Ensemble Deep Learning Network Model for Dropout Prediction in MOOCs

Gaurav Kumar

Lovely Professional University,
Jalandhar-Delhi, G.T. Road, Phagwara,
Punjab, India -144411.
gaurav.20323@lpu.co.in

Amar Singh

Lovely Professional University,
Jalandhar-Delhi, G.T. Road, Phagwara,
Punjab, India -144411.
amar.23318@lpu.co.in

Ashok Sharma

University of Jammu, Bhaderwah Campus,
Jammu and Kashmir -180006.
ashoksharma@jammuuniversity.ac.in

Abstract – *In the online education field, Massive open online courses (MOOCs) have become popular in recent years. Educational institutions and Universities provide a variety of specialized online courses that helps the students to adapt with various needs and learning preferences. Because of this, institutional repositories creates and preserve a lot of data about students' demographics, behavioral trends, and academic achievement every day. Moreover, a significant problem impeding their future advancement is the high dropout rate. For solving this problem, the dropout rate is predicted by proposing an Ensemble Deep Learning Network (EDLN) model depending on the behavior data characteristics of learners. The local features are extracted by using ResNet-50 and then a kernel strategy is used for building feature relations. After feature extraction, the high-dimensional vector features are sent to a Faster RCNN for obtaining the vector representation that incorporates time series data. Then an attention weight is obtained for each dimension by applying a static attention mechanism to the vector. Extensive experiments on a public data set have shown that the proposed model can achieve comparable results with other dropout prediction methods in terms of precision, recall, F1 score, and accuracy.*

Keywords: Deep learning; MOOC; feature extraction; dropout prediction; activity patterns.

1. INTRODUCTION

With the assistance of big data technology and artificial intelligence, an innovative and rapidly growing educational strategy is MOOCs [1]. Through online courses, MOOCs connect participants in global education and give students, instructors, and academic institutions access to an interactive Internet platform [2]. MOOCs now have a significantly larger student population, particularly in the current pandemic with their affordability and convenient features [3]. The high dropout rate currently in place, however, is severely impeding the growth of MOOCs. According to numerous research, less than 10% of MOOC courses are completed [4]. Only 7% of students finish the University of California's courses offered on the Coursera platform, according to statistical data [5]. Significant possibili-

ties for early reversal of the alarming student dropout and higher retention rates are predicted by the MOOC dropout prediction models [6]. These predictions are used to keep students motivated to learn and stop students from dropping out of course instructors through interventions [7].

Depending on the current learning behavior of the students, the chances of course dropout are examined by the MOOC dropout prediction [8]. For MOOC dropout prediction, traditional deep learning and machine learning methods are currently used [9] [10]. Most machine learning-based classification techniques are used in traditional machine learning research [11]. A large amount of time and effort must be expended manually for extracting features [12] [13]. Moreover, the lack of large-scale datasets for training these tech-

niques restricts their application in the MOOC present context [14]]. Higher predictive results are produced by deep learning models than the traditional machine learning models [15] [16]. The feature information is automatically extracted from input data by using the convolutional neural network (CNN), which is the most popular current dropout prediction model. However, it is unable to utilize the data from the time series [17]. The dropout prediction is effectively improved by using Faster RCNN models in certain researchers, and the time information is also captured by this network [18]. Furthermore, several recent research discovered that various characteristics should be handled differently because they have various consequences on the decision to drop out. So, to accomplish this concept, attention becomes a useful focus [19] [20].

The innovative MOOC dropout prediction model is proposed in this research based on previous research. The proposed model is called Ensemble deep learning network (EDLN) model. Faster RCNN and attention mechanisms are integrated with this proposed model. Automatic local feature extraction from the source data is done by the proposed model. Then these features are combined with time series information and predicted by multiplying the combined features by feature-wise weights. At the end of the course, in contrast to existing models, the proposed model's advantage is that it also predicts students' status. The learners' status is predicted by additionally fully exploiting the learner's key feature information and the learner's time series information during every week of the learning process. The essential information is provided by the proposed network model for instructors at risk of dropping out to select when and how to deliver personalized instruction to students.

The main contributions of the research are

- The input for the MOOC dropout model is a time series matrix in two dimensions. The original data's time series state is efficiently preserved while the weekly learner's input features are recorded in this matrix. During the course learning process, the learners' weekly status is predicted by this approach and makes timely interventions and it provides instructors intervention in time.

- The temporal relation between student behavior characteristics weekly is examined using the Faster RCNN. To weigh the characteristics, a static attention method is used by the significance of the behavioral characteristics. In dropout prediction, the effective features are extracted by using the ResNet-50. The efficiency of dropout prediction is effectively improved by the proposed model.

- Comparison experiments established the EDLN model's validity. While compared to the existing models, the proposed EDLN model predicts dropout effectively in the KDD CUP 2015 dataset.

2. LITERATURE REVIEW

This section, review some existing DL techniques for dropout prediction of MOOC learners.

A new feature extraction method is proposed by Jin et al [21] for behavior data of students for learning in this paper. The weekly characteristics of student learning behaviors are used for the experiment analysis. Then, the student dropout is predicted by developing the new support vector regression (SVR). An improved quantum particle swarm optimization (IQPSO) algorithm is used for optimizing the parameters in this paper.

A different integrated structure for MOOCs dropout prediction is proposed by Qiu et al [22]. A feature selection (FSPred) is proposed in this paper. Feature generation, feature selection, and dropout prediction are included in the proposed model. The features are generated by applying a fine-grained feature-generation method and then the valid features are selected by using the hybrid feature selection method. After the generation and selection of features, the logistic regression model is used for the dropout prediction.

A novel supervised ML algorithm is proposed by Panagiotakopoulos et al [23] to predict the dropout of students in MOOC. Six well-known metrics were used to evaluate several predictive models. The learning algorithm's performance is improved by using random search to automatically optimize the hyperparameter. The classification performance is further improved by applying stacked generalization approach was applied to further improve the classification performance.

The novel dropout prediction model is proposed by Xing et al [24]. The intervention personalization was examined for improving the effectiveness of the model in MOOCs. The dropout prediction model is constructed by developing the deep learning model in this research. After that individual student dropout probability is predicated on a temporal prediction mechanism. For at-risk students in MOOCs, individual dropout rates to personalize and prioritize intervention are examined.

In online short courses, a new methodology is examined by Chen et al [25] for dropout prediction of students. The creation of predictive learning analytics is complicated due to the limited enrollment in this course and the absence of intermediate assessments. Only behavior-based machine learning features that have been processed from measurements gathered throughout the learning process are used in this method.

A novel feature extraction method is done by Wan et al [26] for predicting the effectiveness of the students. Then, a model for transfer learning based on TrAdaBoost was proposed. It was applied to the current course iteration's pre-trained model using the data from the previous iteration of the course. In addition, this research contrasted how latecomers changed their learning behavior between the controlled group and the experimental group.

Deep learning is used for increasing the model's performance from the investigation of the above studies across many fields. A MOOC dropout prediction model based on this study's concept to combine the static attention mechanism with Faster RCNN is presented. By assigning the extracted features weights based on static attention, the model's accuracy is increased while identifying important features.

3. PROPOSED METHODOLOGY

3.1. PROBLEM STATEMENT

For five weeks, the analysis was done on the students' records for this research. Whether the learners dropped out is accessed by using the five-week activity records. If there were ten consecutive days without any learning activities, students were classified as dropouts; otherwise, they were classified as non-dropouts. Problems with categorical prediction were established from the dropout problems in this research, with those who had not dropped out represented by 1 and who had dropped out represented by 0.

3.2. PROPOSED EDLN MODEL

The Faster RCNN and static attention mechanisms are combined in the MOOC dropout prediction model based on EDLN. First, the original data is sent to the model as a two-dimensional temporal matrix. For the behaviors of the learners, the two-dimensional convolu-

tion kernel of ResNet-50 is used to extract the local high-dimensional feature information automatically. Then the computational load of the model and dimension of invalid features is reduced by adding a max pooling layer.

Using the local feature data, the time series' hidden long memory features are then retrieved by Faster RCNN and it uses a time series encoding algorithm to encode the data. The feature information is assigned by weight using a static attention mechanism. The key feature information is also highlighted by the static attention method and also the model's effectiveness is further enhanced by this mechanism. Finally, a sigmoid function representing the results of the MOOC dropout classification prediction is output. Fig. 1 shows the structure of the EDLN model. 1 fully connected layer, 7 pooling layers, 7 convolutional layers, 1 RCNN layer, and 1 Static Attention layer are presented in the proposed detection network model.

An EDLN-based model for predicting MOOC dropouts is proposed in this research. In Fig. 2, the model-based prediction process is described. Preprocessing of data, prediction, and evaluation of the model are the main three parts of the proposed model. The KDD 2015 dataset's clickstream data is first processed, and the weekly data on the behavioral characteristics of every student is used as the original data. The time series information and local feature learning of the source data is then automatically extracted and learned using the EDLN MOOC dropout prediction model. Finally, the model's performance was assessed using precision, recall, F1-score, and accuracy.
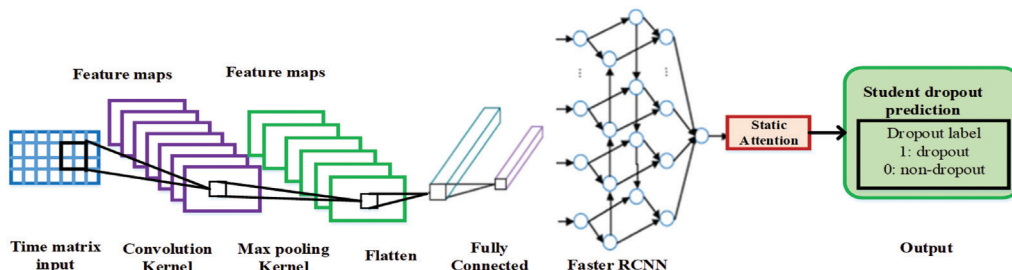

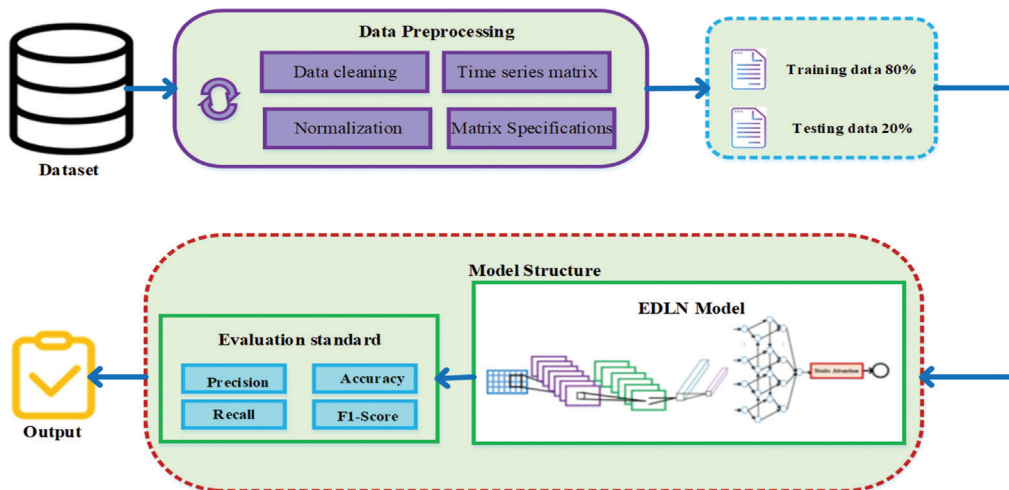**Fig.1.** EDLN model structure.

Fig. 2. Schematic diagram of the proposed methodology

3.2.1. Preprocessing

The clickstream data in this data set provides a behavior log that specifically records the course ID, student ID, occurrence time, and click event. In this research, the student data is first cleaned, and removed dropout labels from the dataset. From 12,000 students, the original data is selected randomly using student ID numbers during course learning which consisted of 7 different feature data types totaling 60,000 pieces during course learning. The multivariate time series data is presented in the characteristic behavior of

learner's datasets with many behavioral features like wiki, discussion, video, page close, traverse, access, and problem. Therefore, a two-dimensional temporal matrix is used to preprocess the data to properly utilize the time series and the information about hidden features between different behavioral variables. Data on a student's d behavioral characteristics over s weeks are contained in the time matrix, starting from week t. Each model input's time matrix is described in Eq. (1).

$$X_t = \begin{bmatrix} x_t^1 & x_t^2 & \dots & x_t^d \\ x_{t+1}^1 & x_{t+1}^2 & \dots & x_{t+1}^d \\ \dots & \dots & \dots & \dots \\ x_{t+s-1}^1 & x_{t+s-1}^2 & & x_{t+s-1}^d \end{bmatrix} \quad (1)$$

The frequency of the d distinctive behaviors is shown for each row in the matrix for the corresponding week, according to the equation above.

For MOOCs with various temporal dimensions, this research generated five separate time series matrices with varying specifications as input to the EDLN model by successively segmenting the datasets and using the data from various weeks' models to examine the dropout prediction performance of the EDLN model. The week dataset utilized and each specification time series matrix are shown in Table 1.

Table 1. Specifications for time series matrices.

Datasets	Specifications for time series matrices
Week 1 data	1 x 7
Week 1–2 data	2 x 7
Week 1–3 data	3 x 7
Week 1–4 data	4 x 7
Week 1–5 data	5 x 7

The data on the characteristics of the student's behavior is divided into weeks in this experiment. $n \times$ 7-dimensional input matrices are generated from the combined weekly data, where $n = 1, 2, 3, 4, 5$. The EDLN MOOC dropout prediction model employed the normalized time series matrix as its input.

3.2.2. Feature extraction using ResNet-50 network

The process of feature extraction is carried out manually and it needs researchers with specialized knowledge. The labor-intensive and time-consuming process of manual feature extraction occurs due to the very low frequency of behavioral characteristics of students and very complex potential dropout patterns in the course. The ResNet-50 algorithm is used for solving the limitations of manual extraction. To extract local feature information, in which the number of input features is equal to the number of convolution kernels.

Full connection, pooling, and convolution are the three levels of a ResNet-50 network's neural architecture. In the convolutional layer, the time matrix's input features are used to calculate the k convolutional kernels. The convolution kernels are enabled for extracting each input feature's characteristic information for each dimension, and the convolution kernel's size is set at (u, v), where Eq. (2) shows the convolution kernel calculation formula.

$$CF = q(w^T X^{(i+u-1, j+v-1)} + b) \quad (2)$$

Where the activation function is represented by , how much behavioral feature data for learners overlap with the convolution kernel is denoted and the bias term is represented by b.

Convolutional and fully connected layers comprise the proposed ResNet-50 feature extraction network. There are trainable weights set for each layer. The time series matrix of size is the model's input. Using a kernel method, the combination of both behavioral dimension features (7) and temporal dimension features (n weeks) are extracted by the proposed EDLN model.

The convolutional layer is the first layer of the feature extraction network. After the last convolutional layer, a flattening operation is performed. Pooling, activation, and convolution are present in the convolutional layer. From the original data, only seven behavioral features were taken into consideration during the convolution process. The convolutional operations are performed with a 3×3 kernel. While focusing on the neighbors, using a smaller kernel navigates the input fields more than large kernels. It allows the model to fully utilize the limited information. Testing each layer with a range of 5 to 10 kernels, the different information is captured by employing the 7 distinct kernels finally for providing the highest performance.

After each convolution, a function of Rectified linear unit (Relu) is selected. We utilize a max-pooling technique in the pooling process. Then the feature maps in the original data are less distinct and close to zero. Because data on the behavior of the dropout students are typically 0 (meaning none). This observation leads to the process of max-pooling, which retrieves the greatest value. This highest value is better suitable for the proposed analysis. After that, the model performs the flattening operation of the generated feature maps.

The fully connected layer is the second element of the CNN module. The flattened convolutional outputs are represented more densely by the fully connected layer. For the next Faster RCNN module, these representations are used as the inputs.

However, learners' learning behavior features are automatically extracted by CNN. These features are in the form of time series data, and the significant time series correlation information range is present in the data. As a result, time series data cannot be extracted only with ResNet-50. Moreover, the time series relationship between the features of learners' behavior learners' time series behavior feature's relationship is extracted by using the Faster RCNN model in this research.

3.2.3. Dropout prediction using Faster RCNN

Faster RCNN is a development of CNN-based RCNN and fast RCNN networks. Several object detection processes are performed by this network. How regions are chosen for processing is the significant difference between them. A region selection algorithm is used by both RCNN and fast RCNN for object detection like Selective Search or Edge Boxes that are different from the CNN network. When training and detecting CNN, the region selection is performed by faster RCNN. The dataset related to this application is used to train the last fully connected layer.

The gradient disappearance problem is improved by developing the Faster RCNN model in recurrent neural networks (RNNs) caused by long input sequences. The input gates, output gates, forget gates, input layer, and output layer are all components of a faster RCNN. The "gate" control mechanism is used by the Faster RCNN for adding or discarding part of the information. Then the memory cell state is updated by combining the current input, historical memory, and historical state.

The neural unit's input information is currently controlled by the input gate. The neural unit's output information is currently controlled by the output gate. The historical data previously stored by the neural unit is controlled by the forget gate.

The information is selectively filtered by the "gate" structure and it consists of the dot product operation's sigmoid function. The sigmoid function produces an output in the range [0, 1], where complete passing is represented by 1 and complete rounding is represented by 0. The below equation is used to determine the dropout prediction,

$$IG_t = \sigma\left(WT_i \cdot [h_{t-1}, x_t] + BS_i\right) \qquad (3)$$

$$FG_t = \sigma\left(WT_f \cdot [h_{t-1}, x_t] + BS_f\right) \qquad (4)$$

$$OG_t = \sigma\left(WT_o \cdot [h_{t-1}, x_t] + BS_o\right) \qquad (5)$$

$$G_t = \sigma\left(WT_c \cdot [h_{t-1}, x_t] + BS_c\right) \qquad (6)$$

$$CS_t = FG_c \cdot CS_{t-1} + IG_t \cdot G_t \qquad (7)$$

Where the sigmoid function is represented by $\sigma(\bullet)$, the weights and biases of the input gate IG_t, forget gate FG_t, and output gate OG_t is represented by WT_i, WT_f, WT_o, BS_i, BS_f, and BS_o from which output h_t at current moment t and cell state C_t at current moment t is calculated.

3.2.4. Attention Mechanism

In several deep-learning fields, the attention mechanism has been extensively employed in recent years. Assigning larger weights to information by using the attention mechanism, is more important for the proposed model. The attention method used in this research is implemented using static attention. Faster RCNN uses the simple, efficient, and typically designed static attention method. While compared to soft attention, the model efficiency is improved by achieving a data vector representation with only one calculation.

Evaluating important features and ignoring unimportant features are done by using the static attention mechanism. Information on each feature's weight is determined by the static attention mechanism for the time series-based generation of local feature information. For adaptive learning, the weights are multiplied by the input feature data. The first hidden state of the RCNN's first layer typically uses the effective lightweight attention module known as static attention. Combining feature information with its output provides a weight value calculation. The following formula represents the main calculation.

$$O_d = \vec{O}_d(t) \,\|\, \bar{O}_d(t) \qquad (8)$$

$$q(t) = \tanh\left(w_{om}O_d(t) + w_{Om^o}\right) \qquad (9)$$

$$p(t) = \exp(w_{qp}{}^T q(t)) \qquad (10)$$

$$Z = O_d p \qquad (11)$$

At t the moment, the output of the Faster RCNN for each feature information is $O_d(t)$ for the input feature information d, and after static attention processing, the feature information's weighted vector is represented by q, where the weights represent the static attention network's level of attention provided to feature information.

4. RESULTS AND DISCUSSIONS

This section presents the environment settings, experimental datasets, and relevant software and hardware. The criteria for both the evaluations and performance analysis are also both clearly described.

4.1. DATASET DESCRIPTION

In this research, the EDLN dropout prediction model's effectiveness is assessed using data from the Cup 2015 KDD, which evolved from "XuetangX," China's largest MOOC platform. Over five months in 2013–2014, the 120,542 clickstream data points are recorded by the dataset in 39 courses from 79,186 students, with each

course lasting five weeks. Seven behavioral factors were identified in this dataset that describes students' behavior such as page close, navigate, discussion, wiki, access, video, and problem. A label was given to each chosen student indicating whether or not they had dropped out.

4.2. EXPERIMENTAL PARAMETERS

For this experiment, the training and test sets were divided into the dataset at an 8:2 ratio. In the model parameters, the empirical values are chosen for the optimum hyperparameters. The hyperparameters are tuned by using the Adam optimizer. The proper model training is ensured by setting the model's dropout is 0.2 to address the overfitting issue. 200 batches are being processed, with a learning rate of 0.0025. The parameter training is done by using the logarithmic loss function and adaptive learning rate optimization is performed by using the Adam optimization function. The best results are achieved by setting the model's epoch to 20.

4.3. EVALUATION METRICS

The most used metric for evaluating the effectiveness of deep learning models is accuracy. The larger teaching accidents are obtained by misclassifying dropout samples as non-dropout samples could result in more significant adverse effects. Therefore misclassifying non-dropout samples as dropout samples is preferable. Precision, recall, F1-score, and accuracy are used for the evaluation of the proposed model for predicting the MOOC student samples. Table 5 displays the confusion matrix used to define the MOOC dropout prediction model.

Higher priority is given to the precision and recall metrics for predicting the dropout samples because of the MOOC dropout prediction problem's cost-sensitive nature. A model with higher precision will accurately predict more samples. A model with a higher recall misses fewer data when making predictions. The model's higher accuracy demonstrates that it avoids making inaccurate predictions. The symmetrical mean of precision and recall is the F1 score.

4.4. EXPERIMENTAL RESULTS

Table 2 shows the five-time matrices that are compared in terms of evaluation values. When the student behavioral features within the first five weeks are used as input, the proposed model performs at its best. 97.5% accuracy can be attained. When the EDLN model uses the input data as a time series matrix of 1 x 7, the poorest classification performance is achieved than other inputs, it has obtained 87.7% accuracy. This is because fewer course tasks are in for learners, causes producing relatively few behavioral features. The more the course continues, the more behaviors the students produce.

Five-time matrices are compared in terms of evaluation values, as shown in Fig. 3. The model performed better than the time series matrix of 1 x 7 by about 1.7%

when the time series matrix of 2×7 was used as input data. While comparing the inputs of the 5×7 time series matrix and the 4×7-time series matrix, there is just about a 2.9% difference between the two experiments. The five-week length of each course in the sample provides for this result. In the last week, several learners decided for learning offline. In the fifth week, it results in learners displaying behavioral characteristics significantly less commonly. This affects the model's assessment results. Fig. 4 displays the proposed model's loss and accuracy and Fig. 5 displays the proposed model's confusion matrix.
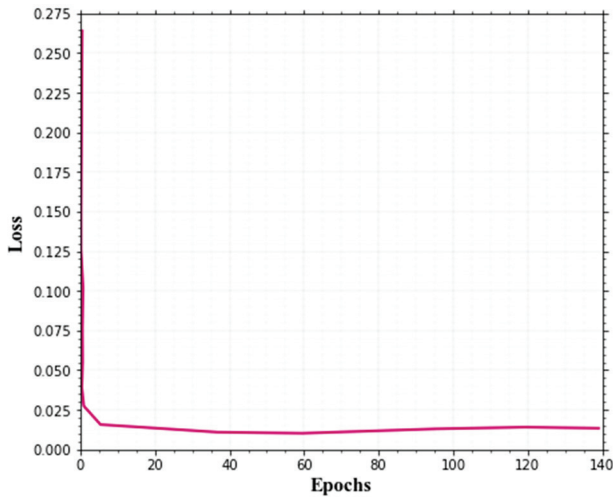
Table 2. The performance result comparison over five-time matrices

| Model	F1-Score (%)	Recall (%)	Precision (%)	Accuracy (%)
EDLN with 1 × 7 matrix	86.2	83.5	85.2	87.7
EDLN with 2× 7 matrix	88.6	85.2	86.6	89.4
EDLN with 3 × 7 matrix	92.3	90.9	90.2	92.9
EDLN with 4× 7 matrix	94.3	94.2	94.6	94.6
EDLN with 5 × 7 matrix	97.2	96.1	97.3	97.5
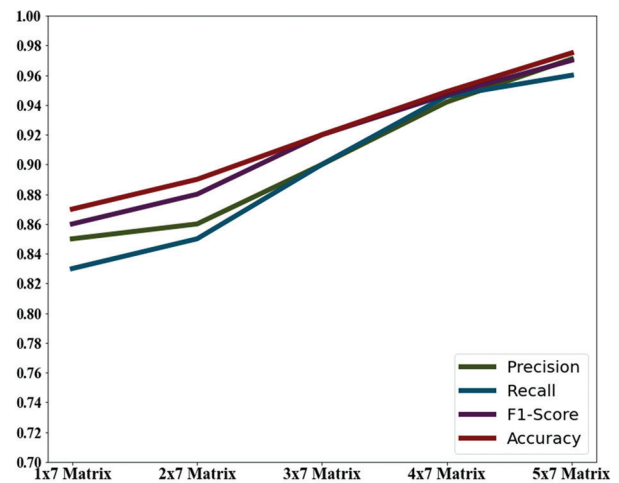
**Fig. 3.** Performance result comparison over five-time matrices

(a)
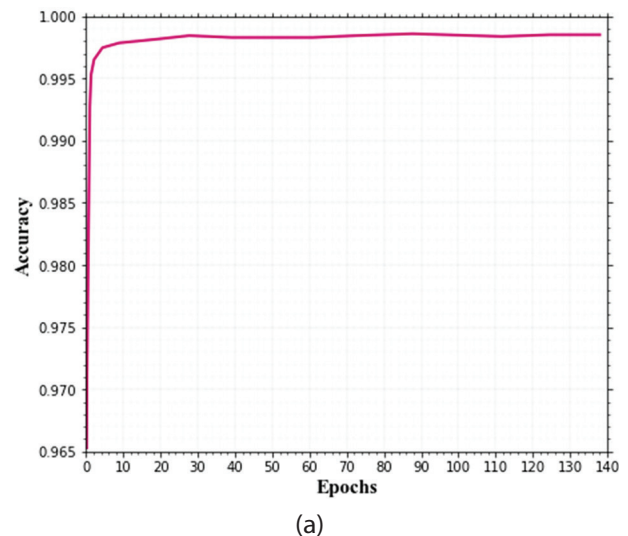
Fig. 4. Proposed model (a) accuracy (b) loss during the training epochs

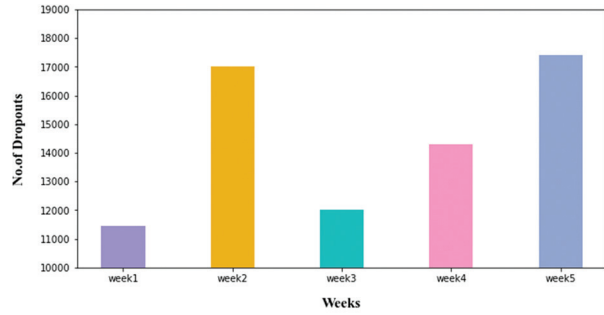Confusion Matrix: [[2131 1647]]
  [669 13652]
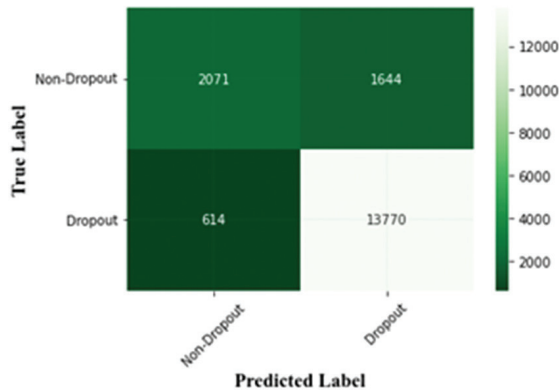
Fig. 5. Confusion matrix of the proposed model

The user's dataset was used to test the system for 5 weeks, and it successfully predicted how many students would drop out over that time. The model produces two classes of output: finished learners, and unfinished learners.
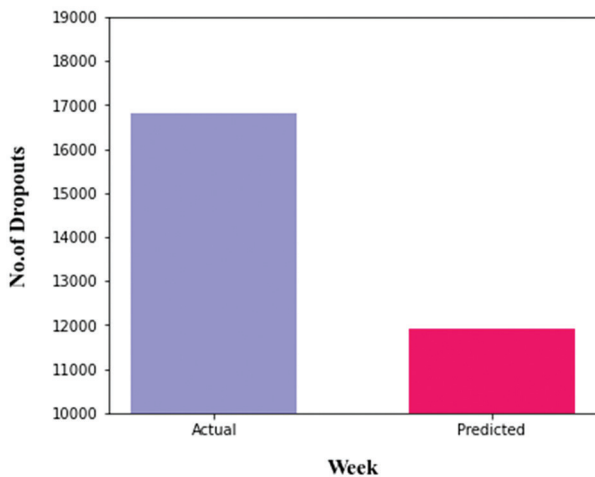
Fig. 6. Actual vs Predicted Dropout in a week

Fig. 7. The performance of Week-wise dropout

The performance of each student is also given a weekly rank. A steady gain will encourage the learner to finish the course early, whereas a consistent decrease will warn the learner about his or her likelihood of leaving the course in the approaching weeks. From the proposed model's performance, we have achieved a reduced dropout rate and this proposed method improves the engagement level of MOOC learners.

```
Week:  1 - Rank:   11306
Buck up buddy

Week:  2 - Rank:   918
That was a great improvement

Week:  3 - Rank:   6036
Watchout! You are slipping

Week:  4 - Rank:   11419
Watchout! You are slipping

Week:  5 - Rank:   11454
Watchout! You are slipping

At this rate you are more likely to dropout
```

Fig. 8. Student Intervention after every week

4.5. QUANTITATIVE EVALUATION

While comparing the baseline model with the EDLN model, the proposed model's advantages and efficiency are validated. Table 3 shows the performance effectiveness of the several existing models and the proposed model.

Table 3. Performance comparison of various models

Model	F1-Score (%)	Recall (%)	Precision (%)	Accuracy (%)
LSTM	78.7	78.8	78.7	79.2
CNN	81	81.2	80.9	81.3
MMSE	92.6	89.5	86.3	87.7
SVM-SGD	93.1	93.1	93.7	91
AdaBoost	94.2	93.8	94.9	92.9
The proposed model (EDLN)	97.2	96.5	97.1	97.4

The proposed technique achieved an accuracy, precision, recall, and F1 measure of 97.4%, 97.1%, 96.5%, and 97.2%, respectively, which indicates that the pro-

posed technique outperforms all other state-of-the-art methods. When compared to the SVM-SGD model, the proposed model accuracy is 6.4% higher. From this analysis, the large-scale MOOC dropout prediction is performed more effectively by the proposed model than the SVM model.

The MMSE, LSTM, and CNN deep learning models are also included in the baseline models. The accuracy of the proposed model is higher than LSTM and CNN by approximately 18.2% and 16.1%. The local receptive field feature is used for performing the feature learning for CNN models in such disordered data. When comparing the proposed EDLN model with the CNN model in terms of all performance analysis values, the proposed EDLN model performs better than the standard CNN model. The performance analysis graph is shown in Fig. 8. Using the time matrix to predict MOOC dropouts, excellent performance is achieved by the proposed model in this research.
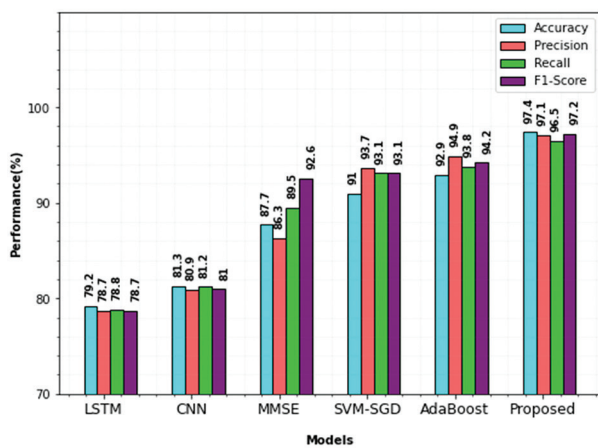

Fig. 9. Performance analysis

The source data for the five models were taken from the KDD CUP 2015 dataset. For dropout prediction, when comparing the existing deep learning model, the proposed EDLN model outperformed. It is shown that the proposed EDLN model, which incorporates the attention process and model input as a temporal matrix, is successful in increasing the dropout prediction accuracy in MOOCs. The proposed EDLN MOOC dropout prediction model more effectively extracts and learns the time series information and local feature learning of the source data automatically and also the proposed model successfully predicted how many students would drop out over that time.

In [21], the SVRQ model is proposed for student dropout prediction, the parameters are optimized by the IQPSO algorithm. The optimization algorithm improves the performance of the student's dropout prediction. An accuracy of 92% and an F1-score of 95% are achieved by this SVRQ model. Qiu et al [22] utilize the FSPred model for student dropout prediction and it achieves an accuracy of 86.34%. Panagiotakopoulos et al [23] proposed an early dropout prediction model

and it achieve an accuracy of 91.00%. In comparison to this, Xing et al [24] and Chen et al [25] achieve an accuracy of 95.01% and 92.5% for dropout prediction. This approach is used to personalize and prioritize intervention for at-risk students in MOOCs by using individual dropout probabilities. While compared to the literature review dropout prediction models, the proposed EDLNet model achieves better results and it effectively predicts the dropout of the student at an early stage. The proposed EDLNet model has the main benefit of preventing overfitting and having no negative effects on network performance due to the classification and segmentation process.

The following conclusions are obtained by empirical analysis. First, when it comes to the length of the course, shorter courses have lower dropout rates than longer ones. To lower the online learning dropout rate, instructors should employ particular subjects. Second, the dropout rate is reduced by social-interactive engagement. To reduce students' feelings of isolation and disconnection, it is important to encourage students to engage in more online activities. For dropout rate reduction, additionally, learner experience is extremely beneficial. The dropout rate is also reduced by increasing experience because collecting experiences is the process of engagement. Lastly, the course itself has an impact on the dropout rate. Less difficult courses have lower dropout rates than more challenging ones.

5. CONCLUSION

For MOOC dropout prediction, this research proposed an EDLN model. The process of training and testing is performed successively after being converted into a two-dimensional temporal matrix form from the public dataset for the 2015 KDD Cup. The EDLN model was compared to five baseline models, and tests with varying time matrix specifications and attention mechanism-based ablation were carried out. During the first five weeks, the dropout situation is accurately predicted by the EDLN model based on learner behavior and characteristic data. The results of the experiment demonstrate that with significant information. Based on better predictive performance, less time required, F1 values, recall, precision, and accuracy, the proposed EDLN model outperforms other baseline models.

While comparing to the baseline models, in addition to creating dropout prediction models that are more accurate, the deep learning methodology will also provide a reliable method to help with intervention design for reducing the dropout rate. The research object for this study was a MOOC dropout prediction problem and ResNet-50, Faster RCNN, and static attention are combined in the proposed model for the prediction of dropout in MOOCs to effectively resolve online learning platforms with a high dropout rate.

The interpretability of the MOOC dropout prediction will be the main area of research in the future. We also

examine the model's basis further to produce predictions that match the learner's states, specifically, the correlation between the learner behavior features and the model's prediction results as well as the learner behavior features. Furthermore, based on a more detailed search of inaccurate predictions, we will try to significantly enhance the current model.

6. REFERENCE

[1] R. B. Basnet, C. Johnson, T. Doleck, "Dropout prediction in Moocs using deep learning and machine learning", Education and Information Technologies, Vol. 27, 2022, pp.1-15.

[2] H. Aldowah, H. Al-Samarraie, A. I. Alzahrani, N. Alalwan, "Factors affecting student dropout in MOOCs: a cause and effect decision-making model", Journal of Computing in Higher Education, Vol. 32, No. 2, 2020, pp. 429-454.

[3] W. Wang, L. Guo, L. He, Y.J. Wu, "Effects of social-interactive engagement on the dropout ratio in online learning: insights from MOOC", Behaviour & Information Technology, Vol. 38, No. 6, 2019, pp. 621-636.

[4] Y. Mourdi, M. Sadgal, H. El Kabtane, W. B. Fathi, "A machine learning-based methodology to predict learners' dropout, success, or failure in MOOCs", International Journal of Web Information Systems, Vol. 15, No. 5, 2019, pp. 12-24.

[5] S. Ardchir, M. A. Talhaoui, H. Jihal, M. Azzouazi, "Predicting MOOC dropout based on learner's activity", International Journal of Engineering & Technology, Vol. 7, No. 32, 2018, pp.124-126.

[6] J. Chen, J. Feng, X. Sun, N. Wu, Z. Yang, S. Chen, "MOOC dropout prediction using a hybrid algorithm based on a decision tree and extreme learning machine", Mathematical Problems in Engineering, 2019, pp. 20-39.

[7] S. Dass, K. Gary, J. Cunningham, "Predicting student dropout in self-paced MOOC course using random forest model", Information, Vol. 12, No. 11, 2021, pp.476.

[8] Y. Zheng, Z. Gao, Y. Wang, Q. Fu, "MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series", IEEE Access, Vol. 8, 2020, pp. 225324-225335.

[9] A. A. Mubarak, H. Cao, I. M. Hezam, "A deep analytic model for student dropout prediction in massive open online courses", Computers & Electrical Engineering, Vol. 93, 2021, p.107271.

[10] T. Y. Yang, C. G. Brinton, C. Joe-Wong, M. Chiang, "Behavior-based grade prediction for MOOCs via time series neural networks", IEEE Journal of Selected Topics in Signal Processing, Vol.11, No. 5, 2017, pp.716-728.

[11] C. Jin, "Dropout prediction model in MOOC based on clickstream data and student sample weight", Soft Computing, Vol. 25, No. 14, 2021, pp. 8971-8988.

[12] M. Youssef, S. Mohammed, E. K. Hamada, B. F. Wafaa, "A predictive approach based on efficient feature selection and learning algorithms' competition: Case of learners' dropout in MOOCs", Education and Information Technologies, Vol. 24, No. 6, 2019, pp. 3591-3618.

[13] M. Şahin, "A comparative analysis of dropout prediction in massive open online courses. Arabian Journal for Science and Engineering, Vol. 46, No. 2, 2021, pp. 1845-1861.

[14] J. Kabathova, M. Drlik, "Towards predicting student dropout in university courses using different machine learning techniques", Applied Sciences, Vol. 11, No. 7, 2021, p.3130.

[15] K. Coussement, M. Phan, A. De Caigny, D. F. Benoit, A. Raes, "Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model", Decision Support Systems, Vol. 135, 2020, p.113325.

[16] S. Sood, M. Saini, "Hybridization of cluster-based LDA and ANN for student performance prediction and comments evaluation", Education and Information Technologies, Vol. 26, No. 3, 2021, pp. 2863-2878.

[17] S. Lee, J. Y. Chung, "The machine learning-based dropout early warning system for improving the performance of dropout prediction", Applied Sciences, Vol. 9, No. 15, 2019, p. 3093.

[18] A. A. Mubarak, H. Cao, W. Zhang, "Prediction of students' early dropout based on their interaction logs in the online learning environment", Interactive Learning Environments, Vol. 30, No. 8, 2020, pp. 1-20.

[19] P.M. Moreno-Marcos, P.J. Muñoz-Merino, J. Maldo-nado-Mahauad, M. Perez-Sanagustin, C. Alario-Hoyos, C. D. Kloos, "Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs", Computers & Education, Vol. 145, 2020, p. 103728.

[20] C. Ye, G. Biswas, "Early prediction of student drop-out and performance in MOOCs using higher granularity temporal information", Journal of Learning Analytics, Vol. 1, No. 3, 2014, pp. 169-172.

[21] C. Jin, "MOOC student dropout prediction model based on learning behavior features and param-eter optimization", Interactive Learning Environ-ments, 2020, pp. 1-19.

[22] L. Qiu, Y. Liu, Y. Liu, "An integrated framework with feature selection for dropout prediction in mas-sive open online courses", IEEE Access, Vol. 6, 2018, pp. 71474-71484.

[23] T. Panagiotakopoulos, S. Kotsiantis, G. Kostopou-los, O. Iatrellis, A. Kameas, "Early dropout predic-tion in MOOCs through supervised learning and hyperparameter optimization", Electronics, Vol. 10, No. 14, 2021, p. 1701.

[24] W. Xing, D. Du, "Dropout prediction in MOOCs: Us-ing deep learning for personalized intervention", Journal of Educational Computing Research, Vol. 57, No. 3, 2019, pp. 547-570.

[25] W. Chen, C. G. Brinton, D. Cao, A. Mason-Singh, C. Lu, M. Chiang, "Early detection prediction of learn-ing outcomes in online short-courses via learning behaviors", IEEE Transactions on Learning Technol-ogies, Vol. 12, No.1, 2018, pp.44-58.

[26] H. Wan, K. Liu, Q. Yu, X. Gao, "Pedagogical interven-tion practices: Improving learning engagement based on early prediction", IEEE Transactions on Learning Technologies, Vol. 12, No. 2, 2019, pp. 278-289.