Business Systems Research | Vol. 16 No. 2 |2025

sciendo

Early Identification of At-Risk Students in Online Education: A Deep Learning Approach to Predictive Modelling

Lediana Shala Riza
South East European University, Faculty of Contemporary Sciences and Technologies, North Macedonia
Lejla Abazi Bexheti
South East European University, Faculty of Contemporary Sciences and Technologies, North Macedonia
Jovana Zoroja
Faculty of Economics and Business, University of Zagreb, Croatia

Abstract

Background: Predicting student performance in online learning is difficult due to class imbalance and limited model interpretability. At-risk students are fewer than high performers, biasing predictions, and methods like SMOTE fail to preserve temporal patterns. Although black-box models are accurate, they lack transparency for actionable insights. **Objectives:** This study proposes a deep learning framework combining LSTM networks and attention mechanisms to address these issues using the OULAD dataset. LSTMs capture temporal dependencies, while attention improves interpretability by emphasising key features. Advanced resampling mitigates class imbalance for robust at-risk student detection. **Methods/Approach:** The methodology applies the KDD framework to process data, uncover patterns, and build models that predict student success risk, ensuring efficient data handling, robust modelling, and actionable insights to improve retention. **Results:** The BiLSTM-RNN achieved the best performance, effectively capturing temporal dependencies and attaining the highest accuracy, precision, recall, and F1-score. **Conclusions:** The findings support more effective and targeted interventions in online education, offering valuable insights for research and practice.

Keywords: student performance prediction; online learning environments; LSTM; deep learning framework; SMOTE

JEL classification: C88
Paper type: Research article

Received: Jan 14, 2025
Accepted: Aug 15, 2025

Citation: Shala Riza, L., Abazi Bexheti, L., & Zoroja, J. (2025). Early Identification of At-Risk Students in Online Education: A Deep Learning Approach to Predictive Modelling. *Business Systems Research, 16*(2), 69-91.
DOI: https://doi.org/10.2478/bsrj-2025-0019

Introduction

In recent years, the rapid evolution of educational technology and the increasing adoption of online learning platforms have significantly transformed the landscape of higher education. Online and blended learning environments, characterised by their flexibility, accessibility, and scalability, have become an essential component of modern educational systems. These platforms offer students worldwide vast opportunities to access quality education, breaking traditional barriers such as geographic location, financial constraints, and institutional limitations. Behavioural engagement, cognitive engagement, emotional engagement, or the expression of delight towards the course materials are just a few examples of the various ways students might engage on online platforms (Brahim, 2022). In any learning system, student engagement serves as a primary metric for assessing how well students are performing. It also plays a significant role in determining academic outcomes, including dropout and withdrawal rates, as well as low achievement (Soobramoney, 2021). However, with the advantages of online education come significant challenges, one of the most pressing being identifying and supporting students at risk of poor performance or early dropout. Timely identification and intervention can play a crucial role in improving academic outcomes and ensuring student retention. However, many educational institutions continue to struggle to predict student success in online learning environments effectively (Alruwais & Zakariah, 2023).

One of the widely recognised datasets in educational data mining and learning analytics is the Open University Learning Analytics Dataset (OULAD). This dataset provides a wealth of information about student demographics, interaction with online course materials, assessment results, and overall performance across multiple modules. The availability of such datasets has enabled researchers to apply machine learning (ML) and deep learning (DL) techniques to uncover patterns, predict outcomes, and propose interventions to improve student success. Traditional ML models, such as logistic regression, decision trees, and random forests, have demonstrated moderate success in predicting student performance. However, these models often fail to fully capture the temporal dependencies and sequential patterns inherent in student interaction data, which are crucial for understanding how student engagement evolves.

Deep learning approaches, particularly recurrent neural networks (RNNs) and their advanced variants like Long Short-Term Memory (LSTM) networks, have shown considerable promise in addressing these limitations. LSTMs are specifically designed to capture long-term dependencies in sequential data, making them particularly suitable for analysing time-series generated by students' online interactions. Despite their advantages, LSTMs still face challenges such as overfitting, vanishing gradients, and difficulty effectively highlighting critical temporal features in long sequences. To address these issues, attention mechanisms have emerged as a powerful enhancement to traditional LSTM architecture. By allowing the model to focus on the most relevant parts of the input sequence at each prediction step, attention mechanisms improve both the interpretability and predictive power of LSTM models.

In the context of student performance prediction, integrating attention mechanisms with LSTMs offers a novel approach to capturing nuanced behavioural patterns in students' learning activities. For example, certain activities, such as completing an assessment or participating in a discussion forum, might carry more weight in predicting performance compared to passive activities like reading course materials. It can be challenging to choose which engagement component or factors to represent. Effective intervention can be supported without detecting every element, or every facet of every factor. Depending on the learning context, certain aspects are more crucial to recognise and adjust to than others (Inder, 2022). Attention-enhanced LSTMs can dynamically assign importance to these activities, providing more accurate and interpretable predictions. Moreover, such models can help identify critical intervention points where educators can provide timely support to at-risk students.

While student performance prediction models have made significant progress, several challenges persist. One of the most notable challenges is class imbalance, where the number of students at risk of failing or dropping out is significantly lower than the number of students performing well. This imbalance can bias models toward overpredicting the majority class, reducing their effectiveness at identifying at-risk students. Techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) and other resampling strategies have been employed to address this issue. However, they are not always effective in preserving the temporal structure of the data. Another challenge is the lack of interpretability in black-box models like deep neural networks. While these models are highly accurate, educators and administrators often require clear explanations for predictions to make informed decisions.

To overcome these challenges, this study proposes an advanced deep learning framework that combines LSTM networks with attention mechanisms to predict student performance on the OULAD dataset. The proposed model aims to address class imbalance through advanced resampling techniques and optimise model interpretability using attention weights. By leveraging LSTMs' strengths in capturing temporal dependencies and the attention mechanism's ability to highlight critical input features, this approach offers a robust solution for early prediction of student performance. This study aims to determine features that predict at-risk students, build ML and DL models, including LSTM and RNN architectures, evaluate their performance, and identify the optimal academic calendar week for detecting at-risk students to support early intervention.

Furthermore, this research seeks to explore the integration of multi-modal learning approaches, combining numerical, categorical, and textual data from the OULAD dataset. Students' engagement data, assessment scores, demographic information, and text-based feedback collectively offer a rich, diverse set of features that multi-modal deep learning architectures can harness. By integrating these diverse modalities, the proposed model aims to achieve higher predictive accuracy while providing educators with actionable insights.

The specific hypotheses of this research are outlined as follows:

- H1: Student engagement levels can be accurately predicted with a high degree of accuracy by applying Machine Learning/Deep Learning algorithms to data from virtual learning environments.
- H2: Early detection of student engagement can be reliably predicted by utilising certain variables that are collected from student interactions inside virtual learning environments.
- H3: Over-sampling techniques are more effective than traditional methods in synthesising balanced datasets and addressing class imbalance in classification tasks.

By exploring these objectives using the OULAD dataset, this study seeks to contribute novel insights into predictive analytics in higher education. Unlike traditional approaches that rely on static point estimates, advanced ML and DL techniques can offer more dynamic and adaptable predictive models. In conclusion, this research advances the field of learning analytics by integrating ML and DL methods to more accurately predict at-risk students. The findings are expected to inform policy decisions, enhance intervention strategies, and ultimately improve student success outcomes in higher education institutions. The remainder of this paper is structured as follows: Following the introduction, the Literature Review presents key developments in educational data mining, with a focus on deep learning methods applied to predicting student engagement. The Methodology details the use of the OULAD dataset, including data selection, preprocessing, transformation, modelling, evaluation, and the proposed BiLSTM-RNN approach. The Results section presents a critical evaluation of model performance, followed by discussions on Research Limitations, Theoretical Contributions, and Practical Implications. The paper concludes with a summary of key findings and a list of References.

Literature review

Student retention and learning achievement remain significant challenges in Massive Open Online Courses (MOOCs). Accurately identifying at-risk students – those likely to drop out or perform poorly – has been a significant focus of recent research. Early identification enables timely interventions to mitigate dropout rates and improve student outcomes. This section reviews recent advancements in detecting at-risk students in MOOCs using various machine learning and deep learning approaches. Despite significant advancements in predictive analytics, identifying at-risk students remains challenging. Numerous studies have applied ML techniques, including regression and classification algorithms, to improve prediction accuracy. These methods often utilise decision trees, random forests, neural networks, support vector machines, and naive Bayes classifiers (Hafzan et al., 2019). However, traditional ML approaches often fail to account for the inherent uncertainty in educational data.

A significant study by Geigle and Zhai (2017) introduced a two-layer Hidden Markov Model (TL-HMM) to predict student performance by analysing both observable student behaviours and hidden latent factors. Unlike traditional Hidden Markov Models (HMMs), which are typically limited to linear state transitions, TL-HMM allows for more nuanced detection of student behavioural patterns. This method is particularly effective at capturing micro-behaviours, such as the transition from engaging in quizzes to participating in forums. By utilising the TL-HMM, researchers were able to infer underlying patterns in student activities that might predict future performance. For instance, high-performing students exhibited fewer transitions between latent states, indicating that they had already mastered specific skills and thus required less support. This model demonstrated its potential to detect subtle behavioural patterns that may influence student success, offering an advantage over traditional models that rely solely on observable data (Choi et al., 2018).

In a similar vein, Chaplot et al. (2015) applied feedforward neural networks to identify at-risk students in MOOCs, using sentiment analysis and clickstream data as baseline features. This approach utilised a large dataset from the Coursera platform, consisting of 3 million students' click logs and 5,000 forum posts. One of the primary challenges in their work was dealing with an imbalanced dataset, as the number of students at risk of attrition was much smaller than that of those who remained engaged. To address this, the authors employed Cohen’s Kappa metric, which provides a better performance metric for imbalanced datasets than traditional accuracy. The results showed that incorporating sentiment features alongside clickstream data led to an accuracy of 74%. When sentiment features were excluded, the accuracy dropped to 70%, indicating the significant role sentiment analysis played in improving prediction performance. This study highlights the value of integrating multiple feature types, such as sentiment and behavioural data, in predicting student outcomes (Chaplot et al., 2015).

Xing and Du (2018) introduced the ConRec Network, a hybrid deep learning model that combines Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to predict dropout in MOOCs. The dataset used included time-stamped student records, such as event time, event type, and student enrollment date. The ConRec Network consisted of two parts: the CNN component, which automatically extracted features from the raw student data, and the RNN component, which aggregated these features to make predictions. The results showed that the ConRec Network performed similarly to other baseline models, with an F1-score range of 90.74-92.48%. Despite the similarity in performance, the authors argued that the ConRec Network outperformed other models in efficiency, as it could automatically extract features without manual feature engineering. This makes the model highly scalable and adaptable to different learning environments.

Palani et al. (2021) use machine learning techniques in their withdrawal-prediction model, enabling it to identify early signs of low student commitment. According to the study's results, the J48 classifier achieved an average F-measure of 0.796, followed by the Random Forest, MLP, Bayesian, and SVM classifiers.

Another key study aimed to address dropout rates by improving the performance of an early warning system for predicting student dropout. This study used the synthetic minority oversampling technique (SMOTE) to address class imbalance, a common issue in dropout prediction models. The research found that dropout students were more likely to exhibit problematic attendance, low academic achievement, and minimal participation in extracurricular activities, all of which are critical indicators of disengagement and potential dropout. By combining machine learning algorithms, including Random Forest (RF) and Boosted Decision Trees (BDT), the study achieved highly accurate predictions of student dropout behaviour (Lee & Chung, 2019). The results of this study indicated that applying SMOTE to these classifiers significantly improved model performance. The AUC (Area Under the Curve) values for the Receiver Operating Characteristic (ROC) curves were 0.986 for RF, 0.988 for BDT, 0.986 for SMOTE + RF, and 0.991 for SMOTE + BDT, showcasing the effectiveness of these techniques in predicting student dropout with high precision. Similarly, the Precision-Recall (PR) curve also showed improvement with the SMOTE-enhanced models. These findings underscore the importance of early interventions to prevent dropout and improve graduation rates, as the early warning system can help redirect at-risk students onto the path to success. The study used data from a large sample of 165,715 high school students collected from the NEIS database in South Korea, providing a robust dataset for model training and evaluation (Lee & Chung, 2019).

These studies demonstrate the wide range of approaches being applied to detect at-risk students in MOOCs, from deep learning models that automatically extract features to traditional machine learning techniques enhanced by feature optimisation. By leveraging diverse data sources, such as clickstream logs, sentiment data, and behavioural patterns, these models provide valuable insights into student performance and enable timely intervention strategies.

Methodology

This research focuses on predicting potential student failures in online courses by analysing historical data, specifically targeting patterns related to academic performance, engagement, and behavioural data. The methodology utilises the Knowledge Discovery in Databases (KDD) framework as a structured approach to systematically process data, identify relevant patterns, and build machine learning models capable of predicting risks to student success or failure. By leveraging the KDD process, efficient data handling is ensured, a robust model is developed, and

actionable insights are generated to support improvements in student retention and success within online learning environments. The KDD process serves as the foundation for this research, providing a systematic, iterative approach to extracting meaningful patterns from the OULAD dataset. The KDD process is widely adopted across different fields, including education, where large datasets of student behaviour and performance can be analysed to predict academic outcomes. This section describes each stage of the KDD process and its application to predicting early risks of student failure in online courses (Marban et al., 2012).
The KDD process comprises several stages:

- Data Selection
- Data Preprocessing (Cleaning and Integration)
- Data Transformation
- Modelling
- Evaluation and Interpretation

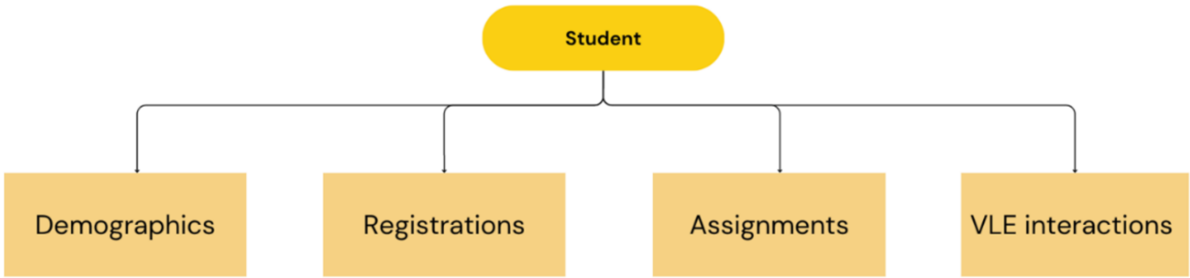
Dataset description

In this study, the publicly accessible OULAD dataset will be used for experimental analysis. The dataset is distinguished by its inclusion of aggregated clickstream data capturing students' interactions within the Virtual Learning Environment (VLE), alongside comprehensive demographic information (Figure 1). This enables the analysis of student behaviour based on their recorded actions (Peach et al., 2019). The ability for instructors to view the activities their students engage in on the VLE and analyse them to understand student behaviour better is one benefit of the platform. Typically, OU students are split up into groups, with an instructor assigned to each group (Hussain et al., 20018). The collection includes details on 32,593 students, 22 courses, assessment results, and records of their interactions with the virtual learning environment (VLE), which are represented as daily summaries of student clicks (10,655,280 entries). Raw data is aggregated from several data files, with newly constructed attributes. The dataset includes logs of student interactions with VLE and the outcomes of student assessments for each triplet of student, module, and presentation. The dataset comprises seven modules: three social science modules (AAA, BBB, GGG) and four STEM modules (CCC, DDD, EEE, FFF). The dataset includes a significant number of failing students.

- Data regarding how students use the VLE may be found in the *studentVLE* database. It contains the number of interactions, interaction dates, student IDs, VLE material identification numbers, presentation codes, and module codes. This table offers a thorough understanding of how students interact with online course materials.
- Details on the registration and registration of student modules are included in the *studentRegistration* table. Module codes, presentation codes, student IDs, and the dates of registration and registration (if any) are all included. The date of un-registration is also noted for students who have withdrawn their enrollment. There are 32,593 rows in it.
- Information on the resources available in the Vin is gVLEn on the *vle* table. It contains activity categories, module, and presentation codes, VLE material identification numbers, and the estimated amount of time the resources will be used.
- Student demographic data, including module and presentation codes, student IDs, gender, region, highest education level, IMD band, age band, number of prior attempts, studied credentials, disability status, and outcomes of the final module presentation, are contained in the *studentInfo* table.

- The outcomes of the students' assessments are listed in the *studentAssessment* table. It contains the assessment IDs, student IDs, submission dates, whether the assessment results were carried over from an earlier presentation, and the student scores.
- Information regarding the assessments included in module presentations is available on the *assessments* table. It contains assessment IDs, module codes, presentation codes, assessment dates, assessment weights, and assessment types (such as tutor-marked, computer-marked, and final exams). Multiple tests and a final exam typically follow every presentation.

Figure 1
Dataset structure



Source: Kuzilek et al., 2017

Data selection

OULAD is a compilation of tabular student data from 2013 and 2014. Each table contains different information, and identified columns can be used to connect data across tables (Figure 2). Since 2012, records of student demographics, completed modules, and virtual learning experiences have been stored in the data warehouse. A selection of representative modules offered at the Open University in 2013 and 2014 was chosen for analysis. The selected module presentation has more than 500 students, and a considerable number of them are failing. Of all the modules that met these requirements, seven were chosen: four STEM (science, technology, engineering, and mathematics) modules, three social science modules, and a total of 38,239 students (Peach et al., 2019). Four classes/grades are created based on students' final performance: 31% were removed, 22% failed, 38% passed, and 9% were awarded distinction status. Students had the option of choosing from seven modules, each taught at least twice during the year (Bittner, 2022).

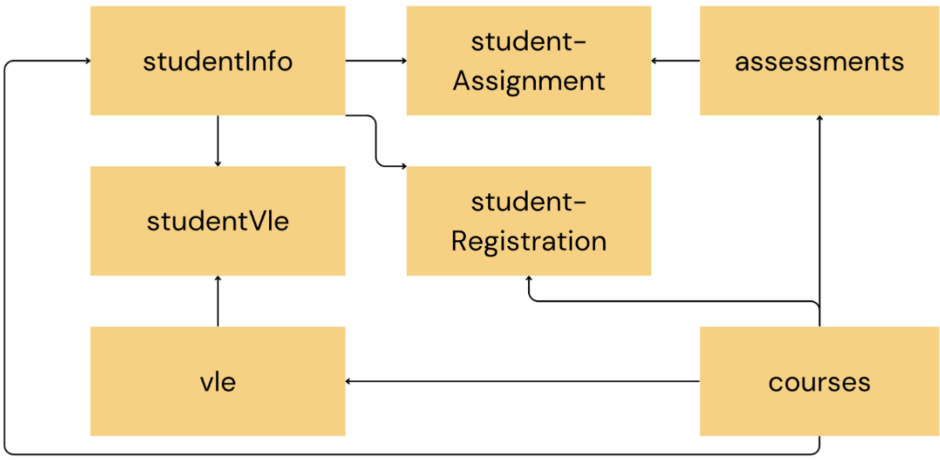
The Open University used SAS technology to build a data warehouse to compile data from these many platforms. By combining information from multiple sources, this data warehouse serves as a central repository, providing researchers with a single data set for study.

- The demographic data comprises details about the age, gender, region, highest level of education attained, and status of disabilities of the students. This data provides researchers with information about the student body and enables them to investigate the potential effects of these factors on academic performance.
- Results and accomplishments of the students are recorded in the performance data. It contains details on the number of prior attempts for a given module,

the total number of credits studied, student identity numbers, presentation and module identification codes, and the outcome of each module presentation.

- Student interactions with the VLE are the primary focus of the learning behaviour data. It contains details on the modules studied, presentation codes, student ID numbers, VLE material ID numbers, interaction dates, and interaction counts. This information sheds light on students' usage habits, levels of involvement in online forums and activities, and interactions with online learning tools (Alnassar, 2023).

Figure 2
Dataset table associations



Source: Kuzilek et al., 2017

- Based on similar approaches, the dataset was deemed appropriate for this study, as it contains the necessary demographic and VLE interaction data required for machine learning to identify student performance and engagement. The attributes included in the dataset are considered common and applicable to any dataset with a comparable structure, regardless of the context or location.

Data selection is the first critical step in the KDD process, in which the relevant data for analysis are identified and collected. In this research, the OULAD dataset is used, which contains detailed information on students enrolled in online courses, including their demographic characteristics, participation in course activities, and academic performance. For the purpose of predicting the risk of failure, the following variables were selected:

- Demographic Information: Includes gender, age, and educational background.
- Course Interaction Data: Includes time spent on course materials, discussion forums, and quizzes.
- Grades: Historical performance data, such as grades on assignments and final exams.
- Engagement Metrics: Metrics indicating student engagement, such as the number of logins and interactions with course resources.

- Time Features: Time-related features such as the date of first login, the time spent on activities, and the duration until the final grade is assigned (Jawad et al., 2022).

The quality of the selected features directly influences the model's predictive power. Therefore, the features chosen to reflect both academic performance indicators and engagement metrics provide a comprehensive dataset that captures the underlying factors contributing to student failure.

Data preprocessing

After selecting the relevant data, the next step is data preprocessing, which involves cleaning, integrating, and formatting the data for modelling. Raw educational data often contains missing values, inconsistencies, and noise that can affect model performance. Key preprocessing tasks in this research included:

- Handling Missing Values: Missing data in the OULAD dataset were addressed using imputation (mean/median for continuous features, mode for categorical variables) or removal for instances with extensive missing values.
- Outlier Detection: Outliers in performance variables (e.g., extreme grades) were capped or transformed to avoid distorting predictions.
- Data Integration: Data from sources like demographics, course logs, and grades were merged to align features with student behaviour and outcomes.
- Feature Encoding: Categorical variables (e.g., gender, course names) were encoded with one-hot or label encoding for compatibility with algorithms.
- Normalisation: Continuous features (e.g., activity time, grades) were normalised using Min-Max scaling to ensure consistent feature ranges.

Data transformation

After cleaning and integrating the data, it must be transformed into a format suitable for analysis to improve input quality and ensure compatibility with machine learning algorithms. Key tasks included:

- Normalisation and Standardisation: Features like activity time, grades, and interaction metrics were normalised to ensure comparable scales, preventing disproportionate influence from larger numerical ranges.
- Feature Engineering: New features were created to enhance insights, such as:
    - Engagement Scores: A composite metric combining login and forum participation for a holistic view of engagement.
    - Time Features: Metrics like time spent on the course and first interaction time to capture behaviour trends.
- Dimensionality Reduction: Techniques like PCA were applied to reduce dataset dimensionality, retaining key features while simplifying and reducing computational demands.

The transformed data is now optimised for modelling, enhancing accuracy and predictive power.

Modelling

Modelling is the central phase of the KDD process, in which machine learning algorithms are applied to prepared data to uncover patterns and make predictions. For this research, several machine learning algorithms were explored to predict student failure using processed data. The models used include both traditional machine learning algorithms (Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbours, Random Forest, and AdaBoost) and more advanced deep learning models (FCN, LSTM, and BiLSTM-RNN).

Logistic Regression: Logistic Regression was chosen as the baseline model for classification tasks. It is simple, interpretable, and effective for binary classification problems, such as predicting student failure. This model estimates the probability of failure based on the relationship between input features and the target variable (Hilbe, 2015).

Random Forest: An ensemble learning method, it was applied due to its ability to handle complex datasets with numerous features. Combining the predictions of multiple decision trees yields a robust, accurate model that can handle nonlinear relationships among features. Random Forest also provides feature importance rankings, helping identify the most influential factors in predicting student failure (Cutler et al., 2012).

Support Vector Machine (SVM): SVM was explored for its ability to handle high-dimensional data and complex decision boundaries. By employing the kernel trick, SVM can effectively classify data that is not linearly separable. This model was beneficial for capturing subtle patterns in the dataset (Kecman, 2005).

K-Nearest Neighbours (KNN): KNN is a simple, yet effective, algorithm for classification tasks. It classifies new instances based on the majority class of the nearest neighbours in the feature space. KNN was included to explore local patterns in student performance and engagement (Zhang, 2016).

Neural Networks (Deep Learning): Neural Networks, including fully connected layers, were tested to capture nonlinear relationships in the data. Deep learning models can learn complex patterns from large datasets, making them well-suited for this predictive task (Van Houdt et al., 2020).

Decision trees: A decision tree classifier is a statistical tool used for clustering and predictions. Its nodes and branches provide a precise flow for classification. In education, decision trees help predict student performance by analysing data like attendance, assessments, assignments, online course completion, test scores, and exam grades. They offer better readability and visualisation than other classifiers (Muraina et al., 2022).

AdaBoost: An ensemble learning technique that iteratively improves the performance of weak classifiers, slightly better than random guessing. It enhances these classifiers by focusing on their errors, assigning higher weights to those that perform better, and combining them to create a strong, accurate model (de Giorgio et al., 2023).

Each of these models was trained on pre-processed and transformed data, and its performance was evaluated using appropriate metrics, including accuracy, precision, recall, F1 Score, and the area under the ROC curve (AUC).

Proposed Model BiLSTM-RNN

In this study, a deep learning model based on the Long Short-Term Memory (LSTM) network is proposed, which is a specialised type of Recurrent Neural Network (RNN) designed for processing sequential and time-series data. This model is particularly suited for handling the time-dependent patterns inherent in student engagement data. The model leverages these temporal patterns, such as how previous interactions and assessment performances influence future outcomes, to predict student performance. The use of LSTM networks for such predictions is advantageous because they are designed to effectively capture long-term dependencies in sequences, making them ideal for modelling student engagement over the course of an academic term or online course. Advantages of this model include:

Handling Sequential Data: One of the primary benefits of LSTM networks is their ability to process and learn from sequences of varying lengths. This is essential for predicting student performance, as students' interactions with course material and their assessment history accumulate over time, impacting future performance.

Long-Term Dependencies: LSTM networks can learn long-term dependencies in data, enabling them to identify how earlier engagement patterns (e.g., participation in the initial modules or early assessments) may significantly influence later outcomes (e.g., final grades or course completion). For example, a student's early struggles might indicate a higher likelihood of failure unless early interventions are implemented, which the model can detect (Van Houdt et al., 2020).

Stability in Learning: Traditional RNNs are often prone to the vanishing gradient problem, in which gradients become too small to contribute to learning, especially with long sequences. LSTMs overcome this issue by utilising a gating mechanism that regulates the flow of information, allowing the model to maintain stable learning even with long sequences of data (Ayinde et al., 2019).

This model is innovative because it focuses exclusively on the predictive power of LSTM-RNN architectures, a technique that has not been widely explored in student performance prediction. While traditional models such as decision trees, logistic regression, or ensemble methods (e.g., Random Forest, XGBoost) are commonly used in many prediction tasks, they are not inherently suited to time-series or sequential data. LSTM-RNNs, on the other hand, are specifically designed to capture sequential dependencies in student behaviour and engagement, making them well-suited to predict how past activities (e.g., participation, performance on early assignments) influence future outcomes, such as final grades or course completion rates. This deep learning-based approach addresses the dynamic nature of online learning, where student behaviour and performance evolve. By focusing on LSTM-RNNs, this study proposes a novel, advanced method to predict student outcomes from temporal engagement data, offering more accurate, timely insights for personalised interventions and improved retention rates.

Results

Modelling results

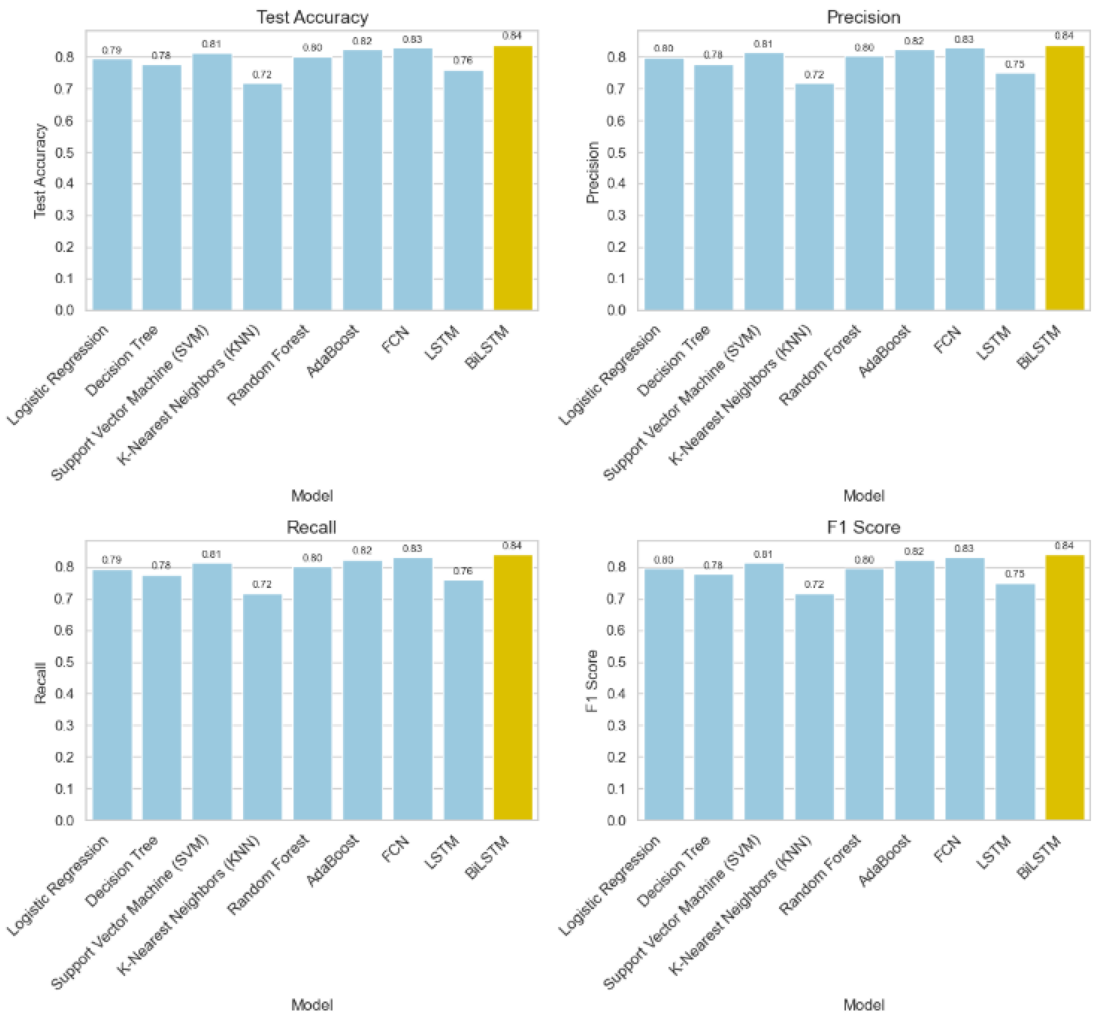
Figure 3 shows the model's results. The models include both traditional machine learning algorithms (Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbours, Random Forest, and AdaBoost) and more advanced deep learning models (FCN, LSTM, and BiLSTM-RNN). Each model's performance is evaluated based on three key metrics: accuracy, precision, recall, and F1 score. The goal of this analysis is to identify which models are best suited for predicting at-risk students and to compare different approaches in terms of predictive power.

Logistic Regression achieved an accuracy of 0.79, with precision, recall, and F1 scores all around 0.80. While these results indicate solid performance, they suggest that Logistic Regression, a simpler linear model, may not fully capture the complex, non-linear patterns in the engagement data. The relatively balanced precision and recall suggest that the model does a fair job at both identifying at-risk students (precision) and minimising false negatives (recall). However, in more complex data sets with intricate feature relationships, more advanced models may outperform Logistic Regression.

The Decision Tree model yielded an accuracy of 0.78, with precision, recall, and F1 scores all at 0.78, indicating similar performance to Logistic Regression. Decision Trees are interpretable, which makes them a popular choice for some applications. However, in this case, the model's performance is slightly lower than that of other more advanced models, such as AdaBoost or BiLSTM-RNN. One reason could be the model's tendency to overfit, especially on complex datasets. While Decision Trees can model nonlinear relationships, they do so by recursively splitting data, which can lead to overfitting on noisy data if not properly pruned.

The Support Vector Machine (SVM) model performed slightly better than both Logistic Regression and Decision Trees, with an accuracy of 0.81. It also achieved a balanced precision, recall, and F1 score of 0.81. SVMs are known for their ability to handle high-dimensional spaces and non-linear decision boundaries. The fact that SVM outperformed both Logistic Regression and Decision Trees suggests it may better capture the complex relationships in the data. However, the SVM model can be computationally expensive, especially with large datasets, and may require careful hyperparameter tuning, such as the kernel function.

Figure 3
Model performance comparison chart



Source: Author's work

K-Nearest Neighbours (KNN) performed worst among the models, with an accuracy of just 0.72. Precision, recall, and F1 scores are also low at 0.72. KNN is a simple, instance-based learning algorithm that classifies data based on the majority class of the nearest neighbours. The poor performance of KNN in this case could be attributed to its inability to capture the temporal dependencies and sequential patterns in the engagement data. Additionally, KNN can struggle with high-dimensional data due to the curse of dimensionality, in which distances between points become less informative as the number of features increases.

Random Forest, an ensemble method that builds multiple decision trees and averages their predictions, performed better than the simpler models, with an accuracy of 0.80. Its precision, recall, and F1 scores were also 0.80, indicating that it can effectively identify at-risk students. Random Forests are less prone to overfitting than individual decision trees because their predictions are averaged, making them a more robust model in many cases. However, despite its performance, Random Forest still lags behind more advanced models, such as AdaBoost and BiLSTM-RNN.

AdaBoost, another ensemble method, showed the highest performance among traditional machine learning models, with an accuracy of 0.82 and precision, recall, and F1 scores of 0.82. AdaBoost combines multiple weak learners (often decision trees) to create a strong learner by adjusting the weights of incorrectly classified instances. The model's performance suggests that it is particularly effective at identifying at-risk students, as it can focus on the harder-to-classify instances. Despite this, AdaBoost still does not match the performance of deep learning models such as FCN and BiLSTM-RNN, which are better suited to handling complex sequential data.

FCN (Fully Connected Network) achieved an accuracy of 0.83, with precision, recall, and F1 scores also at 0.83. This deep learning model uses fully connected layers to process input features and is particularly suited to learning non-linear relationships. FCN’s solid performance suggests that it is a capable model for predicting student performance, capturing more complex patterns in the data than traditional machine learning models. However, while FCN outperforms most traditional models, it still lags behind the BiLSTM-RNN model, which is explicitly designed to capture sequential dependencies in time-series data.

LSTM (Long Short-Term Memory), a type of recurrent neural network (RNN), achieved an accuracy of 0.76, with precision, recall, and F1 scores of 0.75. While LSTMs are particularly well-suited for handling sequential data and capturing long-term dependencies, this model's performance is not as strong as some of the other models, such as BiLSTM-RNN or FCN. This lower performance may be due to insufficient training data, inadequate model tuning, or the LSTM's inability to fully capture the temporal dynamics of the engagement data. Despite its potential, the LSTM model may require further optimisation to improve its performance in this specific application.

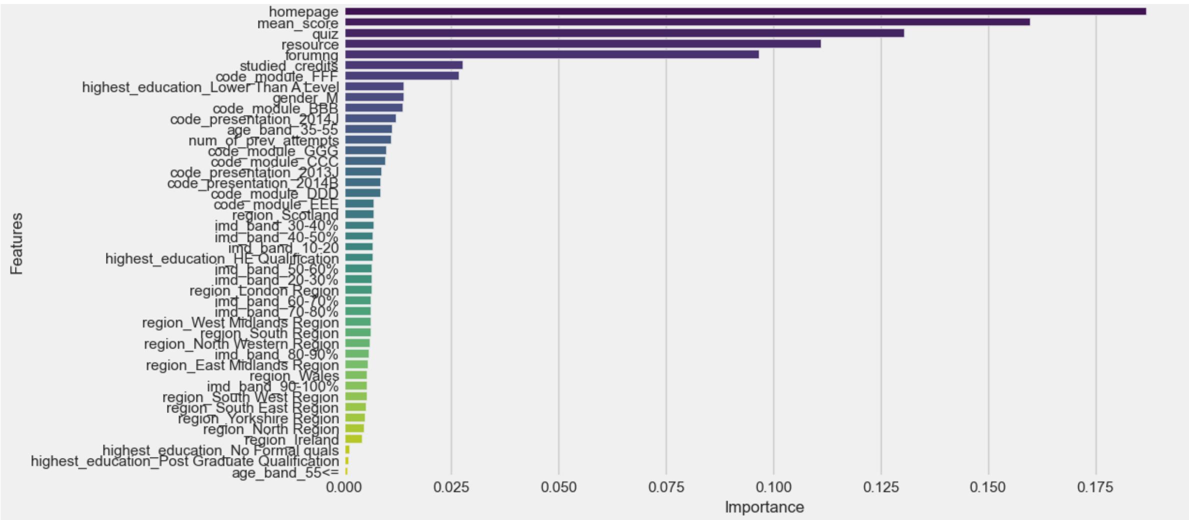
BiLSTM-RNN (Bidirectional LSTM-RNN) model, which combines bidirectional LSTMs with a recurrent neural network architecture, achieved the highest accuracy among all models, with a score of 0.84. Precision, recall, and F1 scores were all 0.84, indicating the model’s ability to predict student performance accurately. The bidirectional nature of the LSTM enables the model to process input sequences in both forward and backward directions, capturing past and future context, which is particularly beneficial in time-series data where context from both ends of a sequence can influence predictions. The high performance of BiLSTM-RNN in this case suggests that it is well-suited to predicting student success based on engagement patterns, providing more accurate insights than simpler machine learning models and other deep learning architectures such as LSTM. The data from the performance of the applied models, particularly the LSTM-RNN model, confirm the first hypothesis of this study and demonstrate that ML/DL approaches are suitable and highly effective for predicting student performance in a virtual learning environment.

Evaluation and interpretation

Once the models were trained, they were evaluated to assess their predictive performance and interpretability. The evaluation process involved comparing the results of different models using metrics such as accuracy, precision, recall, and F1-score. Model interpretability was assessed to ensure that the factors influencing student failure were understandable and actionable. Performance metrics were calculated using cross-validation to ensure the models generalise well to unseen data. The final model selection was based on the trade-off between accuracy and interpretability, ensuring the chosen model provided reliable predictions while being understandable to educators and decision-makers. The insights derived from the evaluation were used to fine-tune the models and identify areas where intervention could help at-risk students. The final model was then used to provide actionable insights to inform early interventions that prevent student failure in online courses.

From Figure 4, it can be observed that the features *homepage*, *mean_score*, *quiz*, *resource*, *forumng*, *student_credits*, and *code_module* are identified as the most significant predictors of student performance in online learning environments. These features provide crucial insights into the factors influencing student engagement and outcomes in the dataset. In this case, the second hypothesis has been confirmed, indicating that certain variables or attributes have a greater impact on predicting student performance in a virtual learning environment than others.

Figure 4
Feature importance



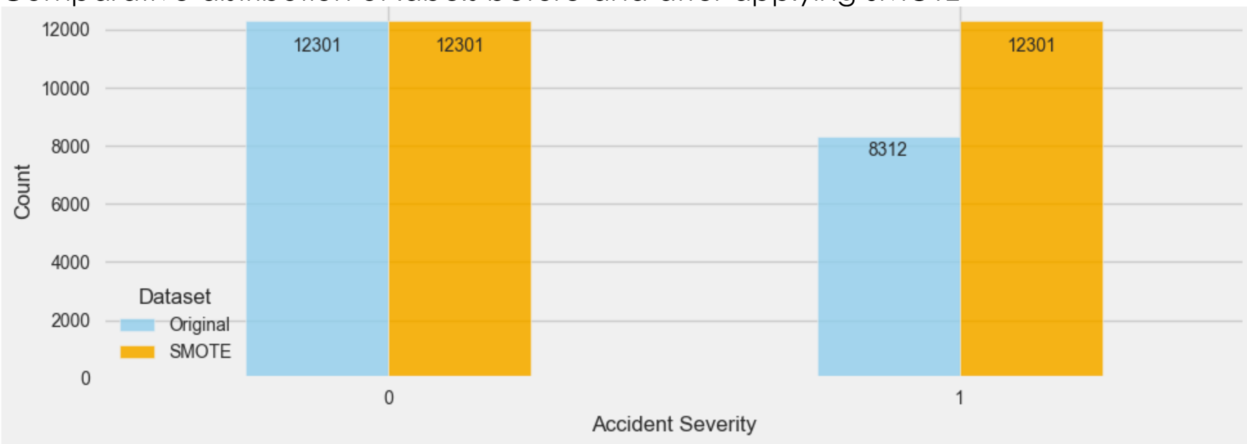
Source: Author's work

Critical evaluation

The results suggest that deep learning models, especially BiLSTM-RNN and FCN, outperform traditional machine learning algorithms in predicting students' performance from engagement data. These models are particularly effective at capturing the non-linear and sequential relationships that are inherent in time-series data, such as engagement patterns and their influence on future performance. The strong performance of BiLSTM-RNN highlights the importance of capturing temporal dependencies and utilising both past and future context in predicting at-risk students. On the other hand, traditional models such as Logistic Regression, Decision Trees, and SVMs perform reasonably well but consistently lag behind the advanced models. These models may be more appropriate for simpler problems where feature relationships are less complex, but they struggle to fully capture the dynamic and sequential nature of engagement data. Despite the strong performance of BiLSTM-RNN and FCN, these models come with their own challenges, including increased computational complexity and the need for more extensive training data. Additionally, while deep learning models can capture complex relationships, they may also suffer from issues like overfitting if not properly tuned or regularised.

A common challenge in student performance prediction is the imbalance in class distributions, particularly in underrepresented categories such as "Distinction" or "Fail." Previous studies have often relied on techniques like oversampling or undersampling to address these imbalances. However, these methods have limitations in improving model generalisation. This research utilises the Synthetic Minority Over-sampling Technique (SMOTE), a generative method to synthesise balanced datasets. By generating synthetic samples, SMOTE has enhanced the model's generalisation capability across all performance categories, leading to more reliable predictions for minority classes without compromising overall accuracy. These results support the third hypothesis of this study, indicating that SMOTE provides a more effective solution to class imbalance than traditional techniques (Figure 5).

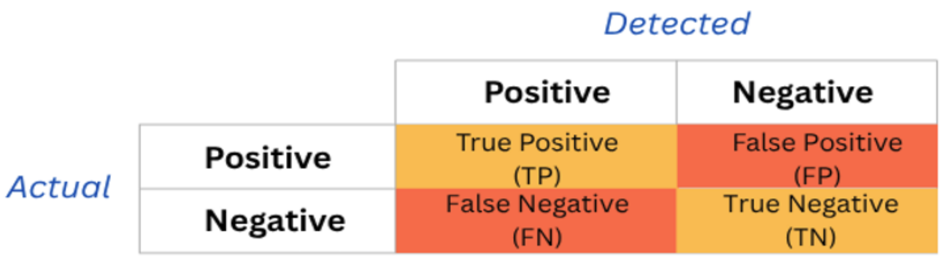
Figure 5
Comparative distribution of labels before and after applying SMOTE



Source: Author's work

The effectiveness of the models was assessed using four evaluation metrics: accuracy, F-score, precision, and recall. These metrics were derived from the binary classification confusion matrix, depicted in Figure 6 and mathematically defined in Equations (1) to (4) (Evangelista, 2021).

Figure 6
The confusion matrix



Source: Authors' work

Precision reflects the proportion of correctly identified positive cases among all cases the model predicted as positive, while recall represents the proportion of actual positive cases correctly identified. All evaluation metrics – accuracy, precision, recall, and F-measure – range from 0 to 1, with higher values indicating better model performance. Accuracy indicates the overall likelihood that a randomly chosen instance will be classified correctly, regardless of its actual class. The F-measure, on the other hand, balances precision and recall through their weighted harmonic mean. In this study, both accuracy and F-measure are used to assess model performance, identifying the model that achieves the highest accuracy and an F-measure approaching 1.

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$F_{1=} \frac{2x(Recall\ x\ Precision)}{Recall+Precision} \tag{4}$$

The results of this study provide insights into the effectiveness of different machine learning models for predicting student performance (Table 1). It was found that advanced deep learning models, particularly the BiLSTM-RNN, offer the highest predictive performance and should be prioritised in applications where time-series data play a critical role. However, traditional machine learning models, such as AdaBoost and SVM, may still be helpful in specific scenarios where interpretability and computational efficiency are of greater concern.0.78

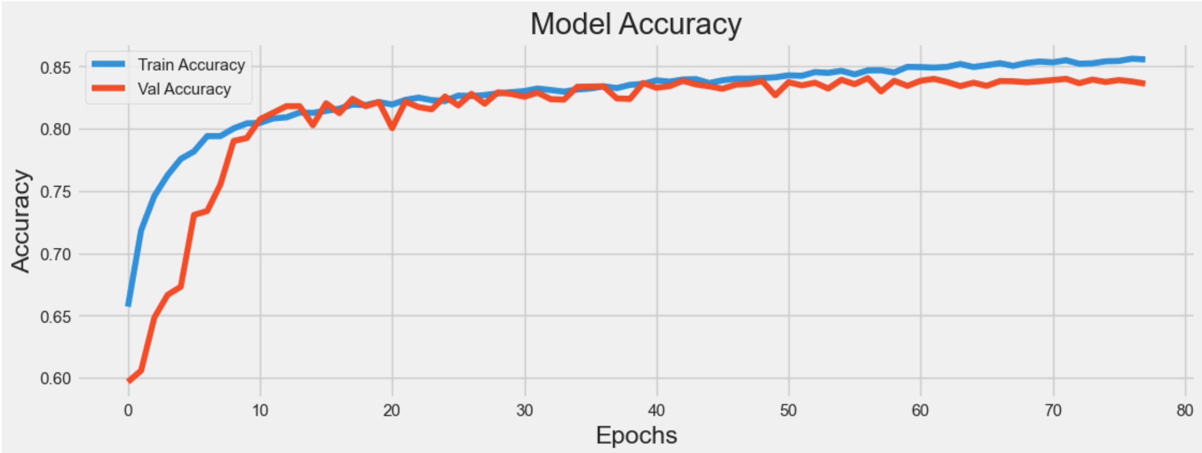
Table 1
Model Performance Comparison

Model	Test accuracy	Precision	Recall	F1 score
Logistic Regression	0.79	0.80	0.79	0.80
Decision Tree	0.78	0.78	0.78	0.78
Support Vector Machine	0.81	0.81	0.81	0.81
K-Nearest Neighbours	0.72	0.72	0.72	0.72
Random Forest	0.80	0.80	0.80	0.80
AdaBoost	0.82	0.82	0.82	0.82
FCN	0.83	0.83	0.83	0.83
LSTM	0.76	0.75	0.76	0.75
BiLSTM-RNN	0.84	0.84	0.84	0.84

Source: Author's work

The graph in Figure 7 displays the model's training and validation accuracy over multiple epochs, showing how well it learns from the training data and generalises to the validation data. The Train Accuracy curve steadily increases over time, indicating that the model is progressively learning to classify the training data more accurately. The Val Accuracy curve follows a similar upward trend, suggesting the model performs well on unseen validation data. There is no noticeable gap between the training and validation accuracies, indicating the model is not overfitting. Both curves rise together, showing that the model generalises effectively, improving performance on both the training and validation sets without memorising the training data. This demonstrates that the model is learning stably and effectively.

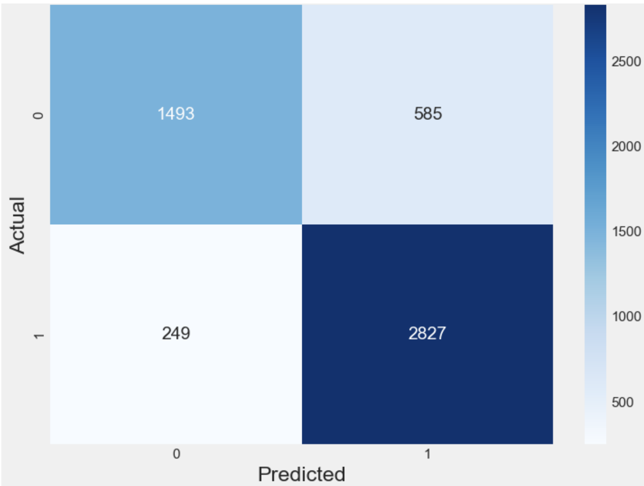
Figure 7
Learning Curve of Bi-LSTM-RNN Model



Source: Author's work

The confusion matrix plot (Figure 8) illustrates the model's classification performance, showing the counts of true positives, true negatives, false positives, and false negatives.

Figure 8
Confusion Matrix of Bi-LSTM-RNN Model



Source: Authors' work

The plot shows the distribution of predicted versus actual labels, with the diagonal elements representing correct classifications (true positives and true negatives), and the off-diagonal elements indicating misclassifications (false positives and false negatives). In this case, the model demonstrates a strong performance, with a high number of correct predictions (diagonal elements) relative to incorrect predictions (off-diagonal elements). This confirms that the model accurately distinguishes between the classes. The relatively low number of misclassifications further suggests that the model is effective in making predictions, with minimal confusion between classes. Overall, the confusion matrix indicates that the model is performing well, with balanced classification accuracy across both classes.

Research limitations

Although the BiLSTM-RNN performed well in terms of predictive accuracy, one challenge with deep learning models is their lack of interpretability. In educational settings, it is crucial to understand the reasons behind a model's prediction, especially when making decisions about student interventions. Future work should explore methods to improve the interpretability of deep learning models, such as using attention mechanisms or explainable AI techniques, to provide insights into why certain students are at risk. Additionally, the results indicated that regularisation techniques such as dropout and L2 regularisation helped prevent overfitting, especially in the BiLSTM-RNN. Future models should focus on model regularisation and continue to utilise regularisation techniques to ensure generalizability, particularly when dealing with large datasets or complex models. Additionally, using learning rate schedules, such as the cosine annealing scheduler, helped maintain model performance across epochs. This approach should be further explored in future work to optimise training processes.

Although the BiLSTM-RNN model demonstrated strong performance, it is important to continuously evaluate and update the model as new data becomes available. Online learning environments are dynamic, and student behaviour can change over time. Periodic model retraining with fresh data can help maintain the model's relevance and accuracy. Additionally, implementing a continuous feedback loop to monitor model predictions and intervention outcomes could help refine the prediction process and intervention strategies.

Conclusion

Summary of findings

As the fields of educational data mining and learning analytics continue to evolve, integrating more diverse data types, developing interpretable models, and continuously evaluating and refining prediction systems will be critical for improving educational outcomes. The future of student performance prediction holds strong potential for improvement, and continued research and innovation will help shape online education, offering personalised learning experiences that cater to the needs of all students.

This study examined the development and evaluation of machine learning and deep learning models for predicting student performance in online courses, with a particular focus on early identification of at-risk students to support timely, targeted interventions. A range of models – including Logistic Regression, Decision Trees, Random Forests, Fully Connected Networks (FCNs), LSTMs, and the BiLSTM-RNN – were assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score.

The experimental findings confirmed substantial performance differences among the models, with the BiLSTM-RNN significantly outperforming all others. Achieving an accuracy, precision, recall, and F1-score of 84%, the BiLSTM-RNN demonstrated its superior ability to capture temporal dependencies in sequential data. Unlike traditional models such as K-Nearest Neighbours or Logistic Regression, which failed to account for the ordered nature of student interactions, the BiLSTM-RNN's bidirectional architecture enabled it to process input sequences in both forward and backward directions. This enabled a deeper understanding of how past engagement influences future outcomes, reinforcing its effectiveness in educational time-series prediction tasks.

The first hypothesis (H1) proposed that student engagement levels can be accurately predicted using ML/DL algorithms applied to VLE data. The findings have strongly supported this. Among all models evaluated, the BiLSTM-RNN achieved the highest predictive performance, with an accuracy of 0.84, and corresponding precision, recall, and F1-scores also reaching 0.84. The superior performance of the BiLSTM-RNN model can be attributed to its bidirectional structure, which enables the model to capture both past and future context in time-series data—an essential characteristic for modelling student behaviour over time. These results demonstrate that advanced deep learning models, particularly those tailored for sequential data, offer substantial benefits for educational prediction tasks and can outperform traditional ML/DL architectures.

The second hypothesis (H2) posited that early detection of student engagement could be reliably achieved by leveraging key interaction variables from the VLE. This was validated using a Random Forest Classifier, which identified the most influential features in the dataset. The analysis revealed that variables such as *homepage*, *mean_score*, *quiz*, *resource*, *forumng*, *student_credits*, and *code_module* played a critical role in predicting student outcomes. These findings affirm that not all features contribute equally to engagement prediction; rather, specific interaction types and demographic indicators serve as strong early predictors. The confirmation of this hypothesis underscores the value of feature importance analysis in developing interpretable and practical educational models.

The third hypothesis (H3) posits that over-sampling techniques, particularly SMOTE, are more effective than traditional methods for addressing class imbalance in classification tasks. The experimental results confirmed this claim. By applying SMOTE to balance the dataset, the models demonstrated improved generalisation across all performance categories, including underrepresented ones (e.g., “Distinction” and “Fail”). Unlike conventional over- or under-sampling methods, which often risk overfitting or data loss, SMOTE can enhance predictive performance for minority classes without degrading overall model accuracy. This reinforces the importance of advanced sampling strategies in handling skewed educational datasets.

The integration of robust deep learning architectures, insightful feature selection, and advanced data balancing techniques enabled accurate, interpretable predictions of student performance. Beyond empirical validation, the findings of this study contribute to both theory and practice. Theoretically, they reinforce the applicability of deep learning models for sequential data in educational contexts. In practice, they provide a foundation for building intelligent early-warning systems that can identify at-risk students with high precision. Such systems can play a critical role in enhancing student retention and optimising support services in online learning environments.

Theoretical contributions

This study has demonstrated the potential of machine learning models, particularly LSTM-based architectures, to predict student performance in online courses. By leveraging sequential data, such models can identify at-risk students early, enabling timely interventions that can improve student success. Unlike traditional approaches that rely on static point estimates, advanced ML and DL techniques offer more dynamic and adaptable predictive models. This research contributes to the field of learning analytics by leveraging ML and DL methods to more accurately and reliably predict at-risk students. Additionally, the study addresses the challenge of highly imbalanced student outcome distributions by employing advanced methods that enhance the model's ability to identify and predict minority classes effectively.

The theoretical contribution of this research emphasises the importance of incorporating sequential data into predictive models. Student engagement and performance often exhibit temporal dependencies, and models such as LSTMs that capture these patterns consistently outperform traditional approaches (Liu et al., 2022; Ding et al., 2019). Unlike earlier studies, this research integrates both demographic and engagement-related features. Future improvements may include real-time feedback, peer interactions, and forum participation, which could offer deeper insights into student behaviour and further enhance predictive power.

Practical implications

The findings of this study are anticipated to have a significant impact on both research and practice, leading to more effective and targeted interventions in online education environments. A key outcome of this study is the development of a system for identifying at-risk students. Predictive models can serve as a foundation for personalised interventions, such as offering tutoring, personalised feedback, or reminders to engage with course materials, which could significantly improve student retention and academic success. To implement this study's findings, machine learning models, particularly LSTM-based architectures, can be integrated into online learning platforms to predict at-risk students using sequential data like engagement and performance history. These models offer dynamic predictions that outperform traditional methods, even when student outcomes are imbalanced. The system would collect diverse data, including real-time feedback and peer interactions, to enhance predictive accuracy. At-risk students would receive personalised interventions, such as tutoring or reminders to engage with course content, improving retention and success. Ongoing evaluations and updates would ensure the models adapt to changing student behaviour. By continuously refining these models and incorporating more data, the system would provide personalised learning experiences, ultimately improving educational outcomes in online education.

Research limitations and future research directions

Although the BiLSTM-RNN model demonstrated strong predictive performance, several limitations should be acknowledged. First, the study relies on the OULAD dataset, which represents a single institution and time period; consequently, the generalisability of findings to other educational contexts remains limited. Replication with diverse datasets and learning platforms would strengthen external validity. Second, despite the use of deep learning methods, model interpretability remains a challenge. While attention mechanisms improve transparency, deep learning models still function primarily as black boxes, which may hinder adoption in educational settings where explainability is essential for justifying interventions. Future work should explore explainable AI techniques to provide clearer insights into model decisions. Third, although SMOTE effectively addressed class imbalance, synthetic oversampling may not fully preserve the complexity of real student behaviour. Alternative rebalancing approaches, such as adaptive sampling or generative models, may offer improvements. Finally, this study focuses primarily on structured numerical and behavioural data; integrating additional modalities such as textual forum content, real-time interaction logs, or affective indicators could yield richer models and more nuanced predictions. Future research should also examine real-time prediction frameworks and longitudinal evaluation of intervention outcomes to ensure that predictive models remain accurate, adaptive, and impactful in evolving online learning environments.

References

1. Alnassar, F. M. (2023). Predicting student performance in a virtual learning environment. Doctoral thesis. Goldsmiths, University of London.
2. Alruwais, N., & Zakariah, M. (2023). Student-engagement detection in the classroom using a machine learning algorithm. *Electronics, 12*(3), 731. https://doi.org/10.3390/electronics12030731
3. Ayinde, B. O., Inanc, T., & Zurada, J. M. (2019). Regularizing Deep Neural Networks by Enhancing Diversity in Feature Extraction. *IEEE Transactions on Neural Networks and Learning Systems, 30*(9), 2650-2661. https://doi.org/10.1109/tnnls.2018.2885972
4. Bittner, A. (2022). Analysis-of-variance (ANOVA) Assumptions Review: Normality, Variance Equality, and Independence. Proceedings of the XXXIVth Annual International Occupational Ergonomics and Safety Conference. https://doi.org/10.47461/isoes.2022_bittner
5. Brahim, G. B. (2022). Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features. *Arabian Journal for Science and Engineering, 47*(8), 10225-10243. https://doi.org/10.1007/s13369-021-06548-w
6. Chaplot, D. S., Rhim, E., & Kim, J. (2015). Predicting student attrition in MOOCs using sentiment analysis and neural networks. In *Workshops at the 17th International Conference on Artificial Intelligence in Education* (AI, Spain). https://ceur-ws.org/Vol-1432/islg_pap2.pdf
7. Choi, S., Lam, S., Li, K. C., & Wong, B. T.-M. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Educational Technology & Society, 21*(2), 273–290. https://www.jstor.org/stable/26388407
8. Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. *Ensemble Machine Learning*, 157-175. https://doi.org/10.1007/978-1-4419-9326-7_5
9. de Giorgio, A., Cola, G., & Wang, L. (2023). Systematic review of class imbalance problems in manufacturing. *Journal of Manufacturing Systems, 71*, 620-644. https://doi.org/10.1016/j.jmsy.2023.10.014
10. Ding, M., Yang, K., Yeung, D.-Y., & Pong, T.-C. (2019). Effective Feature Learning with Unsupervised Learning for Improving the Predictive Models in Massive Open Online Courses. Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 135-144. https://doi.org/10.1145/3303772.3303795
11. Evangelista, E. (2021). A hybrid machine learning framework for predicting students' performance in virtual learning environment. *International Journal of Emerging Technologies in Learning (iJET), 16*(24), 255–272. https://doi.org/10.3991/ijet.v16i24.26151
12. Geigle, C., & Zhai, C. (2017). Modeling MOOC Student Behavior With Two-Layer Hidden Markov Models. Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale, 205-208. https://doi.org/10.1145/3051457.3053986
13. Hafzan, M. Y. N. N., Safaai, D., Asiah, M., Saberi, M. M., & Syuhaida, S. S. (2019). Review on predictive modelling techniques for identifying students at risk in university environment. In *Engineering Application of Artificial Intelligence Conference 2018 (EAAIC 2018)*. https://doi.org/10.1051/matecconf/201925503002
14. Hilbe, J. M. (2015). *Practical guide to logistic regression*. CRC Press, Taylor & Francis Group. https://doi.org/10.1201/b18678
15. Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. *Computational Intelligence and Neuroscience, 2018*, 1-21. https://doi.org/10.1155/2018/6347186
16. Inder, S. (2021). Factors Influencing Student Engagement for Online Courses: A Confirmatory Factor Analysis. *Contemporary Educational Technology, 14*(1), ep336. https://doi.org/10.30935/cedtech/11373
17. Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability, 14*(22), 14795. https://doi.org/10.3390/su142214795
18. Kecman, V. (2005). Support Vector Machines - An Introduction. *Studies in Fuzziness and*

Soft Computing, 1-47. https://doi.org/10.1007/10984697_1

19. Kuzilek, J., Hlosta, M., & Zdráhal, Z. (2017). Open University learning analytics dataset. *Scientific Data, 4*(1). https://doi.org/10.1038/sdata.2017.171
20. Lee, S., & Chung, J. Y. (2019). The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. Applied Sciences, 9(15), 3093. https://doi.org/10.3390/app9153093
21. Liu, H., Zhu, Y., Zang, T., Xu, Y., Yu, J., & Tang, F. (2022). Jointly modeling heterogeneous student behaviors and interactions among multiple prediction tasks. *ACM Transactions on Knowledge Discovery from Data (TKDD), 16*(1), 1-24. https://doi.org/10.1145/3458023
22. Marban, O., Mariscal, G., & Segovia, J. (2012). A data mining & knowledge discovery process model. In *Data mining and knowledge discovery in real life applications,* 18. https://doi.org/10.5772/6438
23. Muraina, I. O. O., Aiyegbusi, E. A., & Abam, S. O. (2022). Decision tree algorithm use in predicting students' academic performance in advanced programming course. *International Journal of Higher Education Pedagogies, 3*(4), 13–23. https://doi.org/10.33422/ijhep.v3i4.274
24. Palani, K., Stynes, P., & Pathak, P. (2021). Clustering Techniques to Identify Low-engagement Student Levels. Proceedings of the 13th International Conference on Computer Supported Education. https://doi.org/10.5220/0010456802480257
25. Peach, R. L., Yaliraki, S. N., Lefevre, D., & Barahona, M. (2019). Data-driven unsupervised clustering of online learner behaviour. *npj Science of Learning, 4*(1). https://doi.org/10.1038/s41539-019-0054-0
26. Soobramoney, R. (2021). Early prediction of students at risk in a virtual learning environment using ensemble machine learning techniques (Doctoral thesis). Durban University of Technology. https://doi.org/10.51415/10321/4072
27. Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review, 53*(8), 5929-5955. https://doi.org/10.1007/s10462-020-09838-1
28. Xing, W., & Du, D. (2019). Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *Journal of Educational Computing Research, 57*(3), 547-570. https://doi.org/10.1177/0735633118757015
29. Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine, 4*(11), 218-218. https://doi.org/10.21037/atm.2016.03.37

About the authors

Lediana Shala Riza is a PhD candidate at the Faculty of Contemporary Sciences and Technologies, South-East European University, North Macedonia. She received her B.S. degree in Computer Engineering from the University of Prishtina in 2018 and her M.S. degree in Subject Teaching with Specialisation in Technology & ICT from the same university in 2021. Online assessment, e-teaching & e-learning, and educational technology are among her research interests. She has written or co-written papers on ICT and education at international conferences and journals. The author can be contacted at **ls31402@seeu.edu.mk.**

Lejla Abazi Bexheti is a Full Professor in Computer Science with extensive experience in academia, research, and quality assurance in higher education. She is affiliated with SEEU University, where she actively contributes to teaching, research, and institutional quality improvement. Her expertise spans e-learning, artificial intelligence, and digital humanities, with a focus on leveraging technology to enhance educational practices. Prof. Abazi Bexheti is a dedicated educator and researcher with a strong commitment to fostering innovation, excellence, and collaboration in higher education and is highly active in international research projects. The author can be contacted at **l.abazi@seeu.edu.mk.**

Jovana Zoroja, PhD, is an Associate Professor at the Faculty of Economics and Business, University of Zagreb, Department of Informatics. She received a PhD in Information Systems at the Faculty of Economics and Business Zagreb with the dissertation thesis "Influence of the ICT on the Competitiveness of the EU Countries". She was also educated at the LSE Summer School in London in Business Development and ICT Innovation. Her main research interests are information and communication technology, e-learning, simulation games and simulation modelling. She is actively engaged in several scientific projects (FP7-ICT, bilateral cooperation, and national projects). The author can be contacted at **jzoroja@efzg.hr.**