DOI: 10.15849/IJASCA.251130.06

Int. J. Advance Soft Compu. Appl, Vol. 17, No. 3, November 2025
Print ISSN: 2710-1274, Online ISSN: 2074-8523

AI-Based Prediction of Student Dropout Using an Enhanced Deep Learning Framework: DeepDropNet Model

Oussama Hamal[1], Nouhaila El Koufi[2], Korhan Cengiz[3], Luka Milić[4]

[1] National School of Architecture (ENA), Marrakech, Morocco
e-mail: hamal.oussama@gmail.com
[2]Laboratory of Mathematical Modeling and Economic Calculations, FEG, Morocco
e-mail:Elkoufi.nouhaila1@gmail.com
[3]Department of Electrical Engineering, Prince Mohammad Bin Fahd University, Al Khobar 31952, Saudi Arabia
e-mail: kcengiz@pmu.edu.sa
[4]Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, 42000 Varaždin, Croatia
e-mail : luka.milic@foi.unizg.hr

Abstract

In the field of online and higher education, student dropout remains a significant challenge, as the success rate and reputation of educational institutions are closely linked to student retention. In increasingly competitive academic environments, institutions must understand the factors that lead to student disengagement and withdrawal. Student dropout not only affects the individual learner's academic journey but also impacts institutional metrics such as graduation rates and funding. Numerous studies have been conducted to predict student dropout; however, evaluations of these methods indicate considerable scope for improvement. Therefore, we propose in this research a new student dropout prediction architecture, DeepDropNet, designed to forecast dropout risk in educational settings. Our proposed DeepDropNet combines a 1D convolutional layer with residual blocks, squeeze-and-excitation blocks, and a spatial attention module to improve predictive accuracy. The residual blocks help address the vanishing gradient problem, while the squeeze-and-excitation and spatial attention components enable the model to capture complex dependencies across temporal and behavioral features. To evaluate the model's performance, we conduct experiments on two real-time educational datasets. Given the common issue of class imbalance in dropout prediction tasks, we employ two data balancing techniques: SMOTE, SMOTEEN, and SMOTETomek. Using 10-fold cross-validation, we thoroughly assess the model, and the results show that DeepDropNet outperforms existing approaches, achieving 94.68% and 96.41% accuracy across the two datasets used.

Keywords: *Artificial Intelligence, Deep Learning, Education, DeepDropNet, Student Dropout, Learning Management System.*

Received 15 July 2025; Accepted 22 October 2025

1 Introduction

The last ten years have seen a radical transformation in the landscape of higher education and distance learning through faster digitization, increased accessibility, and rising enrollments [1]. However, this growth has come with rising concern about student dropout, which is a persistent issue for higher education institutions. Excessive rates of dropout among students not only negatively affect the learning performance of the learners but also institutional performance, budget, and image [2]. The detection of students at risk of dropping out early is necessary to choose the right intervention**.** This intervention provides good outcomes in terms of maximizing retention and success [3].

Various factors influence the decision to drop out among students, making this phenomenon complex. To address and resolve this issue, traditional machine learning techniques, such as SVM and Random Forest, have been utilized. These models are poor at dealing with high-dimensional, non-linear data and depend heavily on manual feature engineering, which is time-consuming and error-prone. Additionally, they lack inherent ways of showing which features contribute or have the greatest contributions to dropout risk, which limits both predictive performance and interpretability.

To address these limitations, deep learning methods have been adopted by recent studies and research, as they allow for automatic feature learning and the ability to model relationships. Complex relationships. However, many existing deep learning models struggle to identify significant predictive features because they lack attention mechanisms that can focus on important patterns in the data. Attention modules are particularly valuable in educational contexts, where both temporal and feature-wise dependencies are critical for understanding why students disengage.

We have proposed a new deep learning architecture called DeepDropNet to bridge these gaps. 1D convolutional layers with residual connections, squeeze-and-excitation (SE) blocks, and spatial attention mechanisms were used in this architecture. It enhances the model's ability to learn and highlight significant characteristics. Due to the high degree of class imbalance in dropout datasets, we employ three oversampling techniques: SMOTE, SMOTE-Tomek Links, and SMOTE with edited nearest neighbors (SMOTEENN), with SMOTEENN yielding the best results. We train and test the model with 10-fold cross-validation on two real student datasets, achieving strong performance across all measures.

The major contributions of this paper are as follows:

- We present DeepDropNet, a novel attention-based deep learning model for student dropout prediction. It combines 1D CNNs, residual learning, SE blocks, and spatial attention to enhance feature representation and classification performance.
- To address the class imbalance problem, we employ and compare three oversampling techniques (SMOTE, SMOTETomek, and SMOTEENN), finding that SMOTEENN delivers superior performance.
- Our model achieved high accuracy rates through extensive experimentation and hyperparameter optimisation on two real-world education datasets, with results of 94.6%, 95.82%, and 96.63%, surpassing traditional machine learning state-of-the-art and current deep learning techniques.

2 Related Work

Student dropout prediction has become an important subject of interest among both the educational data mining and learning analytics research communities [4]. Machine learning, deep learning, and multimodal strategies based on more than one method have been applied to analyze this problem in numerous different studies. Traditional machine learning algorithms such as Logistic Regression (LR), RF, SVM, Naive Bayes (NB), K-Nearest Neighbors (KNN), and DT are commonly employed to classify students into dropout or non-dropout [5], [6]. Although these models are computationally efficient and provide baseline insights, they generally struggle to capture the complex, nonlinear dependencies inherent in students' learning behaviors and longitudinal academic histories. As a result, their predictive performance tends to plateau when confronted with high-dimensional or heterogeneous data [7], [8].

Some researchers have called upon data mining methods such as clustering (e.g., K-Means, DBSCAN) and association rule learning (e.g., Apriori, FP-Growth) to uncover latent patterns of student disengagement [9]. While these methods can provide descriptive insights into dropout risk factors, they are largely exploratory and lack robustness for reliable prediction in real-world academic settings [10]. In order to counter the limitations of traditional classifiers, deep learning models have increasingly been incorporated in current research. Artificial Neural Networks (ANN) based methods, Multilayer Perceptron (MLP) based methods, and Recurrent Neural Networks (RNN) based methods have been shown to work better since they are able to learn feature representations automatically [11], [12]. Other studies have attempted to use Convolutional Neural Networks (CNNs) for temporal and behavior feature extraction [13], while Bidirectional LSTM networks have been used in sequential student data modeling [14]. However, these models are not without shortcomings: CNNs often fail to preserve long-range dependencies, RNNs are prone to vanishing gradient issues, and even attention-based hybrid models can be sensitive to noisy features, reducing generalizability [15], [16].

The class imbalance problem, also common in dropout datasets (due to the rarity of positive dropout cases), has also been widely explored. Certain oversampling methods, such as SMOTE, ADASYN, and MWMOTE, have been proposed by researchers, typically together with classifiers for balancing training data and improving model generalization [17]. More advanced variants like SMOTEEN and SMOTETomek have been shown effective in academic settings, but their integration with deep learning architectures often remains fragmented rather than fully embedded into the model design.

Other work has focused on building comprehensive frameworks for dropout prediction, student profiling, and early warning systems. These frameworks also cover cross-validation of performance for reliability and testing on various datasets to determine generalizability. There are some studies that have employed feature selection techniques such as mutual information, correlation-based filtering to reduce dimensionality and underscore important predictors [18], [19]. However, such frameworks often remain dataset-specific and may not scale across diverse educational contexts.

Attention-based deep learning models have come under the spotlight recently because they have the ability to place weights of importance on features and increase interpretability. For example, applying spatial attention or channel-wise recalibration through squeeze-and-excitation (SE) blocks has proved useful in improving the performance of the model for various educational prediction problems [20]. Nevertheless, most attention-based studies still overlook the joint treatment of feature weighting, residual learning, and class imbalance within a unified framework.

Although there have been a number of models proposed for predicting dropout of students, many traditional ML methods are less capable in expressing intricate feature relationships at deep levels, and even some deep models omit the significance of weighting features. This promotes the development of more complex architectures, such as our DeepDropNet, that combine convolutional layers, residual learning, attention, and oversampling methods to enhance the flexibility of dealing with student data complexity and yielding better prediction outcomes.

3 Problem Formulations

The problem of student dropout prediction has been addressed by numerous studies that have proposed a variety of ML algorithms for modeling student disengagement and withdrawal. However, due to the complex, non-linear, and context-dependent nature of dropout behavior, traditional ML models have often been unable to yield high-performance results. Conventional approaches such as support vector machines, random forests, and decision trees typically entail extensive manual feature crafting, which is not only time and resource intensive but also vulnerable to human prejudice. Additionally, these models lack the ability to selectively focus on salient input features, limiting their potential to identify key dropout predictors in high-dimensional data sets. Although deep learning techniques have been explored in newer research, the majority of existing models fail to incorporate attention mechanisms, with the consequence of providing limited interpretability and worse performance when capturing salient behavioral or academic signals in isolation. To bypass these limitations, we propose a novel deep learning model, DeepDropNet, to automatically uncover complex non-linear feature relationships and learn contextual feature importance using inherent attention mechanisms. The architecture uses 1D convolutional layers with residual blocks, squeeze-and-excitation modules, and spatial attention modules, enabling richer feature extraction and predictive capability. Because of the extreme class imbalance in the datasets, we also applied and compared three oversampling techniques; SMOTE, SMOTETomek, and SMOTEENN; with SMOTEENN giving overall best performance. The suggested model was tested using a 10-fold cross-validation procedure, and on two real-world education datasets, DeepDropNet consistently performed better than state-of-the-art approaches, achieving high performances of 94.68% and 96.41% respectively.

4 Materials and Method

Here, we introduce the components of the proposed SDP framework. It comprises six main stages, as clearly shown in Fig. 1. The first step is data collection, where we gathered data from two sources: a public and a private Moroccan university. Next, we performed data preprocessing, which involves key tasks such as removing irrelevant attributes, addressing missing values, and encoding categorical attributes for both datasets. To address the class imbalance issue, three oversampling methods were applied. Finally, a student dropout prediction model was developed. This model underwent optimization, and we employed various evaluation metrics to assess its performance and compare it with other models.

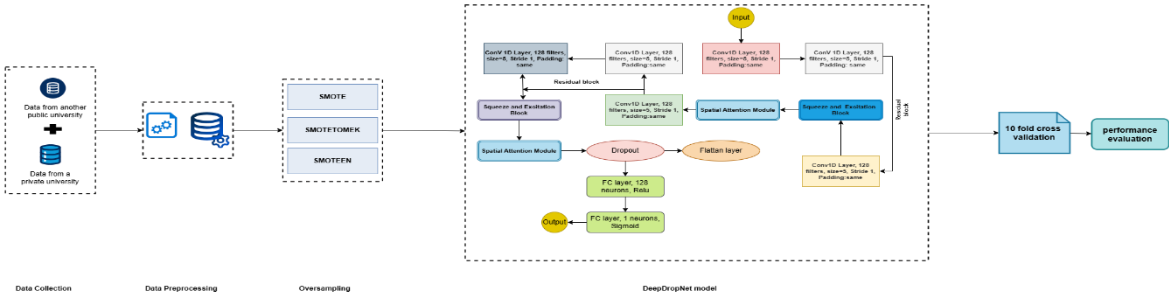

Fig. 1. Proposed System Architecture

4.1 Data collection

In this study, we combined two real-world datasets collected from different Moroccan universities, whose descriptions are provided below. The First dataset comprises 8213 student instances, each representing a student profile. Each instance contains 23 attributes; 22 independent variables and 1 dependent variable (Dropout). Among the independent features, 18 are numerical (e.g., GPA, attendance rate), and 5 are categorical or discrete (e.g., gender, program type). The second dataset includes 6000 student profiles, where each profile is described by 21 variables; 20 independent features and 1 dependent variable (Table 1).

Table 1: Features

Dataset	Records number	Characteristic number
Moroccan public university	8213	23
Moroccan private university	6000	21

4.2 Data preprocessing

During data preprocessing, several steps necessary for that were consistently applied in the two datasets. First, non-informative features such as identifiers (e.g., student ID or phone number) were eliminated because they do not contribute to the learning process. The missing numerical features' values were imputed by zero, while no missing data was encountered in the categorical features. To handle the categorical variables and prepare the data for modeling step, we utilized a mix of Label Encoding and One-Hot Encoding techniques.

4.3 Oversampling

Since the class imbalance is extremely high in all two datasets; where instances of non-dropout far exceed instances of dropout; it was necessary to apply advanced resampling techniques to allow fair model training. To address this, we combined three methods: SMOTE, SMOTETomek, and SMOTEENN. The SMOTE method generates artificial instances of the minority class by interpolating between the existing instances and their nearest neighbors, thereby increasing the dropout class. SMOTETomek combines this artificial oversampling with a cleaning technique called Tomek Links, which removes suspicious instances from the majority class that are closely matched with minority instances and therefore tightens the class border. Finally, SMOTEENN combines SMOTE with edited nearest neighbors (ENN), a filtering technique that eliminates noisy or misclassified instances after oversampling, further improving dataset quality. The resulting balanced datasets are as represented in Fig. 2, and 3, which denote the new distribution after applying each method.

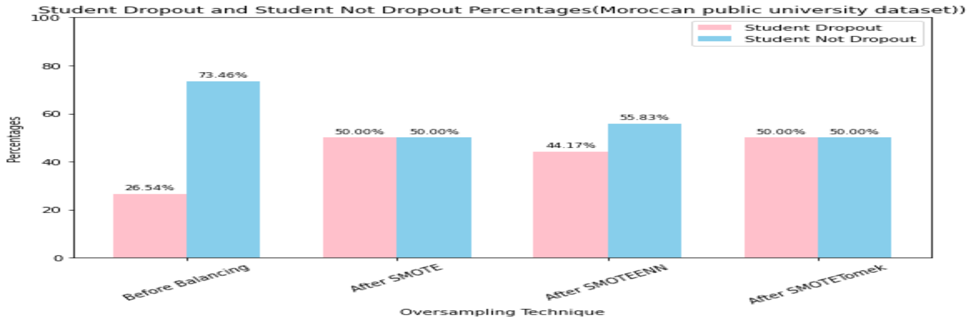

Fig. 2. Data distribution across different oversampling techniques (datasset 1)

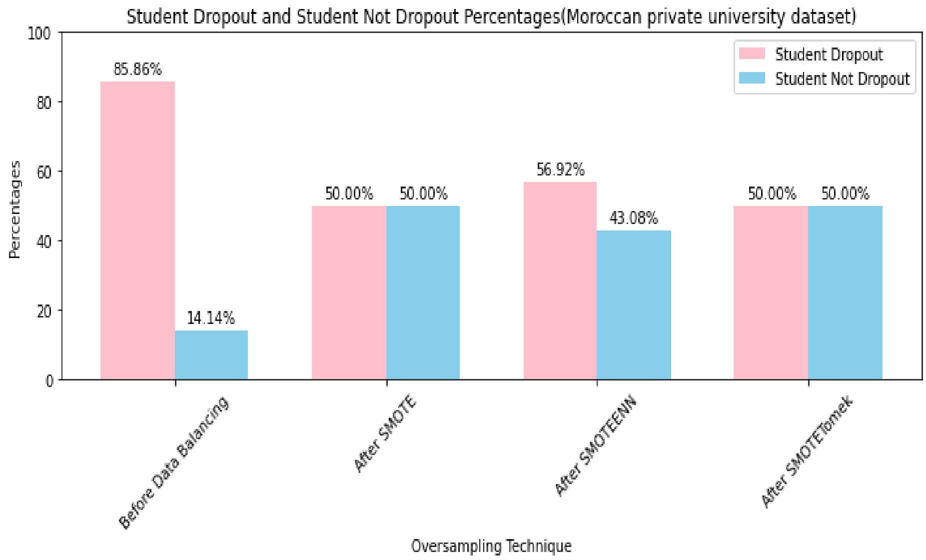

Fig. 3. Data distribution across different oversampling techniques (datasset 2)

4.4 Proposed Model

To accurately forecast student dropout, we introduced a novel deep learning model named DeepDropNet with attention mechanisms for enhanced predictive capabilities. Unlike traditional methods, DeepDropNet is based on a 1D Convolutional Neural Network (CNN) and also strengthened by utilizing residual blocks, squeeze-and-excitation (SE) modules, and spatial attention layers, as illustrated in Fig. 1. The residual blocks allow for increased gradient flow and learning efficiency as they preserve feature mappings layer by layer. The SE and spatial attention modules, meanwhile, learn hard dependencies across and within feature channels so the model will focus on the most significant student features. The architecture design followed extensive experimentation and careful hyperparameters tuning. To our understanding, this represents one of the initial applications of such a paradigm within the arena of student dropout prediction.

DeepDropNet differs from current attention-based CNNs by combining squeeze-and-excitation and spatial attention in a residual learning framework. This dual-attention mechanism, along with dataset-specific hyperparameter tuning and advanced oversampling (SMOTE, SMOTETomek, and SMOTE-ENN), makes the architecture particularly well-suited to the imbalance and diversity of student dropout datasets. These tailored design choices distinguish DeepDropNet from current CNN-based approaches and directly contribute to its enhanced predictive capability.

The DeepDropNet model architecture begins with a 1D Convolutional Neural Network (CNN) that will process sequential student-related features. It is a convolution layer that performs convolution operations, computes dot products between the input and a collection of learnable filters, to find informative local patterns from the input. Specifically, the input tensor with a shape of (1 × W × 1) where W is the number of features in the input per student is passed through a 1D convolutional layer with 128 filters having kernel size of 5, stride of 1, and zero padding to preserve spatial dimensions. This convolutional layer is responsible for generating feature maps that represent lower-dimensional abstractions required for dropout prediction. The resulting tensor (1 × W × 128) from this is then forwarded to subsequent residual learning blocks. Further, down the architecture, the same setup of convolutional layers is used to further refine and extract deeper feature representations.

In order to enhance learning efficiency and broaden the network without being bothered by degradation problems, residual blocks are added to DeepDropNet's structure. Each residual block has two consecutive convolutional layers with configurations mirroring the first CNN layer to facilitate continuous feature extraction. The ReLU activation function is applied after each convolution to introduce non-linearity without affecting computational complexity. A skip connection is established by adding the input of the block directly to the output of the second convolutional layer, forming the expression y = x + G(x). This design solves the vanishing gradient problem by enabling gradients to propagate unimpeded through identity mappings when backpropagating. The output of the residual block remains of shape (1 × W × 128) and is processed by the next squeeze-and-excitation module.

A Squeeze-and-Excitation (SE) block is incorporated within the DeepDropNet architecture to render the model extremely sensitive to the most informative channels. The output of a residual block, U∈R, is first passed through a Global Average Pooling (GAP) operation to compress the information and create a concise channel description spatially. It is given by:

$$\boldsymbol{s_c} = G_{sq}(u_c) = \frac{1}{M}\sum_{j=1}^{M} u_c(j)$$

Where u_c the channel input feature c, and s_c is the scalar sum. The descriptor is fed into an excitation mechanism made up of two fully connected layers: the first one employs a ReLU activation to learn nonlinear interactions, while the second one employs a sigmoid activation to calculate attention weights, which are computed as:

$$\boldsymbol{z} = G_{ex}(z, W) = \sigma(W_2\delta(W_1 s))$$

These weights are used to re-weight the raw input by channel-wise multiplication:

$$\widetilde{\boldsymbol{x_c}} = G_{scale}(u_c, z_c) = z_c u_c$$

Whose output tensor of size 1×W×128 is then fed to the spatial attention module to be further refined.

For further refinement of feature representations, spatial attention module is employed after the SE block to emphasize most salient spatial regions in every feature channel. In this module, the input from the SE block is subjected to average pooling as well as max pooling operations and is designed for extracting complementary details from the provided input tensor. These pooled features are concatenated together to form an aggregated feature descriptor as input to a convolution layer to generate the spatial attention map. The attention map is element-wise multiplied with the input feature tensor to enhance the regions recognized as most important for the task of dropout prediction. This is algebraically captured as:

$$Mk_s(G) = \sigma([AvgPool(G); MaxPool(G)])) = \sigma(g([G^s avg; G^s max])$$

where σ is the sigmoid function and f is a convolution operation. Notably, the output tensor has the same size as the input tensor such that it will be compatible with subsequent layers in the model pipeline.

The spatial attention module's output tensor is passed through a flattening layer, transforming it into a one-dimensional vector for classification. It is protected against overfitting by passing it through a dropout layer of rate 0.5. The vector is then passed through a fully connected dense layer consisting of 128 neurons, where the ReLU activation function introduces non-linearity and helps in learning complex patterns. Then, this intermediate representation is passed through a single-neuron output layer with a sigmoid activation function to produce a binary prediction score. This output value is in the range [0,1], which allows the model to forecast the likelihood of a student dropping out.

During the evaluation phase, we employed a comprehensive set of performance metrics. First, we used accuracy, precision, recall, and the F1-score to provide detailed insights into the classification capability of the model. Next, the ROC-AUC metric was applied to examine the model's ability to distinguish between dropout and non-dropout cases. Finally, the Matthews Correlation Coefficient (MCC) was included, as it offered a particularly valuable measure by delivering a balanced assessment of classification quality, even in the presence of imbalanced class distributions.

4.5 Evaluation Metrics

As the model's performance evaluation is a crucial phase to illustrate its efficiency, we have integrated a set of efficient evaluation measures (Table 2) [21][22][23].

Table 2: Evaluations metrics

Metric	***Methods***
Accuracy	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	$Pr = \frac{TP}{TP + FP}$

Metric	*Methods*
F1-score	$Fs = \frac{(2 * Pr * Rec)}{(Rec + Pr)}$
Recall	$Rec = \frac{TP}{TP + FN}$
Matthews Correlation Coefficient	$MC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TN + FP)(TP + FN)(TN + FN)}}$

5 Results, Analysis and Discussions

5.1 Case Study

We have employed real data gathered from two Moroccan universities to validate the proposed approach. The datasets contains 8213, 6000 records and 23, 21 variables respectively.

The proposed DeepDropNet was evaluated on two real-world student dropout datasets using 10-fold cross-validation. Table 3 and 4 present the empirical results of the DeepDropNet model for both datasets. Across both datasets, the proposed architecture showed strong performance in terms of accuracy and AUC, even without applying any data balancing techniques.

Initially, DeepDropNet achieved 78.44% and 91.75% accuracy for the Moroccan public university Dataset and the Moroccan private university Dataset, respectively, when label encoding was used to preprocess categorical attributes. One important observation from the experiments is that the accuracy for the Moroccan public university Dataset increased slightly (78.55%) when one-hot encoding was applied instead of label encoding. However, for the Moroccan private university dataset, label encoding resulted in better performance. Once data balancing techniques were introduced, performance metrics improved significantly across both datasets, not only in terms of accuracy but also across other evaluation metrics.

For the Moroccan public university dataset, the use of SMOTEEN led to an accuracy improvement of up to 92.40%. Moreover, SMOTEEN consistently outperformed SMOTE and SMOTETomek across all evaluation metrics. Similarly, for the Moroccan private university dataset, SMOTEEN yielded the highest accuracy (94.03%) compared to the other oversampling methods.

After hyperparameter tuning, the performance of DeepDropNet improved sharply. When the kernel size was increased from 3 to 5, accuracy rose by approximately 1% for both datasets. Likewise, tuning the number of filters produced further performance gains. The best results; 92.68% and 96.41% accuracy for the Moroccan public university and Moroccan private university dataset, respectively; were achieved using 128 filters instead of 64 or 32. As shown in Tables 3 and 4, using a Flatten layer in the final architecture produced better results than using a Global Average Pooling layer. Additionally, replacing the SE block with the Basic Channel Attention module [21] yielded good results, though the SE block ultimately offered the highest overall performance on both datasets.

Regarding optimizers, DeepDropNet performed robustly with several configurations at a learning rate of 0.001. The ADAM optimizer achieved the best accuracy: 94.68% (Moroccan public university) and 96.41% (Moroccan private university). NADAM closely followed with 94.24% and 96.37% accuracy, respectively. RMSprop also provided competitive results, with metrics comparable to ADAM and NADAM across both datasets.

Interestingly, a consistent pattern was observed: as the learning rate decreased, the performance of DeepDropNet declined. When the learning rate was reduced to 0.001, accuracy dropped to 94.68% for the Moroccan public university and 96.41% for the Moroccan private university. Evaluation metrics (precision, recall, F1-score, and AUC) also showed a similar decreasing trend at lower learning rates.

We conducted an initial interpretability analysis using SHAP values and Grad-CAM. SHAP results showed that GPA, attendance rate, and programme type were the most influential variables for predicting dropout. For Grad-CAM applied to intermediate feature maps, they indicated behavioural signals such as attendance frequency were strongly highlighted by the attention modules. The initial findings suggest that DeepDropNet detects useful patterns, thereby enhancing its transparency and accuracy.

Table 3: Our Proposed Model Performance on Moroccan Public University Dataset

Encoding	*Size of filter*	*Data balancing technique*	*Number of filters*	*Flatten Layer/GAP*	*Channel attention*	*Optimzer*	*Learning rate*	*Precision*	*F-measure*	*Recall*	*Accuracy*	*AUC*	*MCC*
Label Encoding	3	Not used	64	GAP	SE Block	ADAM	0.001	61.37	57.42	54.43	78.44	82.35	44.10
One-hot Encoding	3	Not used	64	GAP	SE Block	ADAM	0.001	65.03	54.00	47.20	78.55	82.87	42.46
One-hot Encoding	3	SMOTE	64	GAP	SE Block	ADAM	0.001	86.04	78.51	72.35	80.11	90.02	62.18
One-hot Encoding	3	SMOTETomek	64	GAP	SE Block	ADAM	0.001	86.51	81.62	77.37	82.72	91.35	66.62
One-hot Encoding	3	SMOTEEN	64	GAP	SE Block	ADAM	0.001	94.17	92.40	90.55	91.76	97.22	84.66
One-hot Encoding	5	SMOTEEN	64	GAP	SE Block	ADAM	0.001	93.93	92.65	91.58	92.31	97.23	85.32
One-hot Encoding	5	SMOTEEN	32	GAP	SE Block	ADAM	0.001	91.50	92.22	92.22	91.50	97.03	84.18
One-hot Encoding	5	SMOTEEN	128	GAP	SE Block	ADAM	0.001	94.38	93.22	93.22	92.68	97.45	86.48
One-hot Encoding	**5**	**SMOTEEN**	**128**	**Flatten Layer**	**SE Block**	**ADAM**	**0.001**	**94.77**	**94.02**	**95.03**	**94.68**	**97.67**	**90.07**
One-hot Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	ADAM	0.001	94.86	94.70	94.70	94.43	97.62	89.50
One-hot Encoding	5	SMOTEEN	128	Flatten Layer	Basic Channel Attention	ADAM	0.001	94.50	94.32	94.32	93.87	97.50	88.75
One-hot Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	NADAM	0.001	94.57	94.68	94.68	94.24	97.53	89.52
One-hot Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	RMSprop	0.001	94.40	94.43	94.43	94.04	97.35	88.76
One-hot Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	ADAM	0.001	94.14	94.61	94.70	94.31	97.35	89.22

Table 4: Our Proposed Model Performance on Moroccan Private University Dataset

Encoding	*Size of filter*	*Data balancing technique*	*Number of filters*	*Flatten Layer/GAP*	*Channel attention*	*Optimizer*	*Learning rate*	*Precision*	*F-measure*	*Recall*	*Accuracy*	*AUC*	*MCC*
Label Encoding	3	Not used	64	GAP	SE Block	ADAM	0.001	81.70	70.01	61.83	91.75	88.80	67.17
One-hot Encoding	3	Not used	64	GAP	SE Block	ADAM	0.001	79.21	65.53	57.00	90.87	87.44	62.64
Label Encoding	3	SMOTE	64	GAP	SE Block	ADAM	0.001	94.67	90.61	87.04	91.07	96.27	83.36
Label Encoding	3	SMOTETomek	64	GAP	SE Block	ADAM	0.001	93.60	90.16	87.15	90.48	96.10	82.33
Label Encoding	3	SMOTEEN	64	GAP	SE Block	ADAM	0.001	95.47	94.55	93.67	94.03	97.72	89.10
Label Encoding	5	SMOTEEN	64	GAP	SE Block	ADAM	0.001	95.36	96.13	96.71	95.61	98.26	92.21
Label Encoding	5	SMOTEEN	32	GAP	SE Block	ADAM	0.001	95.20	95.10	93.84	94.47	98.02	90.03
Label Encoding	5	SMOTEEN	128	GAP	SE Block	ADAM	0.001	96.41	96.26	96.12	96.00	98.27	92.81
Label Encoding	**5**	**SMOTEEN**	**128**	**Flatten Layer**	**SE Block**	**ADAM**	**0.001**	**96.81**	**96.70**	**96.62**	**96.41**	**98.46**	**93.87**
Label Encoding	5	SMOTEEN	128	Flatten Layer	Basic Channel Attention	ADAM	0.001	96.14	96.64	97.25	96.31	98.37	93.63
Label Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	NADAM	0.001	96.12	96.60	97.06	96.37	98.34	93.56
Label Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	RMSprop	0.001	96.36	96.56	96.76	96.33	98.25	93.48
Label Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	ADAM	0.001	95.82	96.52	97.25	96.30	98.38	93.46
Label Encoding	5	SMOTEEN	128	Flatten Layer	SE Block	ADAM	0.001	96.05	96.62	97.62	96.40	98.48	93.61

Table 5: Comparison of the Proposed Model with State-of-the-Art Methods

Approach	***Methods***	***Accuracy***
Janka et al.[24]	NB, DT, SVM, LR	77%, 90%, 91%, 92%
Noviandy et al.[25]	Stacked LightGBM + RF + LR	80.2%
Alice et al. [26]	Gradient Boosting	83%
Mohamed Sayed [27]	CNN	93%
Wu et al [28]	CNN-BiGRU	90%
Rabelo et al. [29]	LR + NN + DT	89%
Bouihi et al.[30]	Lasso regression	89%
Elbouknify et al.[31]	ML ensemble	88%
Silva et al.[32]	ANN	84%
Vaarma et al. [33]	XGBoost	90.3%

Table 5 compares the performance of DeepDropNet with state-of-the-art models. It is evident that the proposed model outperformed classical ML methods such as SVM [24], Logistic Regression [30], and ensemble classifiers. DeepDropNet also outperformed deep learning baselines such as ANN [32], CNN-BiGRU [28]. Moreover, it surpassed ensemble learning strategies like Gradient Boosting in terms of accuracy, AUC, and generalization ability [26].

6 Conclusion

Student dropout prediction is a significant issue in education, as it directly affects institutional performance and student achievement. To address this challenge, this research introduces a new deep learning model, DeepDropNet, which integrates 1D convolutional layers with residual learning, squeeze-and-excitation blocks, and spatial attention to enhance predictive power. We have tested the introduced model using two real academic datasets collected from two Moroccan universities characterized by high class imbalance. To mitigate this challenge, we employed three oversampling strategies: SMOTE, SMOTETomek, and SMOTEENN. Performance evaluations using 10-fold cross-validation and various metrics (accuracy, precision, recall, F1-score, MCC, and AUC) confirmed that DeepDropNet outperformed existing methods across all benchmark datasets. While this success was achieved primarily through data-centric approaches, future research could explore federated learning for privacy-compliant dropout prediction across multiple institutions. Additionally, applying various feature engineering and data transformation methods may further enhance the model's accuracy.

The proposed model (DeepDropNet) can be easily integrated into existing Learning Management Systems (LMS), such as Moodle, to support early intervention activities. In deployment contexts, the model can run behind an LMS, continuously monitoring student profiles and activity logs to automatically flag at-risk students. These warnings enable instructors and academic advisors to initiate early interventions, such as personalized feedback, additional tutoring, or counseling support. This integration not only enhances institutional decision-making but also allows teachers to provide targeted interventions, ultimately improving student retention in real school settings.

Finally, as DeepDropNet is a black-box model, incorporation of explainable AI (XAI) techniques would be an interesting direction towards enhancing transparency and trust in the prediction result.

References

[1] Salem, M., & Shaalan, K. (2025). Unlocking the power of machine learning in e-learning: A comprehensive review of predictive models for student performance and engagement. *Education and Information Technologies*, 1–24.

[2] Hamal, O., El Faddouli, N. E., Harouni, M. H. A., & Lu, J. (2022). Artificial intelligent in education. *Sustainability, 14*(5), 2862.

[3] El Koufi, N., & Belangour, A. (2024). Toward a recommender system for assisting customers at risk of churning in e-commerce platforms based on a combination of social network analysis (SNA) and deep learning. Journal of Open Innovation: Technology, Market, and Complexity, 10(4), 100425.

[6] Mustofa, S., Emon, Y. R., Mamun, S. B., Akhy, S. A., & Ahad, M. T. (2025). A novel AI-driven model for student dropout risk analysis with explainable AI insights. *Computers*

and Education: Artificial Intelligence, 8, 100352. https://doi.org/10.1016/j.caeai.2025.100352

[7] Pecuchova, J., & Drlik, M. (2023). Predicting students at risk of early dropping out from course using ensemble classification methods. *Procedia Computer Science, 225*, 3223–3232. https://doi.org/10.1016/j.procs.2023.11.407

[8] Lottering, R., Hans, R., & Lall, M. (2020). A machine learning approach to identifying students at risk of dropout: A case study. *International Journal of Advanced Computer Science and Applications, 11*(10), 417–422. https://doi.org/10.14569/IJACSA.2020.0111055

[9] Niu, K., Cai, J., Zhou, Y., Tai, W., & Feng, X. (2025). Hybrid neural network model for MOOC dropout prediction. *Complex Systems Modeling and Simulation*. https://doi.org/10.32604/csms.2025.050419

[10] 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

[11] Choque-Soto, V. M., Sosa-Jauregui, V. D., & Ibarra, W. (2025). Characterization of the dropout student profile using data mining techniques. *Revista de Gestão Social e Ambiental, 19*(2), 1–20. https://doi.org/10.24857/rgsa.v19n2-011

[12] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education, 53*(3), 950–965. https://doi.org/10.1016/j.compedu.2009.05.010

[13] Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences, 11*(7), 3130. https://doi.org/10.3390/app11073130

[14] Jimenez Martinez, A. L., et al. (2024). Early detection of at-risk students using machine learning. *arXiv preprint arXiv:2412.10375*. https://doi.org/10.48550/arXiv.2412.10375

[15] Marcolino, M. R., et al. (2025). Student dropout prediction through machine learning optimization: Insights from Moodle log data. *Scientific Reports, 15*, 9840. https://doi.org/10.1038/s41598-025-64050-8

[16] Aljohani, N. R. (2019). A LSTM deep learning model to predict student dropout in online courses. *International Journal of Advanced Computer Science and Applications, 10*(11), 417–423.

[17] Marcolino, M. R., et al. (2025). Student dropout prediction through machine learning optimization. Scientific Reports, 15, 9840. https://doi.org/10.1038/s41598-025-64050-8

[18] (2023). Student dropout prediction for university with high precision and recall. Applied Sciences, 13, 6275. https://doi.org/10.3390/app13116275

[19] Zapata-Medina, D., Espinosa-Bedoya, A., & Jiménez-Builes, J. A. (2024). Improving the automatic detection of dropout risk in middle and high school students: A comparative study of feature selection techniques. Mathematics, 12(12), 1776. https://doi.org/10.3390/math12121776

[20] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7132–7141). IEEE.

[21] El Koufi, N., & Belangour, A. (2024). Toward a decision-making system based on artificial intelligence for precision marketing: A case study of Morocco. *Journal of Open Innovation: Technology, Market, and Complexity, 10*(1), 100250.

[22] El Koufi, N., Belangour, A., & Sdiq, M. (2022). Research on precision marketing based on big data analysis and machine learning: Case study of Morocco. International Journal of Advanced Computer Science and Applications, 13(10).

[23] El Koufi, N., & Almazroei, E. E. (2024). Sentiment analysis and deep learning-based approach for e-commerce accurate marketing: Case study of Morocco. In International Conference on Advances in Communication Technology and Computer Engineering (pp. 347–352). Springer Nature Switzerland.

[24] Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences, 11*(7), 3130. https://doi.org/10.3390/app11073130

[25] Noviandy, T. R., et al. (2024). Machine learning for early detection of dropout risks and academic excellence: A stacked classifier approach. *Journal of Education, Management and Learning, 2*(1), 28–34. https://doi.org/10.58188/jeml.v2i1.47

[26] Alice, V., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence, 4*(1), 2. https://doi.org/10.1007/s44163-023-00066-0

[27] Sayed, M. (2024). Student progression and dropout rates using convolutional neural network: A case study of the Arab Open University. *Journal of Advanced Computational Intelligence and Intelligent Informatics, 28*(3), 668–678. https://doi.org/10.20965/jaciii.2024.p0668

[28] Wu, X., Yu, Z., Zhang, C., & Yang, Z. (2025). Research on MOOC dropout prediction by combining CNN-BiGRU and GCN. In *Proceedings of the International Conference on Computer Vision, Applications, and Algorithms (CVAA 2024)*, Lecture Notes in Computer Science, 13486 (pp. 683–690). Springer. https://doi.org/10.1007/978-981-97-0042-4_56

[29] Rabelo, A. M., & Zárate, L. E. (2025). A model for predicting dropout of higher education students. *Data Science and Management, 8*(1), 72–85. https://doi.org/10.1016/j.dsm.2025.01.007

[30] Bouihi, B., Bousselham, A., Aoula, E., Ennibras, F., & Deraoui, A. (2024). Prediction of higher education student dropout based on regularized regression models. *Engineering, Technology & Applied Science Research, 14*(6), 17811–17815. https://doi.org/10.48084/etasr.7081

[31] Elbouknify, I., et al. (2025). AI-based identification and support of at-risk students: A case study of the Moroccan education system. *arXiv preprint arXiv:2504.09312*. https://doi.org/10.48550/arXiv.2504.09312

[32] Moreira da Silva, D. E., Solteiro Pires, E. J., Reis, A., de Moura Oliveira, P. B., & Barroso, J. (2022). Forecasting students dropout: A UTAD university study. *Future Internet, 14*(3), 76. https://doi.org/10.3390/fi14030076

[33] Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society, 76*, 102346.

Notes on contributors

Oussama Hamal is an Assistant Professor at the National School of Architecture (ENA), Morocco. He holds a Ph.D. in Artificial Intelligence from the Mohammedia School of Engineers (EMI), Rabat. Dr. Hamal also holds a Master's in Information Systems Architecture from the University of Toulouse, France, and a Master's in Mathematics and Computer Science from Abdelmalek Essaâdi University, Tetouan.

His research focuses on applying AI to education, e-learning, and smart learning environments, with an emphasis on smart campuses, IoT, and pedagogical innovation. He has published widely in international journals and conferences on machine learning, deep learning, data mining, and big data, with applications extending to smart cities, BIM, e-health, and e-agriculture.

Nouhaila El Koufi is an Assistant Professor in Computer Science and Artificial Intelligence at the Faculty of Economics and Management (FEG), Morocco. She obtained her Ph.D. in Computer Science and Artificial Intelligence in October 2024. She received her Engineering Degree in Computer Science from the National School for Computer Science and Systems Analysis (ENSIAS), Rabat, and her Bachelor's Degree in Mathematics and Computer Science from Hassan II University.

Her primary research interests include Generative AI, Artificial Intelligence, Big Data, Machine Learning, Deep Learning, Recommender Systems, the Internet of Things (IoT), and Multi-Agent Systems, with applications in diverse fields such as healthcare, epidemiology, marketing, e-learning, agriculture, and economics.

Korhan Cengiz is an Associate Research Professor at the Department of Electrical Engineering, Prince Mohammad Bin Fahd University, Saudi Arabia, and an Associate Professor at Biruni University, Turkey. He also serves as an Adjunct Professor at the University of Jordan. He holds a Ph.D. in Electronics Engineering from Kadir Has University, Turkey, an M.Sc. in Electronics and Communication Engineering from Namık Kemal University, and dual B.Sc. degrees in Electronics and

Communication Engineering (Kocaeli University) and Business Administration (Anadolu University).
His research interests include wireless sensor networks, wireless communications, statistical signal processing, indoor positioning systems, the Internet of Things, power electronics, and 5G technologies.

Luka Milić received the Ph.D. degree at the Faculty of Electrical Engineering and Computing at the University of Zagreb in 2022, with a thesis in the area of IoT. Since 2013 he has been working at the Faculty of Organization and Informatics at the same University, where he is currently a Senior Teaching Assistant. He is a member of the IoT Laboratory and the Laboratory for Design of Program Interfaces, Internet Services and Videogames. His research interests include IoT, operating systems, computer networks, and gamification. He was a commitee member of several conferences, as well as a reviewer for multiple journals and conferences, and an associate of multiple expert and research projects.