# Learning Analytics Model for Predictive Analysis of Learners Behavior for an Indigenous MOOC Platform (Tadakhul System) in Oman

Amala Nirmal Doss[1,*], Reshmy Krishnan[1], Aruna Devi Karuppasamy[2], and Baby Sam[3]

[1]Department of Computing, Muscat College, Sultanate of Oman
[2]Faculty of Computing Sciences, Gulf College, Sultanate of Oman
[3]New York City Department of Civil Administration Services, USA
Email: amala@muscatcollege.edu.om (A.N.D.); reshmy@muscatcollege.edu.om (R.K.); arunadevi@gulfcollege.edu.om (A.D.K.); mail@babysam.biz (B.S.)
*Corresponding author

Manuscript received October 31, 2023; revised November 22, 2023; accepted December 10, 2023; published July 15, 2024

***Abstract*—The Massive Open Online Courses (MOOCs) platforms are widely used by the learner community all over the world. Using MOOC, the learners can choose the course of interest and learn it in their own pace. The main problem encountered in most MOOC based learning is the lack of a learning analytic system that monitors the learners' interaction with their enrolled courses. This problem leads to incompletion or discontinuation of the course. In this paper, 'Tadakul' system, an original bilingual (English and Arabic) MOOC platform is developed for students of higher education institutions of the Sultanate of Oman. Using the Tadakhul system, learners of various higher education institutions of Oman can enroll themselves in various courses of their interest and complete the course at their own pace. This research aims in understanding the impact of learning analytics system used in the Tadakhul system. A novel deep-learning approach used in the system monitors the learning process of the learners based on their interaction with the enrolled courses. The feedback thus obtained, using the deep-learning approach, can be used to improve the student learning experience. To analyze the feedback obtained from the learners, an innovative approach of combining Bidirectional Long Short-Term Memory (BiLSTM) network with a Convolutional Neural Network (CNN) is proposed. The BiLSTM model performs well in analyzing the sequential data whereas the CNN model is good in extracting the spatial features at each hidden layer that are important for evaluating the student's learning patterns. Thus the proposed model can identify learner's learning behavior and learning styles that help teachers better understand the individual needs. The results of the experimental study revealed that the proposed model outperformed conventional machine learning approaches in predicting learner's learning behavior. Hence, the results obtained by integrating BiLSTM and CNN models on the Tadakhul platform can improve the student experience by making teaching more efficient and effective.**

***Keywords*—Tadakhul, learning analytics, MOOCs, deep-learning, student's review analysis**

## I. Introduction

The emergence of online e-learning platforms has altered the learning pattern in the educational sector. These platforms addressed the issues of offline teaching methods and had a significant impact on the educational sector in the Sultanate of Oman. The e-learning platform provides a wide range of learning opportunities for everyone around the world and provides flexibility in the learning pathway to achieve their goals and enhance their skills. The students have a facility to seek additional help to recall the concepts at their own pace. The proposed Tadakhul platform is intended to address several areas including instructional skills, grading and assessment, student retention, classroom supervision, personalized learning, and assessment for the learners in the Sultanate of Oman.

The Massive Open Online Courses (MOOCs), provide easy access to learning materials through online platforms. MOOCs made a huge impact in the higher educational sector, and the people in industry are allowed to learn new skills or enhance their skills at their own pace. The MOOC allows the learners to personalize their courses and provides timely support by considering their learning preferences. The prediction of student learning behavior, addressing the issues related to dropouts and providing personalized learning are the key challenges of MOOC. This can be addressed by incorporating learning analytics to the MOOC [1–3].

The Tadakhul system is an online learning platform proposed to facilitate interactive and personalized learning for multiple users in both Arabic and English languages. Our proposed Tadakhul system is an e-learning environment that provides learning materials for different learning styles. This study focuses on using learning analytics on the Tadakhul platform, which uses the data based on the student's previous learning experience to predict future learning processes. It helps in providing a personalized learning experience. The Tadakhul system provides a platform where the classroom discussions focus on the most important concepts and personalize learning resources to each student based on their performance during their learning sessions. Applying learning analytics to the data obtained in the online e-learning environment can help in classroom management where students face many issues, and the teachers can help by adjusting the data or providing additional support when needed. The advantage of using learning analytics in the MOOC platform is to predict student failure, identify at-risk students, assess assignments, predict grades, and predict results [4]. In this study, we explore different Artificial Intelligence models such as BiLSTM and CNN to identify the learning behavior of the student utilizing the Tadakhul platform. The BiLSTM model is very good at processing time-series data and understanding the relationship between the data. This is important for tracking and interpreting student progress and interaction patterns over time. The ability to detect physical activity is important for identifying long-term learning and behavior. On the other hand, CNN models are effective at extracting spatial features, which

doi: 10.18178/ijiet.2024.14.7.2123

allows them to identify the learning patterns of the students that help in analyzing their student behavior: level of participation and responses to different types of learning. The deep learning models provide a thorough analysis of student's academic behavior, providing insights that were previously difficult to obtain. With a deeper understanding of these learning behaviors, teachers can have a better idea to design courses and content to ensure a more personalized and meaningful experience for students studying on the Tadakhul platform. For example, it may show that student participation or performance is better after using the teaching strategies presented by the model analysis. This empirical evidence will strongly support the results of combining BiLSTM and CNN models in analyzing and improving student learning in Tadakhul.

The paper's structure is as follows: In Section II, an overview of related studies is presented, and Section III provides detailed insights into the deep learning models we propose. Section IV outlines the results of various experiments conducted on the dataset. Our paper concludes by summarizing the key findings from the deep learning models and proposes future research directions for investigating the learning behavior of students who utilize the Tadakhul system.

## II. Related Work

Learning analytics in Massive Open Online Courses (MOOCs) plays an important role in improving the educational journey by providing valuable insights into student behavior and academic achievement. Predictive learning analytics have been employed in research for purposes encompassing dropout prediction, anticipation of final grades, identification of at-risk learners, and projecting course completion. Since the inception of MOOCs, learning analytics has primarily focused on a significant aspect: predicting dropout rates [5–12], predicting student's grades [13–16], and assessing student engagement in courses [17, 18].

In this context, few studies concentrate on the analysis of student's feedback through the utilization of learning analytics. Conventional machine-learning techniques were employed to examine and interpret the feedback provided by students. The algorithm extracts meaningful insights, patterns, and trends from the feedback data, enabling a deeper understanding of student's opinions about the courses and preferences. Various machine learning algorithms including Naive Bayes, Support Vector Machine, Logistic Regression, Linear Regression, k nearest neighbor, Decision Tree, and Random Forest were used to extract valuable insights from the student's comments [19–24]. Graesser [25] proposed an Intelligent Tutoring System that asks questions and provides tasks that contain problem descriptions, comments, and answer summaries. They utilize artificial intelligence algorithms to adapt learning experiences to the needs of individual students. Intelligent tutoring systems can assess and adjust to student's knowledge and abilities with the use of artificial intelligence [26]. Effective early interventions and a thorough understanding of the underlying causes are essential to address the retention and dropout issues among students. To identify students who are at risk of dropping out and to organize early interventions, machine learning models have received a lot of attention [27].

Artificial Intelligence (AI) can enhance the efficiency of student study sessions, allowing students to focus more effectively on the learning material. AI can leverage student's past academic achievements to design future learning experiences, promoting a more personalized learning journey [28]. Through artificial intelligence, students receive quick and individualized feedback on student's academic progress, which in turn promotes the students learning progress. Though artificial intelligence acts as a valuable tool for learning, it is important to remember the irreplaceable role of human interaction in the educational process. Artificial Intelligence often makes suggestions that do not align well with students learning needs. Recently, deep learning algorithms have received much attention and shown better performance compared to existing machine learning methods. Long-Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) represent widely used deep learning techniques for evaluating student feedback [29–37]. These methods help students make informed decisions about course selection and instructors and improve their knowledge and learning experience.

Applying the learning analytic model to Tadakhul enhances student's learning experience by providing personal insights and feedback. The deep learning models: BiLSTM and CNN can analyze more complex input data like review texts to understand student's learning preferences and student's personal issues in understanding the concepts. Students receive recommendations of learning resources or assignments that are personalized based on their specific learning needs. Researchers demonstrated the advantages of deep learning over traditional methods in terms of superiority in processing big data and analyzing complex patterns. The studies [29, 30, 36, 37] found that deep learning models can predict the student performance with greater accuracy than other models. From this literature review, these deep learning models provide information that can transform educational technology, making it more responsive to student's needs. The deep learning models provide a thorough analysis of student's academic behavior, providing insights that were previously difficult to obtain.

## III. Methodology

The proposed method combines two different deep neural networks with the main objective of predicting ratings based on reviews as shown in Fig. 1. The following section discusses the proposed model that combines Convolutional Neural Network (CNN) with Bidirectional Long Short-Term Memory (BiLSTM) Network.

### A. Pre-Processing

In the pre-processing step, the input data needs to be refined to structured data. Initially, all the input review texts are changed into lowercase followed by the elimination of any special characters, tabs, newline characters, and extra whitespaces. The texts are split into separate words or tokens using tokenization process. The lemmatization process is used to identify similar words by rooting the words to their simplest meaning.

To ensure efficient batch processing in LSTM models, it is essential to pad or truncate these sequences to a consistent length. Second, in neural network models that use GloVe (generalized vectors to represent words) as an embedding layer, GloVe helps transform text data into dense vector representations, effectively capturing the semantic essence of words, making it easier for neural networks to process and understand the underlying context. These vectors representations are designed to retain the semantic meaning of the individual words they represent. In other words, each vector acts as a numerical sum that combines the contextual and textual variations of the corresponding word.

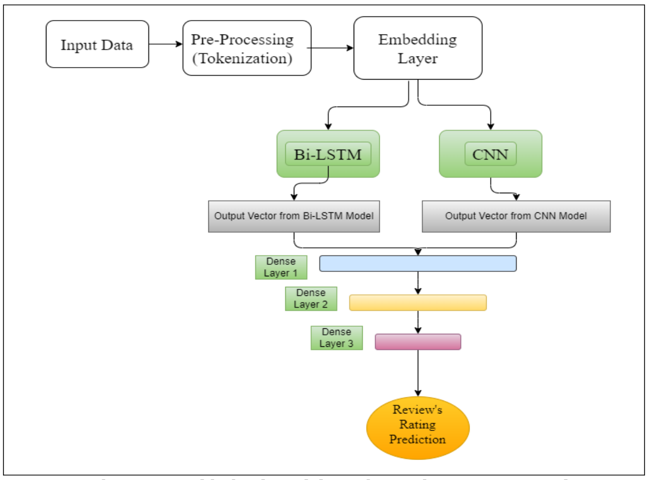


Fig. 1. The proposed hybrid model combines the Bi-LSTM and CNN networks to predict the ratings of course reviews.

### B. Convolutional Neural Network (CNN)

Convolutional Neural Network is one of the most important deep learning techniques in computer vision that has received great attention since 2012 and is widely used in image classification tasks. A CNN architecture of this model consists of three important layers such as a convolutional layer, a pooling layer, and a fully connected layer:

1) Convolutional Layer: Being the first layer of CNN, it performs convolutional operations on the input data represented as an embedding matrix with the predetermined filters or weights. These filters slide across the input to produce an output known as feature maps. The size of these filters is important in the feature extraction which determines the types of features that are extracted from the layer. The output of the convolutional layer is obtained through the following equation,

$$FeatureMap_i = X_i \times W_i + b_i \tag{1}$$

where $X_i$ represents the raw input data (review text) or feature maps from the previous layer, and $W_i$ represents the weights to be learned and applied in the convolution operation and $b_i$ represents the basis of the current layer.

2) Pooling Layer: Pooling layer is followed by each convolutional layer to reduce the size of the feature maps. The pooling window is used over the feature map to reduce the dimensions and retain the important features that are very important in the prediction tasks. Max-pooling and average pooling are commonly used pooling layer to extract the important features.

3) Fully Connected (Dense) Layer: The last layer in the CNN was the fully connected layer, where each neuron in this layer is connected to every neuron in the previous and next layers. This dense layer serves as the place for in-depth feature extraction and is often followed by a classifier like Softmax for prediction tasks.

### C. Bi-Long Short-Term Memory Network

The Bidirectional Long Short-Term Memory (BiLSTM) network is a type of Recurrent Neural Network (RNN) designed to reduce the problems of learning long-term dependencies by traditional RNNs, such as the vanishing-gradient and exploding gradient problems. BiLSTM is particularly useful for sequence prediction problems, natural language processing, time-series forecasting, and many other applications where the order of the input data is important. The main components of Bi-LSTM network are as follows:

1) LSTM Layers: The LSTM units are the core of LSTM layers and for more complex representation of LSTM, the stack multiple LSTM layers are used. Each LSTM layer accepts a sequence of vectors {x1, x2, x3, ..., xn} as its input. Only one vector, xt is passed at each time step t. There are two states employing a major role in LSTM: (i) Hidden States: The hidden state ht for each time step t encapsulates information from the current input xt, and the preceding elements in the sequence. It can be used as the output for that time step or serve as an input for subsequent layers; (ii) Cell States: The cell state Ct serves as the "long-term memory" for the LSTM unit, storing critical information that the network might need for later time steps. It's updated based on what the LSTM decides to remember or forget. LSTM uses the following equation to learn features.

$$C_t = f_t \times C_{t-1} + i_t \times C'_t \tag{2}$$

$$h_t = o_t \times \tanh(C_t) \tag{3}$$

where $C_t$ is the current cell state, $f_t$ is the forgot gate, $i_t$ is the input gate, $o_t$ is the output gate, $C'_t$ is the candidate gate and $h_t$ is the hidden state.

2) Dense (Fully Connected) Layers: Final layer that extract features from the previous layer and helps in performing classification based on the features extracted by the LSTMs.

## IV. Experimental Analysis and the Results

In our experimental analysis, we focus on predicting course review ratings using a hybrid model combining Bidirectional Long-Short-Term Memory (BiLSTM) networks with Convolutional Neural Networks (CNNs). This complex model aims to exploit the strengths of both frameworks by capturing sequential dependencies in text with BiLSTM and local features with CNN. The steps used to process the text for feeding into the model are as follows: (i) Text preprocessing: Before feeding the review text into the model, we preprocess the text, and tokenize sentences into words, (ii) Word Embedding: Pre-trained word embeddings like GloVe is used to convert each tokenized word into a dense vector that captures its semantic meaning, and (iii) Model: The proposed predictive models employs a hybrid architecture that fuses the capabilities of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks.

In this hybrid model, the pre-processed word vectors, also known as embeddings are given as input into a series of convolutional layers. These convolutional layers provide the important function of identifying local patterns or dependencies within text, such as relationships between adjacent or nearby words by effectively capturing features such as bigrams and trigrams. After the convolutional layers have executed their role in local feature extraction, the subsequent pooling layers take over. The pooling operation reduces the size of feature maps produced by the convolutional layers, preserving the most salient aspects while reducing the computational load. Bi-LSTM layers work in this space, processing the raw input data. Bi-LSTMs are particularly good at recognizing long-range dependencies within sequences.

### *A. Dataset*

The dataset we used for our analysis was obtained from Coursera's course review collection available on the Kaggle platform [38]. This dataset is a set of 107,018 unique reviews. Each entry in this dataset is structured to include three key elements: a CourseID that serves as a unique identifier for each course, the Review text that provides quality feedback from users, and the rating, which provides a numerical rating of the course on a specified scale from 1 to 5 as shown in Fig. 2. To prepare for model training and evaluation, we partitioned this extensive dataset into two subsets. Specifically, 80% of the data, or approximately 85,614 reviews, were allocated for training the model. The remaining 20%, constituting about 21,404 reviews, were set aside for testing purposes.

We chose to use course review ratings from Coursera Platform as our dataset for two compelling reasons. First, these estimates closely reflect the type of data we intend to collect from the Tadakhul system once implemented. Second, Course review assessments are extensively engaged by students, making them a rich and relevant data source for our study. Coursera's widespread use among student populations ensures that the insights we gain are relevant and applicable, strengthening the practical relevance of our research. Choosing course review assessments as our dataset was strategically aligned with our research objectives, thereby enhancing both the validity and applicability of our experimental results.

| Id | Review | Label |
|---|---|---|
| 0 | good and interesting | 5 |
| 1 | This class is very helpful to me. Currently, I... | 5 |
| 2 | like!Prof and TAs are helpful and the discussi... | 5 |
| 3 | Easy to follow and includes a lot basic and im... | 5 |
| 4 | Really nice teacher!I could got the point eazl... | 4 |

Fig. 2. The sample dataset which contains course review and the ratings of course reviews.

### *B. Performance Evaluation*

Performance Metrics: The metric used to measure the effectiveness of the model in this study is accuracy. This test is specially designed for situations where the objective is to measure the proportion of correctness among all the data. In the context of predicting a student's academic behavior, accuracy will help us understand how well a pattern is identified or predicts a behavior.

Parameter Adjustments and Tuning: In our research, we conducted experiments using hybrid CNN and Bi-LSTM models on different training parameters and used various optimization algorithm to adjust the network weights: (i) Number of Epochs: The number of epochs refers to the number of times the data set is returned by the neural network. We experimented with different numbers of times to determine the best that the model could learn efficiently, (ii) Batch Size: The batch size is an important parameter that defines the number of samples to work through before updating the internal model parameters. Variety of batch sizes were checked to identify the best optimal batch size for the computational efficiency and model performances, and (iii) Optimizers: The chosen optimizer is very important for minimizing the loss and improving the performance of the models.

We employed different optimizers such as Adam, RMSprop and Stochastic Gradient Descent (SGD). The optimizer that adjusts the learning weights, reduces the convergence speed and minimizes the loss to improve the model performance is chosen as the best optimizer. Adam optimizer uses momentum components which help it smoothly to identify the global minima instead of getting into the trap of local minima like SGD. Also, Adam optimizer uses minimal memory and best for its computational efficiency. Text data, such as course reviews, is typically noisy and exhibits non-stationary characteristics (meaning its statistical properties shift over time). Adam's adaptive learning rate mechanism generally handles such noisy data more adeptly compared to SGD or RMSProp. In essence, Adam's strengths lie in its ability to adjust learning rates for individual parameters, its incorporation of momentum, and its effectiveness with large, noisy datasets, leading to quicker convergence. This makes it particularly apt for tasks like analyzing text data to determine student ratings from course reviews. However, it's crucial to note that the "best" optimizer may vary based on the unique attributes of the dataset and problem. Practically, experimenting with various optimizers and fine-tuning their hyperparameters is often recommended to discover the most efficient strategy for a specific task. In this study, we experimented with 25 training epochs using the hybrid CNN and Bi-LSTM models and applied Adam optimizer to adjust the network weights. In our experiment, the model was trained for a total of 25 times using Adam Optimizer. We compared the performance of optimization methods, including RMSprop and Stochastic Gradient Descent (SGD) in our experiment. However, the Adam Optimizer consistently produced better results.

We evaluated our proposed hybrid model against models such as LSTM-CNN, BiLSTM, LSTM and other traditional machine learning approaches. We used Term Frequency-Inverse Document Frequency (TF-IDF) and Bag-of-Words (BOW) with different classifiers as shown in the Figs. 3‒7. The evaluation result showed that our hybrid model demonstrated better accuracy. This proposed model yielded an accuracy rate of 84% of training accuracy and 78% of testing accuracy, demonstrating the effectiveness of the model in predicting course's review ratings as shown in Table

1. In practice, the information derived from this model can be useful to improve the efficiency of the Tadakhul platform. Teachers can use this information to personalize instruction, and the platform can be optimized to meet student needs and improve the overall learning experience.

Table 1. Performance evaluation

| Model | Training Accuracy | Test Accuracy |
|---|---|---|
| BiLSTM-CNN | 85% | 78% |
| LSTM-CNN | 83% | 76% |
| BiLSTM | 81% | 77% |
| LSTM | 75% | 74% |
| TF-IDF-mL | - | 77% |
| BOW-ML | - | 76% |

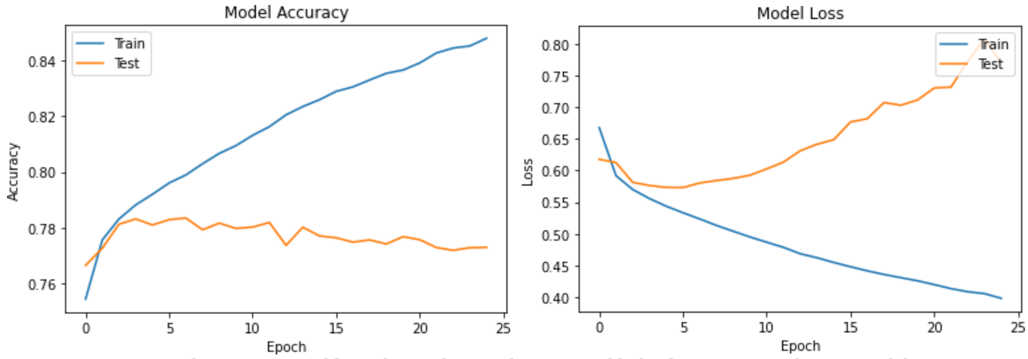



Fig. 3. The accuracy and loss obtained using the proposed hybrid Bi-LSTM and CNN model.

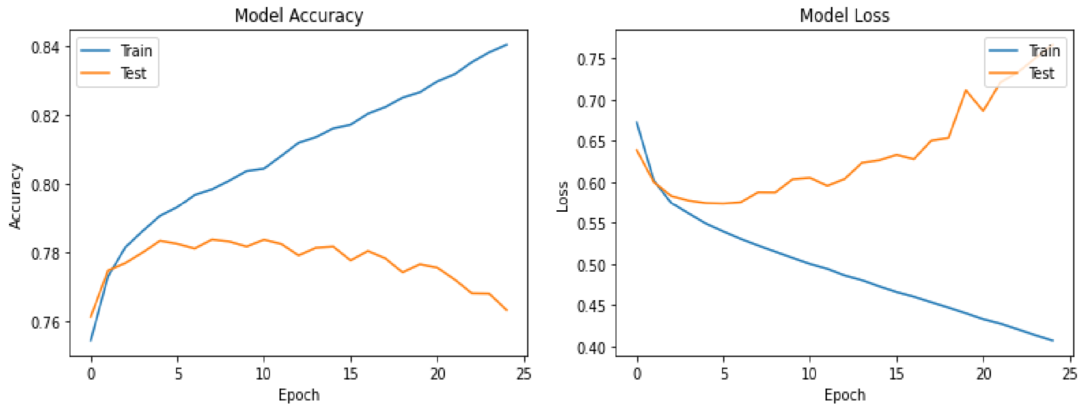



Fig. 4. The accuracy and loss obtained using the proposed hybrid LSTM and CNN model.

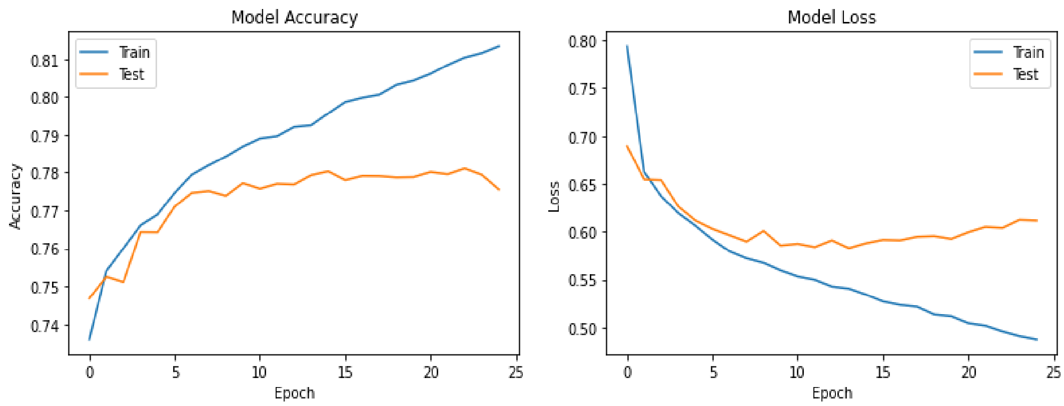



Fig. 5. The accuracy and loss obtained using the Bi-LSTM model.

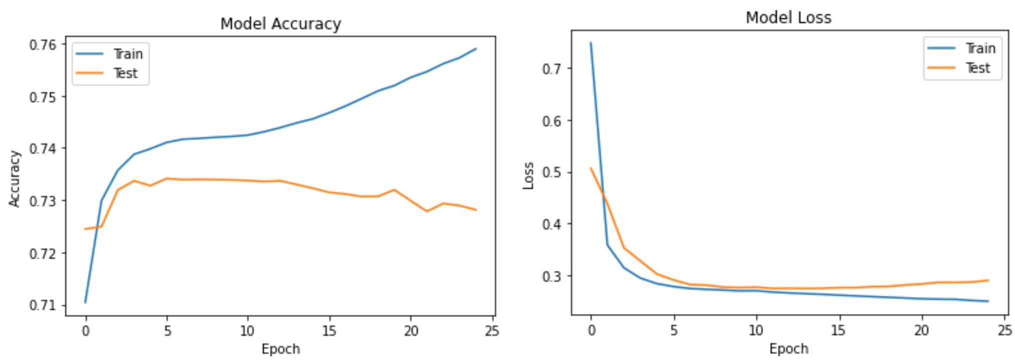
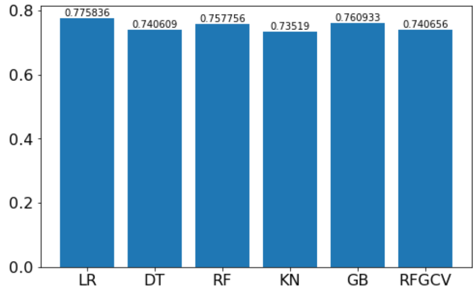



Fig. 6. The accuracy and loss obtained using the LSTM.



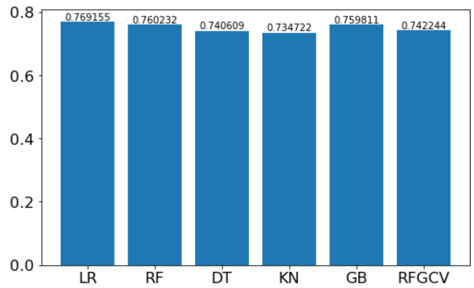



Fig. 7. The accuracy obtained using the TF-IDF and BOW Machine Learning methods with different classifiers Logistic Regression (LR), Random Forest (RF), Decision Trees (DT), K-Nearest neighbor (KN), Gradient Boosting (GB), Random Forest with Grid Search (RFGCV).

### *C. Limitation and Complexity of the Proposed Model*

Our proposed Tadakhul system faces few challenges in predicting certain types of student behavior. (i) Situations in which student engagement vary widely and it is difficult to establish a clear pattern. (ii) The lack of sufficient data due to less or minimal student interaction with the platform will make it difficult for models to accurately predict behavior. (iii) When input data lacks clear indicators of specific learning behaviors, it becomes difficult to make accurate predictions. One limitation is that it relies on information provided by the Tadakhul platform. Future research should aim to include more diverse data to ensure the model is valid across different students and learning environments. The research shows potential, but validating the results requires studying real-world applications.

## V. Discussion and Conclusion

In this research, we introduced a novel hybrid model that combines Bi-directional Long-Short-Term Memory (BiLSTM) with Convolutional Neural Networks (CNN). The motivation behind proposing this hybrid model is to explore the strengths of both sequential processing (provided by Bi-LSTMs) and spatial feature detection (provided by CNNs) to capture the complex patterns present in review data. We experimented with various other models and approaches. Specifically, we evaluated: (i) CNN in combination with LSTM: This approach combines the advantages of LSTM's sequence processing with CNN's ability to capture local patterns within the review text, (ii) LSTM Model: Here, we used LSTM networks to find out their individual performance in processing sequential data especially for text-based course reviews, (iii) BiLSTM model: We also used Bi-directional LSTM model, BiLSTMs to obtain a better understanding of the context in which words appear in course reviews by processing data from both forward and backward directions, (iv) Traditional Machine Learning models: We employed different machine learning algorithms TF-IDF and BOW as a benchmark for comparative analysis. These methods, which include classifiers such as Random Forests and Support Vector Machine, are used to extract features from review evaluations of courses and provide a basis for measuring the performance of our proposed models. The proposed hybrid CNN-BiLSTM model is designed to be computationally efficient and highly effective in capturing the complex features inherent in review texts. By leveraging the strength of CNNs in understanding local text patterns and the efficiency of BiLSTMs in modeling long-term dependencies, our model aims for well-rounded and accurate prediction of course review ratings.

We conclude that our research by presenting an advanced deep learning approach aimed at understanding and predicting student interactions with online course materials and to understand and predicting student interactions with online course materials, conclude that in the context of massive open online courses (MOOCs). Our research is based on a novel combination of Bidirectional Long-Short-Term Memory (BiLSTM) networks and Convolutional Neural Networks (CNN), specifically designed to analyze course feedback collected from students using the Tadakhul system. Our experimental findings strongly indicate that the proposed hybrid model outperforms traditional machine learning approaches. The improved performance in predicting student's learning behaviors, as demonstrated by our empirical analysis, indicates the performance, and promise of integrating Bi-LSTM with CNN for this application. It shows that these observations can be used not only to accurately predict student interactions and learning patterns, but also to provide feedback and support mechanisms that can significantly enrich the overall student learning experience. The performance of our model in using complex patterns in the curriculum for predictive analytics represent a significant advancement in the development of intelligent, data-driven educational platform. The proposed models have shown promising results, investigating other deep learning or ensemble models can provide commendable knowledge from the observation to improve predictive capabilities. Analyzing the data over a period can help to understand information about students' long-term learning patterns and the effectiveness of different learning activities. This study contributes to the field of learning data analysis by demonstrating the effectiveness of hybrid models in determining difficult learning behaviors. On the Tadakhul platform, insights from models can be used to refine training content and strategies for improving personal experiences. This has the potential to transform learning and lead to effective and engaging teaching. Thus, this research serves as a foundation for future work aimed at refining and expanding the use of deep learning techniques to better understand and improve student engagement and success in online learning environments.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

The research idea came from author Amala. Author Amala and Reshmy conducted the research; Author Aruna did the data collection and analysis of the research; Author Baby reviewed and analyzed the results of the findings; Author

Amala and Reshmy wrote the paper; all the authors approved the final version.

## Funding

This research was funded by Ministry of Higher Education, Research and Innovation – Oman (under project proposal ID: MC001/MoHERI/RG/2021)

## References

[1] K. Ahmad, J. Qadir, and l-Fuqaha *et al.*, "Data-driven artificial intelligence in education: A comprehensive review," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 12–31, 2020.
[2] R. Ferguson, "Learning analytics: drivers, developments and challenges," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5–6, pp. 304–317, 2012.
[3] G. Siemens, "Learning analytics: The emergence of a discipline," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013.
[4] Z. Papamitsiou and A. A. Economides, "Learning analytics for smart learning environments: A meta-analysis of empirical research results from 2009 to 2015," *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, vol. 1, 23, 2016.
[5] S. B. Kotsiantis, C. Pierrakeas, and P. E. Pintelas, "Preventing student dropout in distance learning using machine learning techniques," in *Proc. Knowledge-Based Intelligent Information and Engineering Systems: 7th International Conference, KES 2003*, 2003, pp. 267–274.
[6] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Computers and Education*, vol. 53, no. 3, pp. 950–965, 2009.
[7] G. Balakrishnan and D. Coetzee, "Predicting student retention in massive open online courses using hidden markov models," *Electrical Engineering and Computer Sciences University of California at Berkeley*, vol. 53, pp. 57–58, 2013.
[8] K. R. Koedinger, J. Kim, J. Z. Jia, E. A. McLaughlin, and N. L. Bier, "Learning is not a spectator sport: Doing is better than watching for learning from a mooc," in *Proc. the Second (2015) ACM Conference on Learning@ Scale*, 2015, pp. 111–120.
[9] J. Liang, C. Li, and L. Zheng, "Machine learning application in moocs: Dropout prediction," in *Proc. 2016 11th International Conference on Computer Science and Education (ICCSE)*, IEEE, 2016, pp. 52–57.
[10] C. Ye, J. S. Kinnebrew, and G. Biswas *et al.*, "Behavior prediction in moocs using higher granularity temporal information," in *Proc. the Second (2015) ACM Conference on Learning@ Scale*, 2015, pp. 335–338.
[11] J. Dillon, N. Bosch, and M. Chetlur *et al.*, "Student emotion, co-occurrence, and dropout in a mooc context," *International Educational Data Mining Society*, 2016.
[12] W. Li, M. Gao, H. Li, Q. Xiong, J. Wen, and Z. Wu, "Dropout prediction in moocs using behavior features and multi-view semisupervised learning," in *Proc. 2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 3130–3137.
[13] T. C. Russo and J. Koesten, "Prestige, centrality, and learning: A social network analysis of an online class," *Communication Education*, vol. 54, no. 3, pp. 254–261, 2005.
[14] A. Wojciechowski, and L. B. Palmer *et al.*, "Individual student characteristics: Can any be predictors of success in online classes," *Online Journal of Distance Learning Administration*, vol. 8, no. 2, p. 13, 2005.
[15] G. Garman *et al.*, "A logistic approach to predicting student success in online database courses," *American Journal of Business Education (AJBE)*, vol. 3, no. 12, pp. 1–6, 2010.
[16] A. Zafra and S. Ventura, "Multi-instance genetic programming for predicting student performance in web based educational environments," *Applied Soft Computing*, vol. 12, no. 8, pp. 2693–2706, 2012.
[17] M. Cocea and S. Weibelzahl, "Cross-system validation of engagement prediction from log files," in *Proc. Creating New Learning Experiences on a Global Scale: Second European Conference on Technology Enhanced Learning, EC-TEL 2007*, Crete, 2007, pp. 14–25.
[18] M. L. Bote-Lorenzo and E. G. Anchez, "Predicting the decrease of engagement indicators in a MOOC," in *Proc. the Seventh International Learning Analytics and Knowledge Conference*, 2017, pp. 143–147.
[19] R. S. Baker, T. Martin, and L. M. Rossi, "Educational data mining and learning analytics," *The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, And Applications*, pp. 379–396, 2016.
[20] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Sentiment analysis: Towards a tool for analysing real-time students feedback," in *Proc. 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. IEEE*, 2014, pp. 419–423.
[21] X. Chen, M. Vorvoreanu, and K. Madhavan, "Mining social media data for understanding students' learning experiences," *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 246–259, 2014.
[22] A. Ortigosa, J. M. Mart'ın, and R. M. Carro, "Sentiment analysis in facebook and its application to e-learning," *Computers in Human Behavior*, vol. 31, pp. 527–541, 2014.
[23] A. Koufakou, J. Gosselin, and D. Guo, "Using data mining to extract knowledge from student evaluation comments in undergraduate courses," in *Proc. 2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 3138–3142.
[24] J.-A. P. Lalata, B. Gerardo, and R. Medina, "A sentiment analysis model for faculty comment evaluation using ensemble machine learning algorithms," in *Proc. the 2019 International Conference on Big Data Engineering*, 2019, pp. 68–73.
[25] A. C. Graesser, X. Hu, B. D. Nye *et al*., "Electronixtutor: An intelligent tutoring system with multiple learning resources for electronics," *International Journal of STEM Education*, vol. 5, pp. 1–21, 2018.
[26] X. Chen, H. Xie, and G.-J. Hwang, "A multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers," *Computers and Education: Artificial Intelligence*, vol. 1, 100005, 2020.
[27] N. Mduma, K. Kalegele, and D. Machuve, "A survey of machine learning approaches and techniques for student dropout prediction," *Data Science Journal*, vol. 18, no. 14, pp. 1–10, 2019.
[28] Z. Ju and R. Qu, "A research on using artificial intelligence information technology to enhance the effectiveness of English language learning—An example of learning college English," in *Proc. 2021 International Conference on Computers, Information Processing and Advanced Education (CIPAE). IEEE*, 2021, pp. 1–4.
[29] O. B. Coelho and I. Silveira, "Deep learning applied to learning analytics and educational data mining: A systematic literature review," in *Proc. Brazilian Symposium on Computers in Education*, vol. 28, no. 1, 2017, p. 143.
[30] O. Corrigan and A. F. Smeaton, "A course agnostic approach to predicting student success from vle log data using recurrent neural networks," in *Proc. Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017*, Tallinn, Estonia, 2017, pp. 545–548.
[31] V. D. Nguyen, K. Van Nguyen, and N. L.-T. Nguyen, "Variants of long short-term memory for sentiment analysis on Vietnamese students' feedback corpus," in *Proc. 2018 10th International Conference on Knowledge and Systems Engineering (KSE). IEEE*, 2018, pp. 306–311.
[32] C.-W. Tseng, J.-J. Chou, and Y.-C. Tsai, "Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members," *IEEE Access*, vol. 6, pp. 72870–72879, 2018.
[33] L.-C. Yu, C.-W. Lee, H. Pan, C.-Y. Chou, P.-Y. Chao, Z. Chen, S. Tseng, C. Chan, and K. R. Lai, "Improving early prediction of academic failure using sentiment analysis on self-evaluated comments," *Journal of Computer Assisted Learning*, vol. 34, no. 4, pp. 358–365, 2018.
[34] S. Srinivas and S. Rajendran, "Topic-based knowledge mining of online student reviews for strategic planning in universities," *Computers and Industrial Engineering*, vol. 128, pp. 974–984, 2019.
[35] D. Dess`ı, M. Dragoni, G. Fenu, M. Marras, and D. R. Recupero, "Evaluating neural word embeddings created from online course reviews for sentiment analysis," in *Proc. the 34th ACM/SIGAPP Symposium on Applied Computing*, 2019, pp. 2124–2127.
[36] M. L. B. Estrada, R. Z. Cabada, R. O. Bustillos, and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Systems with Applications*, vol. 150, 113265, 2020.
[37] P. Ren, L. Yang, and F. Luo, "Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis," *Education and Information Technologies*, vol. 28, no. 1, pp. 797–814, 2023.
[38] KAGGLE. (2021). Kaggle. [Online]. Available: https://www.kaggle.com/datasets/septa97/100k-courseras-course-reviews-dataset