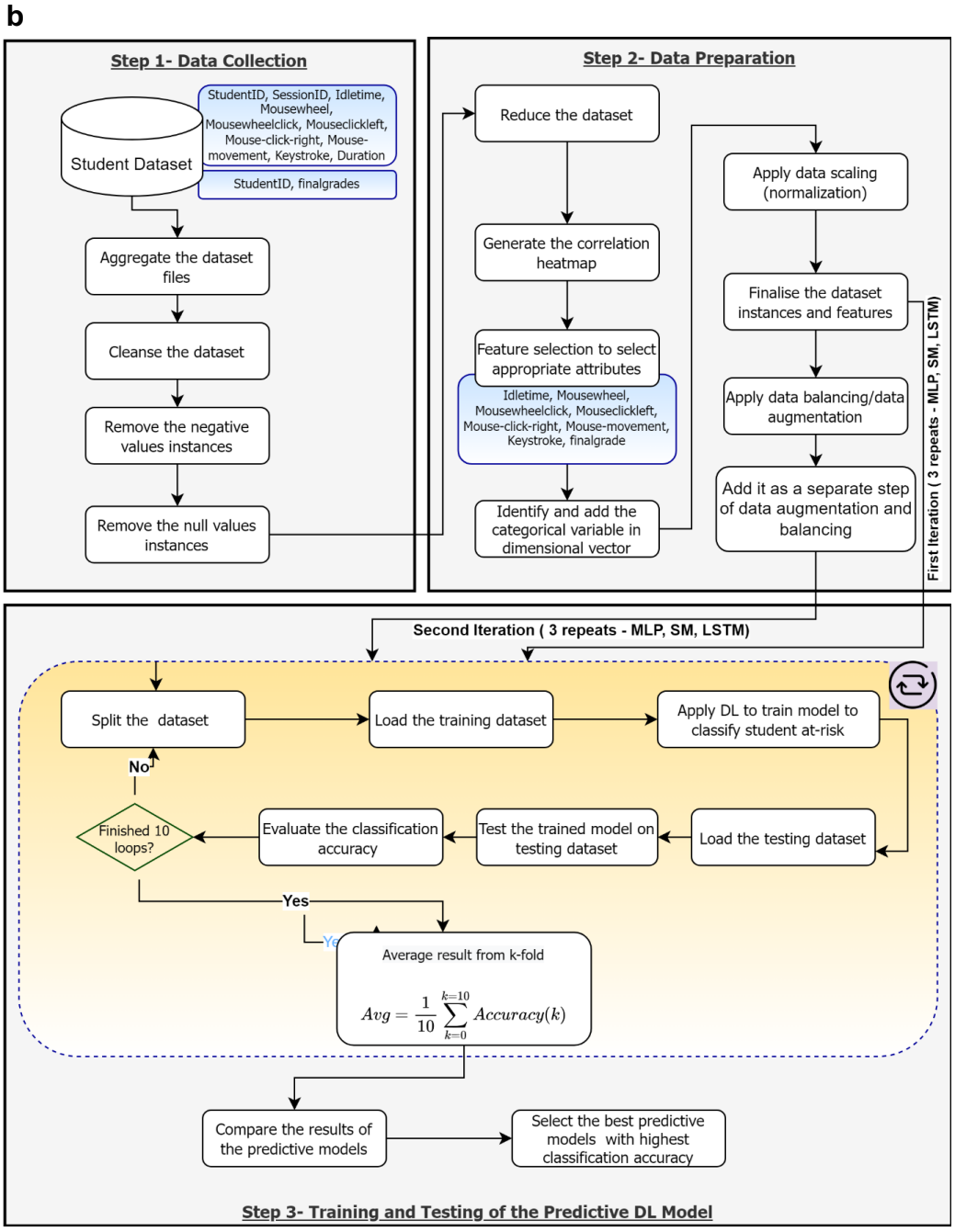
Random oversampling, a naive resampling technique, is a data augmentation technique that randomly resamples the minority class of the dataset and, based on the nature of the dataset, addresses the imbalanced dataset issue. In oversampling techniques like this, the instances of the minority class are duplicated in the dataset which can lead to overfitting. Arriving at the right threshold to avoid excessive overfitting and prevent information loss is difficult. The performance of the classifier is affected due to the bias introduced towards predicting the oversampled class. Under-sampling is another naive resampling technique that is used in combination with oversampling to balance the dataset.

The synthetic minority oversampling technique (SMOTE) is a resampling technique for numeric data introduced by a previous study (Chawla et al. 2002). SMOTE creates new synthesized instances from the existing minority class instances instead of merely replicating instances. It finds the *k*-nearest neighbour of the minority class and then generates new data points for this class along the line segment joining randomly selected *k*-nearest neighbours of the minority class instance. However, it does not consider that the neighbouring data points can be from the majority class, which can add overlapping data points and noise.

A distribution-based algorithm (DBA) (Bermejo et al. 2011) performs multiple tasks on the existing dataset to generate artificial instances. This algorithm resamples the dataset by replacing the class instances as follows and generates the same number of synthesized data points for each class. This algorithm allows using any of the four probability distributions, i.e. Uniform, Poisson, Multinomial, and Gaussian.

- Oversamples the minority class.


a
1. Data Collection
2. Data Pre-Processing
3. Feature Selection
4. Model Training using DL Algorithms
5. Model Prediction Evaluation
6. Model Improvement using Data Scaling


b
Step 1- Data Collection
Student Dataset
StudentID, SessionID, Idletime, Mousewheel, Mousewheelclick, Mouseclickleft, Mouse-click-right, Mouse-movement, Keystroke, Duration
StudentID, finalgrades
Aggregate the dataset files
Cleanse the dataset
Remove the negative values instances
Remove the null values instances
Step 2- Data Preparation
Reduce the dataset
Generate the correlation heatmap
Feature selection to select appropriate attributes
Idletime, Mousewheel, Mousewheelclick, Mouseclickleft, Mouse-click-right, Mouse-movement, Keystroke, finalgrade
Identify and add the categorical variable in dimensional vector
Apply data scaling (normalization)
Finalise the dataset instances and features
Apply data balancing/data augmentation
Add it as a separate step of data augmentation and balancing
First Iteration (3 repeats - MLP, SM, LSTM)
Second Iteration (3 repeats - MLP, SM, LSTM)
Split the dataset
Load the training dataset
Apply DL to train model to classify student at-risk
Load the testing dataset
Test the trained model on testing dataset
Evaluate the classification accuracy
Finished 10 loops?
No
Yes
Average result from k-fold
$Avg = \frac{1}{10}\sum_{k=0}^{k=10} Accuracy(k)$
Compare the results of the predictive models
Select the best predictive models with highest classification accuracy
Step 3- Training and Testing of the Predictive DL Model

◂**Fig. 3** **A** A holistic overview of the flowchart of the proposed predictive model. **B** Workflow of the proposed methodology using DL and data augmentation techniques on a small and imbalanced dataset to predict student academic progress. *MLP* Multilayer perceptron, *LSTM* long short-term memory, and *SM* sequential model. Dataset dimensions are listed in "Appendix 1"

- Undersamples the majority class.
- Minimizes the overlapping by removing or reducing it.
- Fully balances majority and minority classes.

SMOTE is considered a better approach than the other augmentation techniques for handling imbalanced datasets. It has demonstrated its effectiveness over an extended period with variable degrees of imbalance ratio and in mitigating model overfitting (Abid et al. 2022; Reshi et al. 2021; Rupapara et al. 2021). In this study, a DL-based predictive model is developed which compares the classification accuracy of the prediction of students at risk from sample data with and without dealing with class imbalance by using the SMOTE and DBA data augmentation technique. Three DL classifiers are selected to demonstrate the effectiveness of modifying the class distribution and scale of the dataset, i.e. MLP, LSTM, and sequential model (SM). MLP is a type of neural network that consists of input, densely connected layers called hidden layers, and output. SM builds a linear composition of the neural network model by creating a stack of layers. LSTM is a special type of recurrent neural network based on the long short-term memory approach. Each step has the option of three gates, which control the information flow: input, forget, and output. At each step, the input gate determines whether to forget or to learn (write into the memory) the new information from the input sources.

3 Study methodology

The research method of this study consists of data cleansing and transformation, data preparation involving data augmentation, and the application of different DL models to train and test the predictive models to identify students at risk. Figure 3A summarizes the flow of development of the DL model, whereas Fig. 3B depicts the complete and detailed workflow of the proposed work to demonstrate the methodology step-by-step. This method is part of a series of rigorous and iterative phases to develop a DL-based predictive model as an innovative educational artefact to predict student academic performance (Fahd et al. 2021a, b).

3.1 Data transformation

The dataset was downloaded from the University of California, Irvine ML repository (Vahdat et al. 2015). The dataset consisted of 230,318 data points collected from activities and interactions of 112 students which the LMS recorded in multiple comma-separated value (csv) files. These files contain numerical features, i.e. StudentID, SessionID, Exercise, Activity, Start-time, End-time, Idle-time, Mouse-click-left, Mouse-click-right, Mouse-wheel-click, Mouse-wheel, Mouse-movement, Keystroke, and final marks. A summary of dataset dimensions utilized in this study may be perused as supplementary material at: https://github.com/KFVU/DL-C/blob/main/1.pdf. The dataset files are aggregated into one file to be processed in the next step. A few basic statistical features of the dataset such as percentile, mean, and std are given in Table 2. Python as a platform was used to clean and transform the required data points from the original dataset by removing the instances that have negative values, null values, or missing information.

3.2 Data preparation

3.2.1 Correlation analysis

After data transformation, the dataset was reduced into 93 instances by aggregating all instances of each student from multiple sessions and merging them with their total final marks. The correlation analysis was performed on the reduced dataset, depicted by the heatmap in Fig. 4. Python Seaborn, which is a Matplotlib-based data visualization library, was used to create the heatmap for correlation analysis. It was used to identify significant connections such as whether a relationship exists between two numeric numbers x and y, i.e. for our dataset, we are exploring the correlation between each feature and the final marks of the student. The Pearson product-moment coefficient is used to spot the relationships between the features x and y. It is the fraction of the covariance of x and y and the product of the standard deviation of x and y. The mathematical expression is:

$$\frac{\sum (x - \text{mean}(x))(y - \text{mean}(y)))}{\sqrt{\sum \left(x - \text{mean}(x)^2\right)}\sqrt{\sum \left(y - \text{mean}(y)^2\right)}}.$$

The Pearson coefficient ranged between − 1 and 1. A coefficient value greater than 0 shows a positive correlation, and a value less than 0 shows a negative correlation. For example, our dataset demonstrates a positive correlation between keystrokes and final marks, and a negative correlation between idle time and final marks, i.e. the higher the number of keystrokes or the lower the idle time, means the probability of a higher final score is greater.

3.2.2 Data scaling: normalization

In this step, features were selected based on the correlation analysis, and the final total was transformed into a categorical variable with Pass (*P*) and Fail (*N*) classes. The dataset population comprised 61% of the majority class (*P*) and 39% of the minority class (*N*). The final features of the dataset consist of one categorical variable with two classes (*P* and *N*) and eight numerical features.

The selected eight features span varying degrees of magnitude, range, and units. Most of the features consist of quantities and units of some features are in seconds. This difference in ranges of values of features causes different step sizes of gradient descent for each feature. DL techniques use gradient descent as an optimization technique that requires data to be scaled. Dataset scaling helps to use similar steps for gradient descent. Normalization (ranging between 0 and 1) and standardization (centred around the mean (= 0) with a unit standard deviation—ranging between − 1 and 1) are two of the most used feature scaling techniques.

We have used normalization, which is one of the most used scaling techniques, to scale the dataset. The Python minmaxscalar class (based on Min–Max scaling) from the Scikit-learn library was used to transform each feature of the dataset into a range of 0 and 1. The formula for normalization scaling is:

$$x' = \frac{x - x_{\min}}{x - x_{\max}}.$$

Table 2 presents the descriptive statistics of the dataset population after normalization, including central tendency summary and dispersion.

3.3 Data augmentation and balancing

Our dataset consisted of only 93 instances and was an imbalanced dataset. The ratio of the majority class (*P*) to the minority class (*N*) instances, i.e. the imbalance ratio of the dataset, is 61:39, as shown in Fig. 5. To augment and balance the dataset, we considered two algorithms: the SMOTE algorithm and the DBA. In this step, we have only considered the MLP algorithm as a base model to repetitively apply these augmentation algorithms and evaluate the classification accuracy. The performances of the MLP-based algorithm are compared with other DL algorithms in the following step. The main objective of this current step is to evaluate the effect of each augmentation method on the MLP classifier's accuracy and select the best-suited augmentation and balancing technique to generate synthetic data which can then be used to evaluate the performance metrics of DL techniques for predicting at-risk students.

Firstly, the SMOTE algorithm was applied to our dataset. This synthetically created extra data points from the minority class (i.e. *N*). In each iteration, SMOTE augmented the instances of one class, i.e. the minority class. In each iteration, we have applied resampling to uniformly increase the class distribution by 100% each time. Due to the nature of our dataset, the SMOTE algorithm did not completely balance the dataset and thus did not eliminate the issue of bias towards the majority class.

Secondly, we applied the DBA to balance the distribution of the majority and minority instances and augment the size of the dataset. This algorithm not only augments the dataset but also balances the class distribution. This algorithm does the following:

1. Oversamples the minor class (N).
2. Undersamples the majority class (P).
3. Removes overlapped instances from N and P classes.
4. Fully balances both P and N classes.

For SMOTE, each iteration is of a different structure. Therefore, the percentage of instances to generate is 100% for each iteration of current instances. Also, each iteration is configured to auto-detect the minority class and five nearest neighbours to resample. This process is repeated until the highest accuracy is presented by the classifier.

For DBA, we have started the number of instances to resample per class label to 50 to keep the dataset size near to the original dataset size, which was 93. In each iteration, 50 instances are synthetically created on the existing instances and added to the dataset. Next, the MLP algorithm is applied to the augmented dataset to find the accuracy. This process is repeated to find the best classifier with the highest accuracy.

Table 3 and Fig. 6 show that both the SMOTE and DBA techniques were promising approaches for data augmentation to improve accuracy. SMOTE could not effectively handle high-dimensional data and suffered from overfitting issues as it uses the *k*-nearest neighbour algorithm to generate synthesized data. However, we selected the DBA augmentation approach to enlarge the dataset for the next step due to the following reasons.

- DBA has shown the highest classification accuracy when the data count is 1400, whereas SMOTE demonstrated

Table 2 Descriptive statistics of the dataset

	Idle _time	Mouse _wheel	Mouse_ wheel_click	Mouse_click_left	Mouse_ click_right	Mouse_ movement	Keystroke	Session_duration
Mean	0.989	0.207	0.026	0.496	0.439	0.210	0.278	0.643
Standard deviation	0.104	0.193	0.116	0.205	0.252	0.146	0.141	0.221
Min	0	0	0	0	0	0	0	0
25%	1.000	0.080	0.000	0.380	0.197	0.141	0.194	0.524
50%	1.000	0.163	0.000	0.491	0.469	0.191	0.281	0.715
75%	1.000	0.260	0.004	0.648	0.620	0.256	0.352	0.792
Max	1	1	1	1	1	1	1	1

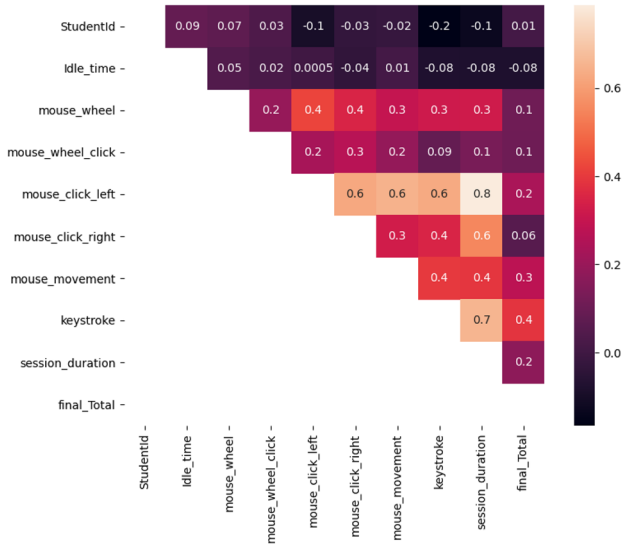


Fig. 4 Heatmap for correlational analysis for dataset set to train a model to identify student academic performance

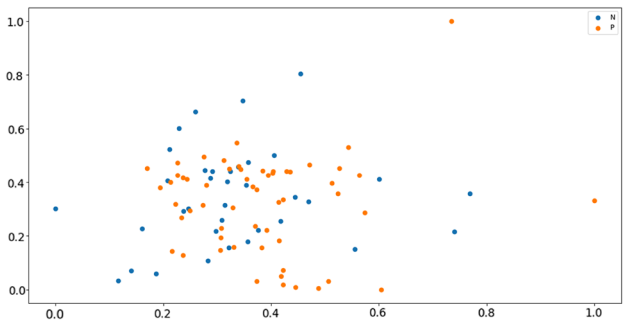


Fig. 5 Data points of original small and imbalance dataset before data augmentation

97.30% accuracy on the nearest population of the dataset of 1472 instances as highlighted in Table 3.

- DBA demonstrated significantly high accuracy (more than 90%) in the third iteration, just after the augmented dataset doubled the scale of the original dataset, i.e. the size of the dataset is equal to 200. But, SMOTE showed similar accuracy in the seventh iteration when total instances were quite large and 10 times the original dataset, i.e. the size of the dataset was more than 1000.
- DBA outputted a balanced dataset set in each iteration, which is not the case for SMOTE outputs.

3.4 Application of DL

This step is an iterative process of applying the DL algorithms on the dataset and evaluating and comparing the results of the experiments of three predictive models on both the original and augmented dataset. We used DL supervised methods to train the model, which learns from labelled classes, i.e. *P* and *N*. The SM, LSTM, and MLP algorithms were explored to identify at-risk students before and after data balancing and data augmentation. In the first iteration, we trained the DL models on the imbalanced dataset. After the correction of imbalance and lesser dataset in the previous step, we re-trained the models on balanced and the larger-scale (augmented) dataset using the same DL algorithms. The training process was executed in three iterations, one for each DL classifier, and evaluated the classification accuracy of all three classifiers. The recall and precision measure of the classifier with the highest accuracy was also evaluated.

In this study, the MLP consisted of three hidden layers (64, 32, 2) and rectified linear activation (ReLU) was used as a nonlinear activation function. The Softmax function was used on the last layers to normalize the MLP model output. In this study, TensorFlow–Keras was used to implement SM and consisted of two fully connected neural layers. In these layers, the first layer comprised 64 nodes, and the last layer had one score to indicate that it is of binary classification. The built-in optimizer Adam was also used. LTSM was configured with two layers of 64 nodes and Sigmoid as the gate activate function.

The predictive model was trained and tested by using *k*-fold cross-validation instead of dividing the data into

Table 3 Comparison of the count of augmented data and accuracy of a DL model over 15 iterations

	SMOTE				DBA			
	Total	Majority (P)	Minority (N)	Accuracy (%)	Total	Majority (P)	Minority (N)	Accuracy (%)
0 Iteration	93	57	36	62	93	57	36	62
1 Iteration	129	57	72	65.11	100	50	50	75
2 Iteration	186	114	72	68.82	200	100	100	90.50
3 Iteration	256	114	114	69.38	300	150	150	93.67
4 Iteration	372	228	144	77.69	400	200	200	97
5 Iteration	516	228	228	82.75	500	250	250	97.50
6 Iteration	744	456	288	86.15	600	300	300	97.50
7 Iteration	1032	456	576	91.08	700	350	350	97.80
8 Iteration	1488	912	576	93.28	800	400	400	98.12
9 Iteration	2064	912	1152	95.15	900	450	450	98
10 Iteration	2976	1824	1152	96.34	1000	500	500	98.20
11 Iteration	4128	1824	23.4	98.09	1100	550	550	98.80
12 Iteration	5952	3648	2304	98.77	1200	600	600	98.91
13 Iteration	8256	3648	4608	99.21	1300	650	650	98.15
14 Iteration	11,904	7296	4608	99.51	1400	700	700	98.65
15 Iteration	16,512	7296	9216	99.70	1500	750	750	98.17

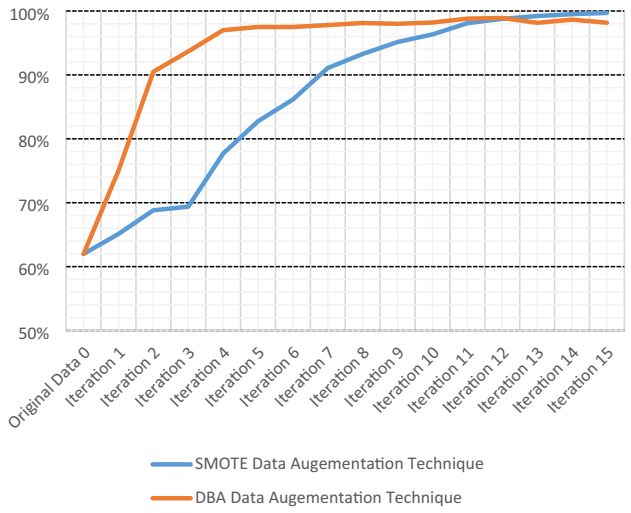


Fig. 6 Classification accuracy of each iteration of the DBA and SMOTE data augmentation technique

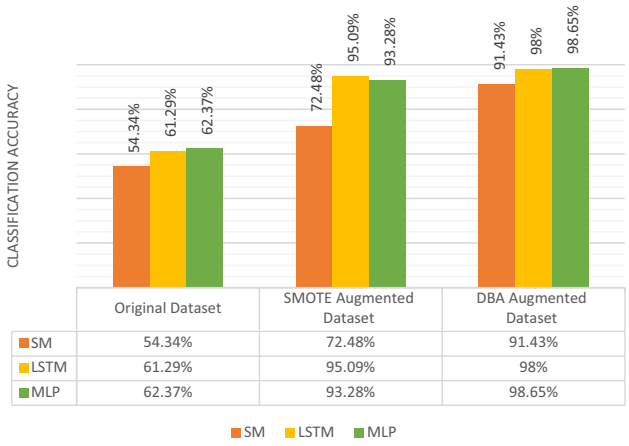


Fig. 7 Comparison of the classification accuracy of three DL algorithms on the original dataset, augmented dataset by using SMOTE and DBA algorithm

training and testing datasets. There were multiple reasons for using cross-validation in this study. One of the reasons was that using *k*-fold cross-validation mitigates the data leakage risk by performing multiple iterations of cross-validation and presenting the average of the overall performance to provide a more reliable model performance. Another significant reason was that it improves model generalization by handling noise in the dataset and preventing overfitting. In *k*-fold cross-validation, the dataset is randomly split into k subsets called folds. The algorithm was iteratively trained on $k-1$ folds of the dataset and used the remaining fold of the dataset for testing. To show the accuracy of the model in prediction, it should be tested on data that it has not seen before, where it must make predictions similar to the actual results. In this study, the dataset was divided into 10 folds: p1 to p10. Therefore, ten different models could be trained; each model was trained on nine folds and tested on the tenth. The first model was trained on p1 to p9 and tested on p10. The second model was trained on p1, p3 to p10 and tested on p2 and so on. This helped to use all instances both for training and for testing.

Classification accuracy was used to evaluate the performance of the DL algorithms as presented in Fig. 7.

Classification accuracy is used to select the DL classifier with the highest accuracy, which is calculated by using confusion metrics. All three DL classifiers performed well with considerably high accuracy after the application of data augmentation techniques, demonstrating their feasibility and effectiveness when used on a large-scale and balanced dataset.

Algorithm 1 combines and fully automates the process of collecting the dataset, applying data augmentation techniques, applying DL algorithms to identify the best classifier using *k*-fold, identifying struggling students, and offering them support by using a recommendation system. It provides an overview of this automated process which re-emphasizes the significance of automating the complete procedure of selecting the best ML classifier to identify students at risk from beginning to the end.

Algorithm 1 Algorithm for automatic identification of struggling students by using data augmentation and DL algorithms.

```
function applyDBA (Z,p) -- where Z is the original dataset and p is number of instances to re-sample per class label
  for each class c in Z do
    for each feature f in Z do
        learn probability distribution P of c and f of Z
    endfor
  endfor
  for each class do
    for i: 1..p do
      newinstance = [][]
      for each feature:f do
        newinstance[i][f] = sample value from P[f]
        add class of newinstance[i]
      endfor
      Z = Z + newinstance[i]
    endfor
  endfor
end function
X ← download training dataset
X ←remove null and negative values
X ←aggregate the data
X ←select features ← perform correlational analysis
X ←transform categorical variable
j ← 50
for loop:1..15 do
  X ←applyDBA(X,j)
  j=j+50
  loop++
endfor
apply 3 DL classifiers
    listofDLalgorithms[]={ "MLP", "SM", "LSTM" }
compare performance metrics dl1-3
k=10
n=1
for each d:listofDLalgorithms do
   CA[n] ← (1/k)x( sum of performance metric of d for k randomly subset of X)
   n++
endfor
for m: 1..3 do
  if CA[m]>CA[m+1] then
    bestdl ← m
  elseif CA[m]<CA[m+1] then
    bestdl ← m+1
  endif
endfor
collect the real student data
apply best classifier (bestdl) on real data
listofstudentsatrisk[] ← predicting students' performance with bestdl # best classifier
for each student :listofstudentatrisk[] do
    support program recommendation system based on student profile
endfor
```

The above algorithm of automatic prediction of student academic performance by using DL-based predictive model creates and verifies the dimensional vector X. Three DL algorithms (MLP, SM, and LSTM) are trained and tested on the X-dimensional vector using a *k*-fold cross-validation. A scaling technique (DBA) is applied to the X vector. The augmented vector X is again trained and tested using three DL algorithms and *k*-fold cross-validation. The prediction accuracy of the three predictive models is then compared, and the most robust and accurate model is selected. Algorithm 1 fully automates the process of creating the dimensional vector X, selecting the best predictive model based on the highest prediction accuracy, and identifying students at risk of failing to offer timely remedial activities and improving student learning.

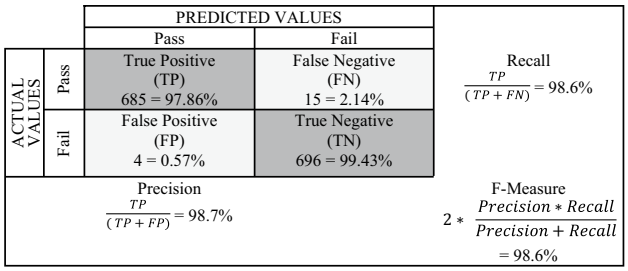


Fig. 8 Confusion metric, recall, precision, and F-measure for MLP-based classifier with DBA augmented dataset. TP = number of positive instances correctly identified as positive (the number of students correctly identified as not 'at risk'); TN = number of negative instances correctly identified as negative (the number of students correctly identified as 'at risk'; FP = number of negative instances identified incorrectly as positive (the number of at-risk students who were incorrectly identified as not 'at risk'; and FN = number of positive instances incorrectly identified as negative (the number of not-at-risk students incorrectly identified as at-risk)

4 Model evaluation and discussion

In this study, we experimented with three DL algorithms on the dataset to identify struggling students. The comparison of classifier performance based on accuracy for the three classifiers is depicted in Fig. 7, with and without the application of data augmentation techniques. All classifiers clearly showed a significant increase in accuracy in identifying students at risk with SMOTE and DBA augmentation techniques. However, MLP outperforms other classifiers for the DBA augmented dataset.

The classification accuracy of the DL technique jumped significantly, i.e. to 90.5%, as soon as the dataset population has an even distribution of majority and minority classes after the application of a data balancing algorithm. By using data balancing and data augmentation techniques, the classification accuracy of MLP models was improved to 98.65% after the data were increased to 1400 data points, with a balanced ratio of majority and minority classes of 50:50. The data augmentation repeats were stopped when the dataset size reached 1400, as further addition of synthesized data improved accuracy, but the accuracy fluctuates between 98 and 99, even when the dataset instances are doubled. In most use cases, the human user will not be able to distinguish a model accuracy difference of 1% (i.e. 98%-99%). Both models are considered good and able to solve the underlying problem of identifying students at risk.

For the predictive model, the combination of different levels of recall and precision measures has different meanings. The combination of high recall and high precision demonstrates that the classes are handled perfectly by the predictive model. The classification accuracy in Fig. 8 demonstrates that the DL classifier based on MLP handles the classes. Figure 8 shows the confusion metric, recall, and precision measures to reveal that the MLP classifier has both performed well based on the combination of recall and precision measures, i.e. recall and precision are high. Also, higher precision is desired, which means lower the number of FP instances identified by the classifier, i.e. a smaller number of at-risk students are identified incorrectly as not-at-risk students. A higher percentage of F-measure means the classifier has identified low FP and FN, i.e. low instances of not identifying students at risk who are actually at risk and high accuracy of correctly identifying students at risk.

To our knowledge, the use of student learning behaviour and the LMS participation dataset has not been previously investigated to develop a predictive model. We are unable to compare the performance of our DL-based predictive model to an existing model. However, this study has explored alternative approaches such as cross-validation and baseline model comparison to provide insight into the efficiency of our DL-based predictive model. This study has employed cross-validation to assess the robustness and generalizability of the predictive model. The *k*-fold cross-validation specified a k parameter to define the number of folds to split data into and provided more robust estimate of prediction accuracy across different data folds. Furthermore, the predictive model accuracy was compared against the baseline model. A common baseline model for predictive models is random forest (RF). A basic version of RF-based predictive model is implemented in an existing study (Fahd et al. 2021a, b) achieving the highest classification accuracy of 85.7%. Our proposed predictive model based on DL algorithms demonstrated superior performance compared to the RF baseline model with a classification accuracy of 98.65%, thus showcasing the value of our proposed predictive model and its potential.

The limitations of this work mainly exist in the synthesis of the data. Synthetic data may not capture real-world patterns and observations which skew the results, but it does

allow us to simulate a theoretical scenario where a proof of concept can be built. This questions the reliability of the data and the use of such models in the real world.

The goal of the study was to support educators to timely predict struggling students and offer them appropriate support programmes to enhance their academic progress. The proposed predictive model can be integrated into an educational decision support system to trigger flags or alerts to identify students at risk of failing automatically or upon request. This allows educators and administrators to effectively monitor student progress and take timely interventions as required. The alerts generated by the education decision support system offer a proactive outreach approach towards at-risk students. These alerts can be integrated into LMS to provide regular incremental feedback to students to encourage their learning. Furthermore, these alerts empower educators to provide targeted instructions and tailored strategies to students for continuous success. Figure 9 shows the key strategies and suggestions that may be included in the interventions.

These timely detections with tailored and targeted interventions would improve the student progress that will result in increased retention and decrease attrition with a positive impact on the student and reputation, and the financials of HE institutions. This eventually impacts the nation's economy, as students would be able to pay back the study loan and reduce their unpaid debt. In a broader sense, our translational research aimed to promote practical problem solving studies through the applications of DL, advancing technology-based innovations (Shee et al. 2021; Sabharwal and Miah, 2022) in many other problem domains. Extension of these studies provide enormous opportunities for creation of new practical knowledge, although it is recommended that adequate research methodology such as design science (Miah, 2009; Miah and Ahamed, 2011; Miah et al. 2016) can be of paramount study task that will offer guidance and supportive framework for research operations. Such design research enable innovations in other aspects of HE, such as for academic records management (Miah and Samsudin, 2017) or expert systems applications (Genemo et al. 2016) for assisting HE providers in delivering high quality outcome.

5 Conclusion

We applied DL algorithms to train a predictive model to predict student academic performance and identify struggling students. In this study, we first modified the class distribution and augmented the dataset to resolve the implications of small-scale and imbalanced datasets by using two different techniques, i.e. SMOTE and DBA. Three DL algorithms were used on the augmented dataset, and all of them showed satisfactory results. It was demonstrated that an increase in data points by using a good augmentation method leads to higher classification accuracy and reduces false prediction. This means a combination of a good augmentation technique and a good classifier, results in a better performing model.

In the future, we will apply the trained model to real data from the education field. Also, it will be beneficial to integrate a recommender system approach to offer appropriate support programmes to struggling students based on their profiles and on programmes that benefit other students with similar profiles.

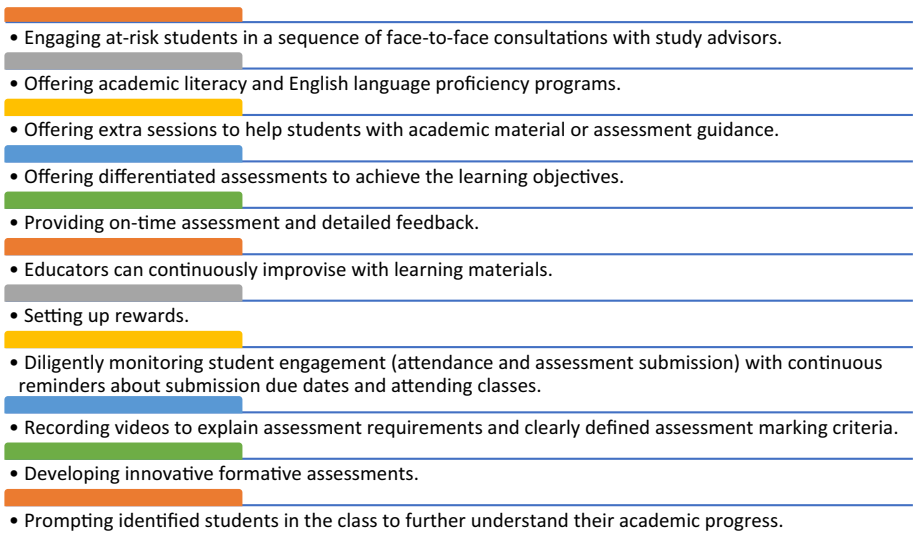


Fig. 9 Key interventions strategies and suggestions

Appendix 1: List of dataset dimensions

	Dimension	Brief description	Range of values
1	Session	Session number	
2	Student_Id	Student identification number	1–115
3	Exercise	Exercise number in a specific session	1–6, i.e. Es_session number_execrise number
4	Activity	Abbreviation of activities categorized into 15 categories	e.g. Diagram or Deeds
5	Start_time	Start date and time of an activity	dd.mm.yyyy hh:mm:ss, e.g. 2.10.2020 10:27:39
6	End_time	End date and time of an activity	dd.mm.yyyy hh:mm:ss, e.g. 2.10.2020 11:20:30
7	Idle_time	The idle time during the period of an activity	0–1,722,305,428 milli sec
8	Mouse_wheel	Number of the mouse wheel	0–1111
9	Mouse_wheel_click	Count of mouse wheel clicks in a specific activity	0–60
10	Mouse_click_left	Count of mouse left clicks in a specific activity	0–338
11	Mouse_click_right	Count of mouse right clicks in a specific activity	0–48
12	Keystroke	Count of keystrokes in a specific activity	0–2307
13	Mouse_movement	Distance covered by the mouse movements in a specific activity	0–32,484

Author contributions All authors worked and reviewed the manuscript

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Competing interests The authors declare no competing interests.

References

Abid M, Ullah DS, Siddique M, Mushtaq M, Aljedaani W, Rustam F (2022) Spam SMS filtering based on text features and supervised machine learning techniques. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-12991-0

Ahmad Tarmizi SS, Mutalib S, Abdul Hamid NH, Abdul Rahman S (2019) A review on student attrition in higher education using big data analytics and data mining techniques. Int J Mod Educ Comput Sci 11(8):1–14. https://doi.org/10.5815/ijmecs.2019.08.01

Ajoodha R, Jadhav A, Dukhan S (2020) Forecasting learner attrition for student success at a South African University

Akour M, Sghaier HA, Al Qasem O (2020) The effectiveness of using deep learning algorithms in predicting students achievements. Indones J Electr Eng Comput Sci. https://doi.org/10.11591/ijeecs.v19.i1.pp388-394

Aldowah H, Al-Samarraie H, Fauzy WM (2019) Educational data mining and learning analytics for 21st century higher education: a review and synthesis. Telemat Inform 37:13–49. https://doi.org/10.1016/j.tele.2019.01.007

Allah AGF (2020) Using machine learning to support students' academic decisions. J Theor Appl Inf Technol 8(10):3778–3796

Barari S (2019) Deep Learning in Python: Different Types of Deep Learning Networks [Video]. SAGE Publications, Ltd, London

Beer C, Lawson C (2016) The problem of student attrition in higher education: an alternative perspective. J Furth High Educ 41(6):773–784. https://doi.org/10.1080/0309877x.2016.1177171

Berens J, Schneider K, Gortz S, Oster S, Burghoff J (2019) Early detection of students at risk—predicting student dropouts using administrative student data from German Universities and machine learning methods. J Educ Data Min 11(3):1–41

Bermejo P, Gámez JA, Puerta JM (2011) Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets. Expert Syst Appl 38(3):2072–2080. https://doi.org/10.1016/j.eswa.2010.07.146

Canty AJ, Chase J, Hingston M, Greenwood M, Mainsbridge CP, Skalicky J (2020) Addressing student attrition within higher education online programs through a collaborative community

of practice. J Appl Learn Teach. https://doi.org/10.37074/jalt.2020.3.s1.3
Chawla NV, Bowyer KW, Kegelmeyer LOHWP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
Doleck T, Lemay DJ, Basnet RB, Bazelais P (2019) Predictive analytics in education: a comparison of deep learning frameworks. Educ Inf Technol 25(3):1951–1963. https://doi.org/10.1007/s10639-019-10068-4
Fahd K, Miah SJ, Ahmed K, Venkatraman S, Miao Y (2021a) Integrating design science research and design based research frameworks for developing education support systems. Educ Inf Technol 26(4):4027–4048. https://doi.org/10.1007/s10639-021-10442-1
Fahd K, Miah SJ, Ahmed K (2021b) Predicting student performance in a blended learning environment using learning management system interaction data. Appl Comput Inform. https://doi.org/10.1108/ACI-06-2021-0150
Fatima EB, Omar B, Abdelmajid EM, Rustam F, Mehmood A, Choi GS (2021) Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection. IEEE Access 9:28101–28110. https://doi.org/10.1109/ACCESS.2021.3056285
Fok WWT, He YS, Yeung HHA, Law KY, Cheung K, Ai Y, Ho P (2018) Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine. In: 4th IEEE international conference on information management
Fonseca A, Cabral B (2019) Designing a neural network from scratch for big data powered by multi-node GPUs. In: Howlett RJ, Jain LC (eds) Handbook of deep learning applications. Springer. https://doi.org/10.1007/978-3-030-11479-4
Gupta A, Anjum GS, Katarya R (2021) InstaCovNet-19: a deep learning classification model for the detection of COVID-19 patients using Chest X-ray. Appl Soft Comput 99:106859. https://doi.org/10.1016/j.asoc.2020.106859
George A-J, McEwan A, Tarr J-A (2021) Accountability in educational dialogue on attrition rates: understanding external attrition factors and isolation in online law school. Australas J Educ Technol. https://doi.org/10.14742/ajet.6175
Genemo H, Miah SJ, McAndrew A (2016) A design science research methodology for developing a computer-aided assessment approach using method marking concept. Educ Inf Technol 21:1769–1784
Hernández-Blanco A, Herrera-Flores B, Tomás D, Navarro-Colorado B (2019) A systematic review of deep learning approaches to educational data mining. Complexity 2019:1–22. https://doi.org/10.1155/2019/1306039
Heublein U (2014) Student drop-out from German Higher Education Institutions. Eur J Educ 49(4):497–513. https://doi.org/10.1111/ejed.12097
Hippel PTV, Hofflinger A (2020) The data revolution comes to higher education: identifying students at risk of dropout in Chile. J High Educ Policy Manag 43(6):1–22. https://doi.org/10.1080/1360080X.2020.1739800
Hoffait A-S, Schyns M (2017) Early detection of university students with potential difficulties. Decis Support Syst 101:1–11. https://doi.org/10.1016/j.dss.2017.05.003
Hosseini M-P, SenbaoLu KK, Slowikowski A, Venkatesh HC (2020) Deep learning architectures. In: JanuszKacprzyk PAOS (ed) Deep learning: concepts and architectures, vol 866. Springer. https://doi.org/10.1007/978-3-030-31756-0
Imran M, Latif S, Mehmood D, Shah MS (2019) Student academic performance prediction using supervised learning techniques. Int J Emerg Technol Learn (iJET). https://doi.org/10.3991/ijet.v14i14.10310
Iqbal Z, Qayyum A, Latif S, Qadir J (2019) Early student grade prediction: an empirical study. In: 2019 2nd International conference on advancements in computational sciences (ICACS), Pakistan, pp 1–7. https://doi.org/10.23919/ICACS.2019.8689136
Issah I, Appiah O, Appiahene P, Inusah F (2023) A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. Decis Anal J 7:100204. https://doi.org/10.1016/j.dajour.2023.100204
Katarya R, Arora Y (2020) Capsmf: a novel product recommender system using deep learning based text analysis model. Multimed Tools Appl. https://doi.org/10.1007/s11042-020-09199-5
Katarya R, Gaba J, Garg A, Verma V (2021) A review on machine learning based student's academic performance prediction systems. In: 2021 International conference on artificial intelligence and smart systems (ICAIS), India, pp 254–259. https://doi.org/10.1109/ICAIS50930.2021.9395767
Kedia P, Katarya R (2021) CoVNet-19: a deep learning model for the detection and analysis of COVID-19 patients. Appl Soft Comput 104:107184. https://doi.org/10.1016/j.asoc.2021.107184
Lehr S, Liu H, Kinglesmith S, Konyha A, Robaszewska N, Medinilla J (2016) Use educational data mining to predict undergraduate retention. In: 2016 IEEE 16th international conference on advanced learning technologies (ICALT), USA, pp 428–430. https://doi.org/10.1109/ICALT.2016.138
Monllaó Olivé D, Huynh D, Reynolds M, Dougiamas M, Wiese D (2020) A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC. J Comput High Educ 32:428–430. https://doi.org/10.1109/ICALT.2016.13810.1007/s12528-019-09230-1
Mngadi N, Ajoodha R, Jadhav A (2020) A conceptual model to identify vulnerable undergraduate learners at higher-education institutions. https://doi.org/10.1109/IMITEC50163.2020.9334103
Munappy A, Bosch J, Olsson HH, Arpteg A, Brinne B (2019) Data management challenges for deep learning. In: 2019 45th Euromicro conference on software engineering and advanced applications (SEAA)
Munguia P (2020) Preventing student and faculty attrition in times of change. In: Radical solutions and learning analytics. Springer. https://doi.org/10.1007/978-981-15-4526-9_8
Miah SJ, Ahamed R (2011) A cloud-based DSS model for driver safety and monitoring on Australian roads. Int J Emerg Sci 1(4):634–648
Miah SJ (2009) End user as application developer for decision support. In: Proceedings of the Fifteenth Americas Conference on Information Systems 2009, AMCIS. vol 2, pp 142
Miah SJ, McGrath GM, Kerr D (2016) Design science research for decision support systems development: recent publication trends in the premier IS journals. Australas J Inf Syst 20:1–14
Miah SJ, Samsudin, AZH (2017) EDRMS for academic records management: a design study in a Malaysian University. Educ Inf Technol 22:1895–1910
Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. J Big Data. https://doi.org/10.1186/s40537-014-0007-7
Katarya R (2019). A review: predicting the performance of students using machine learning classification techniques. In: 2019 Third international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), India, pp 36–41. https://doi.org/10.1109/I-SMAC47947.2019.9032493
Katarya R (2023) Deep auto encoder based on a transient search capsule network for student performance prediction. Multimed Tools Appl 82(15):23427–23451. https://doi.org/10.1007/s11042-022-14083-5
Reshi AA, Ashraf I, Rustam F, Shahzad HF, Mehmood A, Choi GS (2021) Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. PeerJ Comput Sci 7:e547. https://doi.org/10.7717/peerj-cs.547
Rupapara V, Rustam F, Fatima Shahzad H, Mehmood A, Ashraf I, Choi GS (2021) Impact of SMOTE on imbalanced text features for toxic

comments classification using RVVC model. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3083638

Salal YK, Abdullaev SM (2020) Deep learning based ensemble approach to predict student academic performance: case study. In: 2020 3rd International conference on intelligent sustainable systems (ICISS), India, pp 191–198. https://doi.org/10.1109/ICISS49785.2020.9316044

Sabharwal R, Miah SJ (2022) An intelligent literature review: adopting inductive approach to define machine learning applications in the clinical domain. J Big Data 9:53. https://doi.org/10.1186/s40537-022-00605-3

Shcheglova I, Gorbunova E, Chirikov I (2020) The role of the first-year experience in student attrition. Qual High Educ 26(3):307–322. https://doi.org/10.1080/13538322.2020.1815285

Shin H, Lee K, Lee C (2020) Data augmentation method of object detection for deep learning in maritime image. In: 2020 IEEE international conference on big data and smart computing (BigComp), Korea (South), pp 463–466. https://doi.org/10.1109/BigComp48618.2020.00-25

Shingari I, Kumar D, Khetan M (2017) A review of applications of data mining techniques for prediction of students' performance in higher education. J Stat Manag Syst 20(4):713–722. https://doi.org/10.1080/09720510.2017.1395191

Shee H, Miah SJ, de Vass T (2021) Impact of smart logistics on smart city sustainable performance: an empirical investigation. Int J Logist Manag 32(3):821–845

Sultana J, Rani MU, Farquad MAH (2019) Student's performance prediction using deep learning and data mining methods. Int J Recent Technol Eng (IJRTE) 8(1):1018–1021

Sun D, Mao Y, Du J, Xu P, Zheng Q, Sun H (2019) Deep learning for dropout prediction in MOOCs. In: 2019 Eighth international conference on educational innovation through technology (EITT)

Tsiakmaki M, Kostopoulos G, Kotsiantis S, Ragos O (2020) Transfer learning from deep neural networks for predicting student performance. Appl Sci. https://doi.org/10.3390/app10062145

Vahdat M, Oneto L, Anguita D, Funk M, Rauterberg M (2015) A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. In: Conole G, Klobučar T, Rensing C, Konert J, Lavoué E (eds) Design for teaching and learning in a networked world. EC-TEL 2015. Lecture notes in computer science, vol 9307. Springer. https://doi.org/10.1007/978-3-319-24258-3_26

Veluri RK, Patra I, Naved M, Prasad VV, Arcinas MM, Beram SM, Raghuvanshi A (2022) Learning analytics using deep learning techniques for efficiently managing educational institutes. Mater Today Proc 51:2317–2320

Wakelam E, Jefferies A, Davey N, Sun Y (2020) The potential for student performance prediction in small cohorts with minimal available attributes. Br J Educ Technol 51(2):347–370. https://doi.org/10.1111/bjet.12836

Wlodarczak P (2019) Deep learning in eHealth. In: Howlett RJ, Jain LC (eds) Handbook of deep learning applications. Springer. https://doi.org/10.1007/978-3-030-11479-4

Xing W, Du D (2018) Dropout prediction in MOOCs: using deep learning for personalized intervention. J Educ Comput Res 57(3):547–570. https://doi.org/10.1177/0735633118757015

Zhao W (2017) Research on the deep learning of the small sample data based on transfer learning. AIP Conf Proc 1864(1):020018. https://doi.org/10.1063/1.4992835

Zhou Z-H, Feng J (2017) Deep forest: towards an alternative to deep neural networks. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17), Melbourne, Australia, pp 3553–3559. https://doi.org/10.24963/ijcai.2017/497