

A CNN–LSTM-based deep learning model for early prediction of student’s performance

Monika Arya¹, Anand Motwani²,
Kauleshwar Prasad³,
Bhupesh Kumar Dewangan^{4,*},
Tanupriya Choudhury^{5,*} and
Piyush Chauhan^{6,*}

¹Department of Computer Science and Engineering, CMR Engineering College, Kandlakoya, Medchal, Hyderabad, Telangana 501401, India

²Department of Computer Science, School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh 466114, India

³Department of Computer Science and Engineering, BIT, Bhilai House Durg, Chhattisgarh 491001, India

⁴Department of Computer Science and Engineering, OP Jindal University, Raigarh, Chhattisgarh, 496109 India

⁵School of Computer Science, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand, 248007, India

⁶Department of Computer Science and Engineering, Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, 412115, India

*E-mails: piyush.chauhan@sitnagpur.siu.edu.in; Shbichauhan@gmail.com, bhupesh.dewangan@opju.ac.in; bhupesh.dewangan@gmail.com, Tanupriya@ddn.upes.ac.in; tanupriya1986@gmail.com

Received for publication
September 10, 2024.

Abstract

In issues pertaining to higher education, deep learning (DL), and its connection to educational data, it is crucial to forecast students’ success. The ability to predict a student’s success aids in choosing courses and developing future study schedules. Apart from forecasting children’s performance, it also assists educators and administrators in keeping an eye on pupils, offering them support, and incorporating training initiatives to maximize outcomes. Student prediction has the advantage of lowering official warning flags and removing ineffective pupils from the classroom. By helping students select courses and study schedules that are suited for their skill levels, prediction supports the students directly. In the proposed approach, a methodology based on the integration of convolutional neural network (CNN) and long short-term memory (LSTM) is proposed to optimize students’ performance prediction systems. This study utilized a student performance dataset from the UCI ML Repository. It includes information on student achievements in secondary education from two Portuguese schools. The suggested method overcomes the following three problems in model development: an imbalanced dataset, a lack of feedback mechanism to enhance the quality of learning, and an inadequate mechanism to extract the learning patterns/relevant features to predict student performance. The suggested system’s effectiveness has been demonstrated by the accuracy (98.45) and loss (0.1989) obtained to achieve the best prediction.

Keywords

convolutional neural network (CNN), long short-term memory (LSTM), deep learning, machine learning, Prediction

I. Introduction

One of the most important and effective factors in the growth of the nation is education. It now lasts a life-time rather than being a one-time occurrence. One explanation is that people's working life are changing so quickly that they must continue to study throughout their employment. Student performance is a crucial aspect of the learning process. Forecasting learning outcomes is critical for identifying learners who are more likely to have difficulty academically in the future. If these data are converted into knowledge, it may be useful to improve teaching and learning quality while also assisting students in reaching their academic objectives. Because of the critical importance of this issue in the growth of countries all over the world, predicting student performance is becoming increasingly important. Predicting student performance is becoming extremely important because of its critical importance in the overall growth of the country. The development of a country is dependent on its education system because it is the education system that produces individuals capable of taking challenges that will march the country towards growth in all aspects of life. Additionally, the assessment of student's performance reflects the effectiveness of educational establishments, which are in charge of molding future generations to correspond with the various phases of the lives of individuals in each nation. Early performance evaluations of students might help them see their strengths and shortcomings and do better on exams. Since students encounter numerous issues that have a detrimental impact on their entire careers, many researchers have concentrated on creating new methods for evaluating their academic performance. There are many factors on which a student's performance depends, such as, student learning environment, mentor support, family background of a student, and mental state of a student.

a. Need for student performance evaluator

In a traditional classroom where the number of students is limited, numerous characteristics of interactions allow the teacher to assess an individual student's degree of participation rather efficiently. The rapid increase in student numbers makes it challenging for even experienced instructors to perform personalized assessments. In this situation, an automated system that precisely predicts how pupils will perform in real-time could be useful. It would be a valuable tool for determining when and with whom to conduct live educational interventions, aiming to

boost engagement, provide incentives, and support student success. It can also assist policymakers and school administrators in determining the best strategies for improving the educational quality.

Evaluating student performance requires a lot of physical effort, time, and correct information about the learning atmosphere to prevent bias. As a result, it is necessary to automate evaluating student performance by projecting their performance considering every factor.

Teachers and students can adjust their teaching and learning processes with early student performance prediction. Hence, weak individuals can receive an early warning to improve their academic standard. As a result, students can restructure their study methods, producing positive outcomes.

b. Challenges in developing an early student performance detection system

1. Presently, there is an excessive increase in the number of students. Due to the enormous growth in the number of students, even experienced human instructors find it impractical to conduct such personalized assessments of each student.
2. These days, every field generates enormous amounts of data due to the technology development. Since the data sizes may surpass processing volume and storage space limitations or evaluation capacity with standard methods, it is difficult to acquire insight into this massive amount of data for early student performance prediction and to investigate the circumstances that contributed to the failure.
3. Data imbalance exists in the datasets, increasing the danger of model overfitting.
4. The creation and optimization of Artificial intelligence (AI) models—that is, the use of various algorithms to create models with improved prediction accuracy—has received most of the attention in the literature on AI prediction models. Nevertheless, there is a dearth of use of AI prediction models that incorporate real-time feedback and provide teachers and students with pertinent feedback to enhance the quality of learning for students.
5. Various institutes' approaches for predicting student success are inadequate in comprehending the learning patterns of their pupils, which has a negative impact on their performance.
6. Because of distance limitations, determining how well students succeed in virtual learning environments can be difficult.

7. Ignorance of key aspects, analysis restrictions due to the number of facts available, and incertitude in student statistics are among the many of these problems.

c. Motivation

Achieving good performance is always guided by evaluation. Similarly, obtaining positive results depends heavily on the evaluation of student achievement. However, the student's dataset is growing so quickly that it is difficult to get improved outcomes. The lack of proper guidance and related problems cause a lot of students to give up on their professions. Providing the proper guidance on their route is essential since students are the foundation of the future world. Many experts are working to create a way to gauge pupils' performance because of technological advancements. Thus, motivated by this fact, this study focuses on developing an effective method for the early identification of student performance. As a result, a child's enormous potential can be developed and ensure their success throughout life.

d. Contribution

In this study, to optimize students' performance prediction system, a methodology based on the integration of convolutional neural network (CNN) and long short-term memory (LSTM) is proposed to overcome the following three problems in model development: an imbalance dataset, a lack of feedback mechanism to enhance the quality of learning, and an inadequate mechanism to extract the learning patterns/relevant features to predict student performance. The main advantage of LSTM cells is their ability to retain long-term information through a unique memory cell, unaffected by current inputs or outputs. This helps them learn long-term dependencies and avoid issues such as vanishing or exploding gradients. However, CNNs are particularly good at learning spatial organization from the input data. Thus, by combining CNN and LSTM, the raw data's spatial and temporal characteristics may be considerably extracted, increasing the accuracy of the student performance prediction model. Additionally, since LSTMs use feedback connections, they can process entire sequences of data rather than just individual data points. This capability makes them highly effective at recognizing and predicting patterns in data. The main contributions of this study are as follows:

- (a) Handle the imbalanced dataset.
- (b) Extract relevant features and hidden patterns for model development to improve the accuracy of the model.

- (c) Improve the overall accuracy of the model.
- (d) Apply a feedback mechanism to improve learning.
- (e) Determine the performance of the model using various evaluation metrics.
- (f) Compare the suggested methodology with similar methods such as the simple LSTM model, LSTM with dropouts, and bidirectional LSTM.

e. Paper organization

The student performance prediction system and its importance and challenges, along with the motivation and contribution of this study, are briefly presented in Section 1. Section 2 discusses the existing relevant and related works associated with a summary table for comparison of existing methodologies. Section 3 provides the suggested approach. The results and discussion of the proposed method are provided in Sections 4 and 5, respectively, and the overall work and future work is concluded and discussed in Section 6, respectively.

II. Literature review

Kim et al. [1] proposed GritNet, an algorithm based on a deep learning (DL) approach. It is built on bidirectional LSTM. The proposed work takes the student's learning activity as input and searches the most discriminative sequence to predict the student's performance. Aslam et al. [2] used a DL model along with synthetic minority oversampling technique (SMOTE) to resolve the dataset imbalance issue. Deep neural network (DNN) with multiple hidden layers that updated automatically was proposed by Li and Liu [3]. Feedforward and backpropagation were also integrated to predict students' performance. Machine learning (ML) along with data preprocessing techniques to remove redundant and missing values from the dataset was employed by Hussain and Khan [4]. Ouyang et al. [5] suggested an integrated approach of AI with an in-time feedback mechanism for early prediction of student performance. Arashpour et al. [6] proposed a hybridized approach of Support vector machine (SVM) and Artificial Neural Networks (ANN). Simulation of traditional ML techniques and CNN over five datasets to predict student performance was done by Mohammad et al. [7]. An integrated approach using ANN and LSTM was suggested by Al-azazi et al. [8]. Kukkar et al. [9] employed LSTM network consisting of four layers along with Random forest (RF) and GB to forecast whether students will pass or fail. Rahul et al. [10]

introduced a search capsule based on “Deep Auto Encoder” (TSCNDE) to perceive the performance of students. The study was performed by Chen et al. [11] in which college students’ entrepreneurial competency was examined using 10 characteristics that were identified from interviews. The validity and reliability of the data gathered via questionnaires were examined, and the regression analysis verified the efficacy of the components. The most crucial component, according to an analytic hierarchy process (AHP) model, is entrepreneurial knowledge, which is followed by ability and intrinsic potential. Regression analysis and AHP combined to evaluate and rank the factors is new. The adoption of e-learning platforms in biology education and their effects on student performance are also examined in some studies. It was discovered that perceived utility, simplicity of use, attitude, flexibility, and content quality significantly affect students’ desire to adopt e-learning and increase performance using a mixed methods

approach with structural equation modeling (SEM) and independent sample *t*-tests. The main goals of the useful suggestions made for educators and legislators are to improve platform usability and content quality and support for a wide range of student demands [12]. By categorizing students’ learning styles using ML algorithms based on the Felder-Silverman Learning Style Model (FSLSM), a work done by Hasibuan et al. [13] tackles the difficulty of tailoring e-learning systems. Teachers can create personalized learning routes and modify lesson plans with the help of predictive models. Results from the pretests and posttests indicated that students were more motivated and satisfied, with satisfaction increasing from 60% to 87% when learning styles were considered.

Table 1 summarizes the dataset used, the evaluation parameters, and limitations/future scope of the related and relevant work done in past for student’s performance prediction.

Table 1: Summary of related work

Reference	Dataset used	Evaluation metrics	Limitations/future scope
[1]	Datasets of two Udacity ND programs: ND-A and ND-B	ROC	In future, the indirect data can be incorporated into the GritNet.
[2]	WOU, XAPI, UCI, and AV student performance datasets	Precision, recall, F-score, and accuracy	Further validation on other large size imbalanced datasets is required.
[3]	Real data were collected from a multidisciplinary university	MAE and RMSE	Reliability of the system can be improved further by updating layers of NN
[4]	Dataset acquired from the “Khyber Pakh tunkhwa Board of Intermediate & Secondary Education” Peshawar	Accuracy and RMS	In big data environment, the DL models need to be integrated with traditional ML techniques. RNN model that updates learning rate is required to maximize precision of the prediction framework
[5]	Three different datasets	Accuracy	Educational contexts such as course subjects must be taken into account. Sample size aspect that verifies the empirical research results and implications is overlooked
[6]	Open University dataset	Precision, recall, accuracy, F1, and FM	This study basically focused on online mode of study, and in future, other modes can be analyzed. Hybrid models of TLBO optimization, ANN, and SVM were evaluated. Evaluation and comparison with other hybrid models is required to achieve more reliable predictions of academic performance
[7]	House, WOU, XAPI, UCI, and AV dataset	Accuracy	In future, the CNN model can be squeezed to reduce CNN structures

(Continued)

Table 1: Continued

Reference	Dataset used	Evaluation metrics	Limitations/future scope
[8]	OULAD	Precision, recall, F1-score, and accuracy	As a limitation of this study, researchers should consider that MOOC students generate large clickstream records, and DL techniques require significant training time, which can delay data processing and evaluation of results.
[9]	OULAD	Precision, recall, F1-score, and accuracy	With fewer data streams, the system achieves lower accuracy, while more data streams improve prediction performance. In the future, data from various institutions and study areas will be collected to assess performance variations. Additionally, other DL and ML algorithms will be integrated to better understand relationships among student academic attributes and enhance prediction accuracy
[10]	OULA dataset	Accuracy, sensitivity, specificity, and precision	Limitation: It uses only a single dataset and limited performance metrics for predicting student performance, which affects its overall effectiveness. To achieve better results, more data should be considered. Future research can expand this work by using multiple datasets for student performance prediction

AV, analytics Vidhya; CNN, convolutional neural network; DL, deep learning; FM, Fowlkes-Mallows index; MAE, mean absolute error; ML, machine learning; MOOC, MAssive Open Online Courses; ND, nanodegree; NN, Neural Network; OULA, Open University Learning Analytics; OULAD, Open University Learning Analytics Dataset; RMS, root mean square; RMSE, root mean square error; ROC, receiver operating characteristic; TLBO, teaching-learning-based optimizer; UCI, University of California-Irvine; WOU, Western Ontario University; XAPI, experience API.

III. Proposed method

a. Data gathering

In this study, we utilized a student performance dataset from the UCI ML Repository. It includes information on student achievements in secondary education from two Portuguese schools [7]. It includes various attributes, such as student grades, demographics, social factors, and other features (school-related), which were collected through questionnaires and school reports. The dataset consists of two separate subsets representing performance in the following two distinct subjects: Portuguese language (por) and Mathematics (mat). In total, the dataset comprises of 32 attributes and 649 records.

b. Checking and dealing with missing values

The dataset is checked for missing values, and if present, then different methods are used to deal with missing values. In our dataset, no missing value is present.

c. Checking for class distribution ratio and handling imbalanced data

The data are highly imbalanced, with a class distribution ratio of 550:99 for the majority to the minority class. Figure 1 shows the class distribution of the student performance dataset.

The imbalanced class distribution leads to a biased learning model, which results in lower

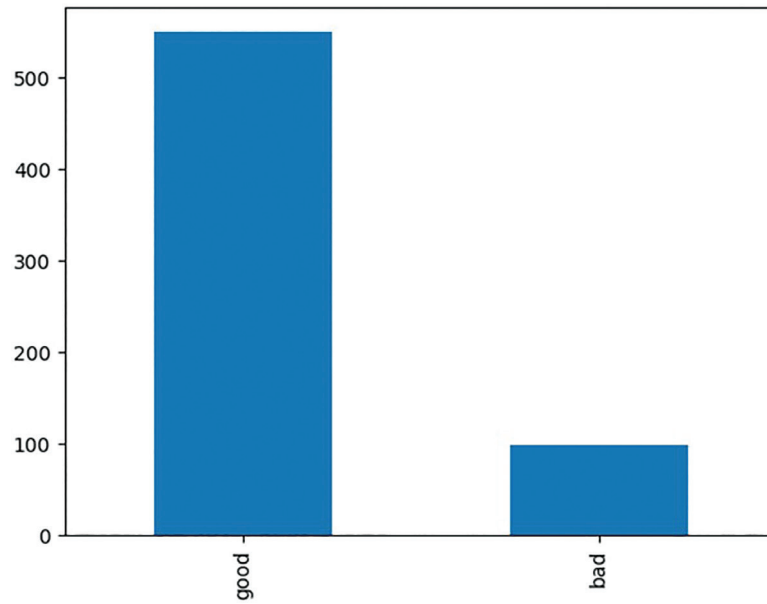


Figure 1: Class distribution of dataset.

Table 2: Descriptive analysis

	School	Sex	Age	Address	Family size	...	Absences	G1	G2	G3	Performance
Count	649	649	649	649	649	...	649	649	649	649	649
Unique	2	2	8	2	2	...	24	27	16	17	2
Top	GP	F	17	U	GT3	...	0	10	11	11	Good
Frequency	423	383	179	452	457	...	244	95	103	104	550

predictive accuracy for the minority classes compared to the majority classes. There are various approaches such as oversampling and under sampling to handle imbalanced data. In this study, adaptive synthetic sampling (ADASYN) method is used. ADASYN is a type of oversampling technique. This method works based on the density of minority class instances [14]. The generation of new samples is inversely proportional to the density of minority class samples. It creates more samples in the feature space region where the density of minority class examples is low or none and less samples in the high-density space. The adaptive approach is beneficial for the model using different datasets

with different class distribution ratios. In this study, after applying ADASYN oversampling technique, the shape of original and resample datasets is as follows:

original dataset shape: Counter ({'good': 550, 'bad': 99})

Resample dataset shape Counter ({'bad': 550, 'good': 550})

d. Hidden pattern recognition

To discover the hidden patterns in the dataset, descriptive analysis is performed. Table 2 depicts the descriptive analysis of data.

From descriptive analysis of the dataset, the following conclusions can be drawn:

- i. The female students are more in number than the male students.
- ii. The family size of most of the student is >3 .
- iii. Most of the students belong to urban area.
- iv. The frequency of G3 is near G1 and G2 where G3 is the target attribute, i.e., final grades of the students. Additionally, G1 and G2 are the grades of first and second periods, respectively. Thus, the final grades correlates highly with grades of first and second periods.

e. Integration of CNN and LSTM for students' performance prediction

There is a one-dimensional spatial structure to the student data, and the CNN could be able to identify invariant patterns that distinguish between good and poor performance [15]. An LSTM layer may then be used to classify this newly learned spatial feature [16].

The CNN model used comprises of the following five layers:

1. The first input layer converts the sample into a (58, 32) shape vector.
2. The second is a pooling layer, also called Max-Pooling1D. It splits the size of the convolutional layer's feature maps into half using a length and stride of 2. The pooling layer produces an output with a shape of (29, 32).
3. The third layer is a flatten layer. It enables the processing of the output by traditional, fully connected layers.
4. At last, two dense layers are used – the first dense layer is a fully connected layer with 64 neurons and a rectifier activation function. The last layer is the output layer. It has three neurons for the three categories of classes and a SoftMax activation function.

The batch size of 32 and epochs 10, 50, 100, and 200 were used.

The embedding layer passes the combined features to the LSTM, followed by the addition of a one-dimensional CNN and max pooling layers. A modest collection of 32 characteristics and a short filter length of 3 are employed in the suggested

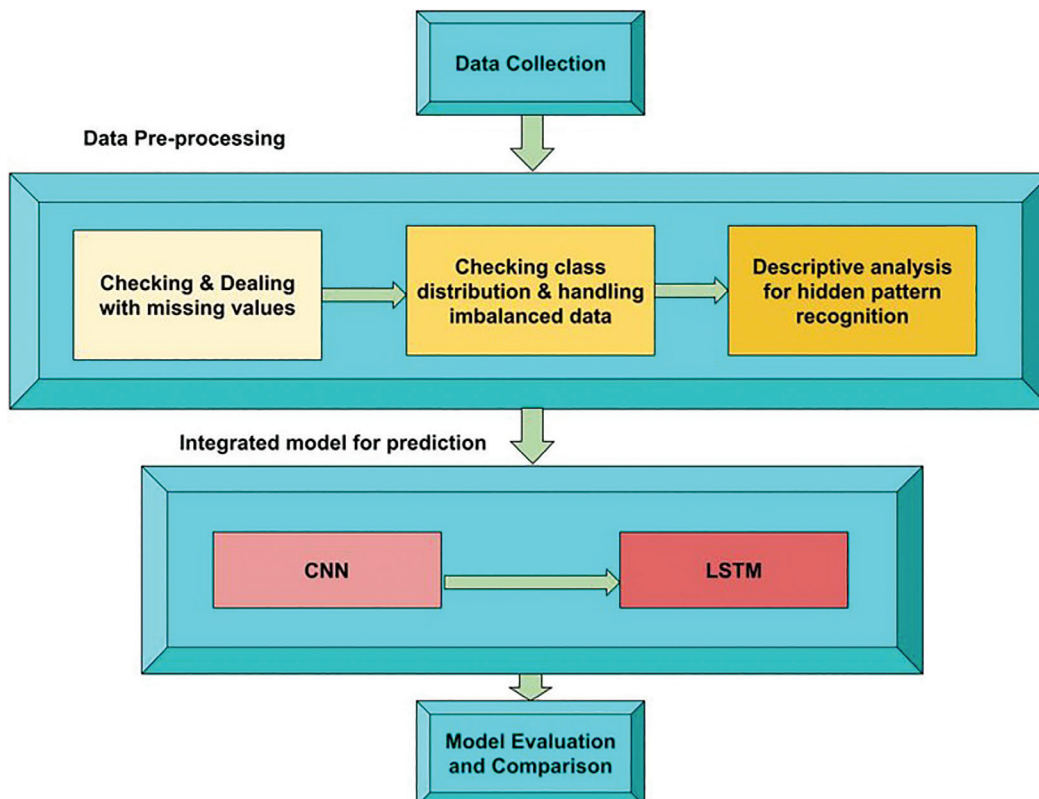


Figure 2: Flowchart. CNN, convolutional neural network; LSTM, long short-term memory.

method. The feature map size is cut into half by the pooling layer using the typical length of 2.

f. Model evaluation and comparison

The suggested model is evaluated using metrics such accuracy, loss, and training time. To provide a comparison, the performance of similar LSTM models such as simple LSTM, LSTM with dropouts [17], and bidirectional LSTM [18] are also summarized using the same dataset. Figure 2 depicts the flow-chart of the suggested model, in which a predictive modeling pipeline utilizes CNN and LSTM models.

Data collection comes first and then comes data preprocessing, where missing values are filled in, class distribution is examined, and patterns that are concealed are found. After that, an LSTM is used to record temporal patterns, and a CNN is used to extract features in an integrated model for prediction. Sequential and geographical data may be handled efficiently with the help of CNN and LSTM. To evaluate the model's correctness and performance in comparison to other models or benchmarks, it is finally subjected to evaluation and comparison. This procedure guarantees a thorough approach to forecasting.

Table 3: Comparison of LSTM-based student performance prediction models

S.No	Model	Average accuracy	Average training time	Average loss
1	Simple LSTM	85.83	333 ms/step	0.2516
2	LSTM with dropouts	86.68	659 ms/step	0.2334
3	Bidirectional LSTM	88.69	2 s/step	0.2561
4	Proposed CNN + LSTM	98.45	183 ms/step	0.1989

CNN, convolutional neural network; LSTM, long short-term memory.

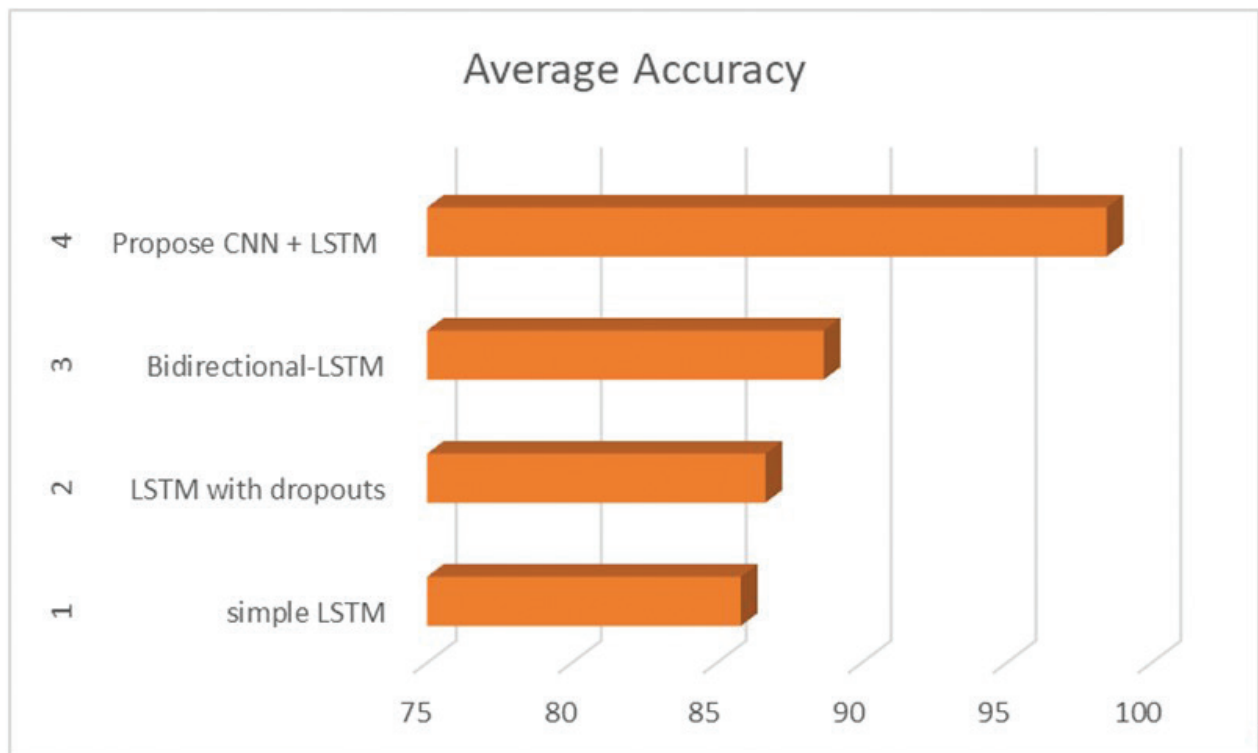


Figure 3: Comparison of average accuracy. CNN, convolutional neural network; LSTM, long short-term memory.

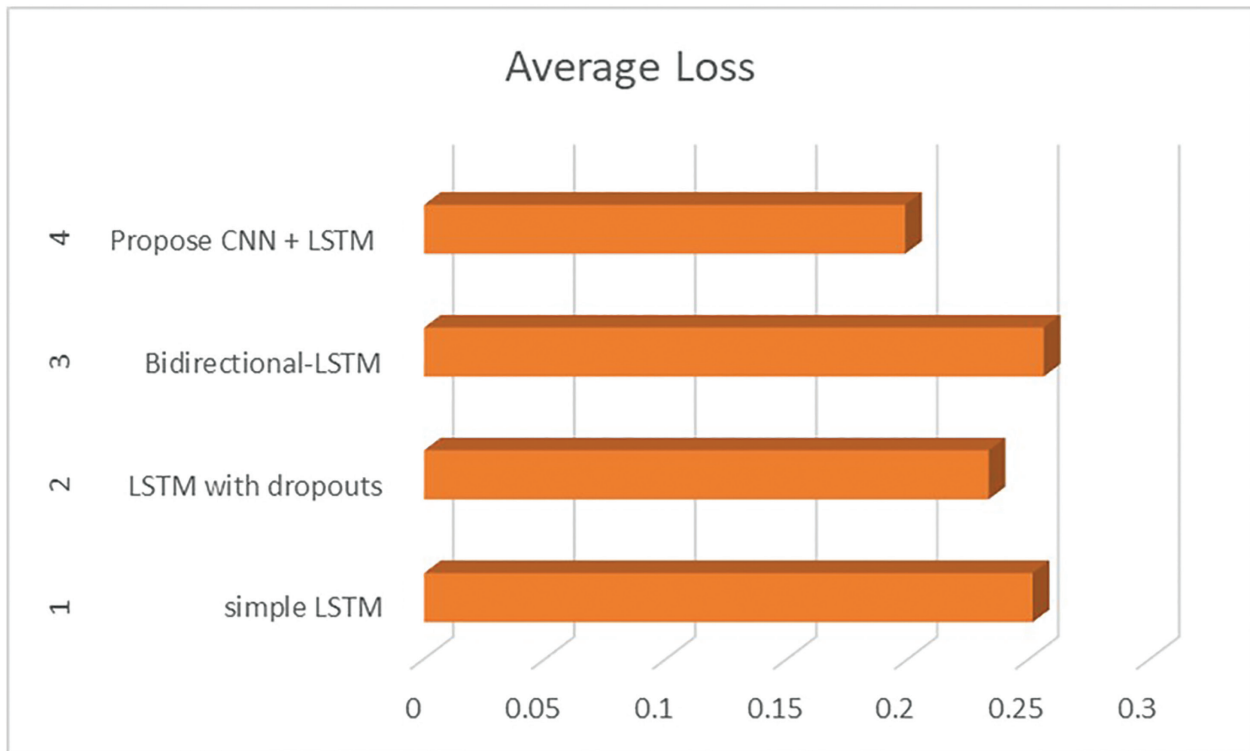


Figure 4: Comparison of average loss. CNN, convolutional neural network; LSTM, long short-term memory.

IV. Results

The evaluation metrics used to evaluate the student performance prediction model are accuracy, loss, and training time. Table 3 shows the performance comparison of the proposed model with a simple LSTM model, LSTM with dropouts, and bidirectional LSTM in terms of average accuracy, training time, and loss for three epochs.

Figure 3 presents a graph comparing the average accuracy of the proposed model with similar approaches and Figure 4 compares the average loss.

V. Discussions

From the above results, it can be concluded that the results are almost at the state of the art when using a basic LSTM with minimal adjustment. LSTM with dropout has the intended effect on training, resulting in a slightly slower convergence trend and, in this instance, a poorer accuracy at the conclusion. Bidirectional LSTM resulted in a little better but requiring a lot more training time. Neither the forward nor the reversed order is always ideal, and using both will still produce superior outcomes. A bidirectional

LSTM network is used in this scenario. In bidirectional LSTM, two independent LSTM networks, one with a forward sequence and the other with a reversed sequence, are combined. After that, the output from the two LSTM networks is concatenated and sent to the network's later layers. Finally, with the proposed method, CNN with LSTM, attained better outcomes, yet faster and with fewer weights.

VI. Conclusion and future scope

In this research, a methodology combining CNN and LSTM was proposed to improve the accuracy of a student academic prediction framework, achieving approximately 98.45% accuracy. To further enhance the prediction accuracy and understand the relationships between different academic features and additional DL and ML algorithms will be integrated. This could help identify critical online learning activities that require further investigation.

This study demonstrated the effectiveness of the DL model, with performance evaluated based on accuracy, loss, and training time. Data imbalance was a challenge, leading to potential overfitting, which was mitigated using ADASYN. The proposed model

outperformed basic LSTM models with the same dataset. Future improvements may include updating extracted features and optimizing neural network weights to enhance system reliability. Additionally, data from other institutions and diverse learning environments will be gathered to assess performance and account for variability.

Funding

We thank for receiving the funding for this from RSF (Research Support Funding) from Symbiosis Institute of Technology, Symbiosis International (Deemed University) (SIU), Lavale Campus, Pune, Maharashtra, 412115, India.

References

- [1] B. H. Kim, E. Vizitei, and V. Ganapathi, "GritNet: Student performance prediction with deep learning," *Proc. 11th Int. Conf. Educ. Data Mining, EDM 2018*, 2018.
- [2] N. Aslam, I. U. Khan, L. H. Alamri, and R. S. Almuslim, "An Improved Early Student's Performance Prediction Using Deep Learning," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 12, pp. 108–122, 2021, doi: 10.3991/ijet.v16i12.20699.
- [3] S. Li and T. Liu, "Performance Prediction for Higher Education Students Using Deep Learning," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/9958203.
- [4] S. Hussain and M. Q. Khan, "Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning," *Ann. Data Sci.*, vol. 10, no. 3, pp. 637–655, 2023, doi: 10.1007/s40745-021-00341-0.
- [5] F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, "Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, pp. 1–23, 2023, doi: 10.1186/s41239-022-00372-4.
- [6] M. Arashpour et al., "Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization," *Comput. Appl. Eng. Educ.*, vol. 31, no. 1, pp. 83–99, 2023, doi: 10.1002/cae.22572.
- [7] A. S. Mohammad, M. T. S. Al-Kaltakchi, J. Alshehabi Al-Ani, and J. A. Chambers, "Comprehensive Evaluations of Student Performance Estimation via Machine Learning," *Mathematics*, vol. 11, no. 14, pp. 1–16, 2023, doi: 10.3390/math1143153.
- [8] F. A. Al-azazi and M. Ghurab, "ANN-LSTM: A deep learning model for early student performance prediction in MOOC," *Heliyon*, vol. 9, no. 4, p. e15382, 2023, doi: 10.1016/j.heliyon.2023.e15382.
- [9] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, "Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms," *Educ. Inf. Technol.*, vol. 28, no. 8, pp. 9655–9684, 2023, doi: 10.1007/s10639-022-11573-9.
- [10] Rahul and R. Katarya, "Deep auto encoder based on a transient search capsule network for student performance prediction," *Multimed. Tools Appl.*, vol. 82, no. 15, pp. 23427–23451, 2023, doi: 10.1007/s11042-022-14083-5.
- [11] Y. Chen and Y. Mei, "Competency Model: A Study on the Cultivation of College Students' Innovation and Entrepreneurship Ability," *HighTech Innov. J.*, vol. 4, no. 4, pp. 821–828, 2023.
- [12] N. Salybekova, S. Abdimalik, G. Issayev, G. Khalikova, A. Berdenkulova, and K. Bakirova, "E-Learning Adoption: Designing a Network-Based Educational and Methodological Course on" Humans and Their Health," *Emerg. Sci. J.*, vol. 7, no. 6, pp. 2097–2119, 2023.
- [13] M. S. Hasibuan, R. Z. A. Aziz, D. A. Dewi, T. B. Kurniawan, and N. A. Syafira, "Recommendation Model for Learning Material Using the Felder Silverman Learning Style Approach," *HighTech Innov. J.*, vol. 4, no. 4, pp. 811–820, 2023.
- [14] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *Proc. Int. Jt. Conf. Neural Networks*, no. 3, pp. 1322–1328, 2008, doi: 10.1109/IJCNN.2008.4633969.
- [15] M. Taye, "Theoretical Understanding of Convolutional Neural Network :," *Computation*, vol. 11, 2023.
- [16] M. Arya and G. Hanumat Sastry, Effective LSTM Neural Network with Adam Optimizer for Improving Frost Prediction in Agriculture Data Stream, vol. 1761 CCIS. Springer Nature Switzerland, 2023.
- [17] J. Billa, "Dropout approaches for LSTM based speech recognition systems," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2018-April, pp. 5879–5883, 2018, doi: 10.1109/ICASSP.2018.8462544.
- [18] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recognition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3697 LNCS, pp. 799–804, 2005, doi: 10.1007/11550907_126.