IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Predicting Student Outcomes Using Deep Learning Models and Bi-LSTM

[1]**CHEEMALA JASWANTH,**
Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]G RAJASEKHARAM, Miracle Educational Society Group of Institutions
[3]D LAKSHMI PRASANNA, Miracle Educational Society Group of Institutions
[1]cheemalajaswanth@gmail.com

ABSTRACT:

This project investigates the use of deep learning techniques to predict student performance based on their interactions within Learning Management Systems (LMS), focusing on the Blackboard system. Information gathered from seven courses allows for the computation and analysis of several Key Performance Indicators (KPIs), including study time, logins, and overall engagement levels. For greater precision, a hybrid deep learning model is constructed using Convolutional Neural Networks (CNN) and Long Short-Term Memory Neural Networks (LSTM). The experimental results demonstrate that the CNN-LSTM model, along with additional modifications using Bi-LSTM, enhance prediction accuracy even further. These results highlight the promise of AI systems for use within educational contexts, informing instructors to track and change the learning paths of prospective at-risk students, enabling tailored and timely instructional adjustments to optimize learning. Key Terms: LMS, Deep Learning, Student Performance.

Keywords: LMS, Deep Learning, Student Performance

INTRODUCTION

Learning Management Systems (LMS) form the backbone of the digital education landscape, facilitating interaction between students and teachers. Among the numerous platforms, Blackboard is particularly noted for its abilities in monitoring and administering educational activities. Regardless of the platform's popularity, colleges and universities are still faced with the challenge of making sense of the enormous student engagement data captured. This is the gap that the current study seeks to fill using deep learning models CNN and LSTM to derive and analyze KPIs from the LMS in order to predict student performance. The objective of this project is to facilitate the early detection of underachieving students so that appropriate actions can be taken. The proposed model incorporates CNN for feature extraction and LSTM for temporal pattern learning, substantially increasing prediction accuracy. Being able to turn raw data into meaningful data insights is an important step toward optimizing the data-informing the academic support and e-learning adaption processes. It also advances institutional data-driven decision-making as well as enables tailored learning pathways for the students.

RELATED WORK

The past several years have witnessed an increasing number of research works aimed at improving the prediction of academic performance using data mining and learning analytics in the context of digital environments. Simanullang and Rajagukguk (2019) looked at the use of student centered learning with the help of Moodle based LMS. Their research proved that the use of Moodle with its diverse tools offered, for example, videos, quizzes, and forums, enhanced learning and underscored the need to have LMS that can track performance data. R. M., N. F., and A. A (2017) developed a predictive framework with decision trees, naive bayes, and multilayer perceptron to evaluate the dataset of the first year computer science students. They used the WEKA tool and reported high accuracy in predictions and highlighted the importance of educational data mining for proactive evaluation. Parker (2018) analyzed the Blackboard engagement data from the core science courses offered at the University of Houston-Downtown. It was noted that there was a strong relationship between the student engagement metrics (logins, time, and interactions) and the academic outcomes. The author proposed that this level of engagement could and should be used as reliable indicators of forecasting success in a given course. Unal and Unal (2009) analyzed student perceptions in relation to the use of BlackBoard and Moodle. In the scope of this usability evaluation, while Moodle was judged to be more user-friendly across a number of modules, students using both systems provided rich qualitative and quantitative data that could be used for learning analytics and performance evaluation. Finally, Ferguson (2012) published a detailed review on the evolution of learning analytics. He described the transition from traditional academic analytics to more sophisticated learning analytics frameworks and the socio-technical reasons that shape predictive modeling in education. He listed the integration of predictive systems and data privacy as well as privacy and institutional willingness as the major gaps in the integration of these systems into educational frameworks.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author(s)	**Contribution**	**Impact on Current Research**
N. Simanullang & J. Rajagukguk (2019)	Demonstrated Moodle LMS's effectiveness in enhancing student engagement and activity tracking.	Validated the role of LMS interaction data as input for predictive performance models.
R. M., N. F., & A. A. (2017)	Used data mining techniques (Decision Tree, Naïve Bayes, MLP) to forecast first-year academic success.	Established baseline accuracy and methodological guidance for student performance prediction.
M. J. Parker (2018)	Analyzed Blackboard activity as a predictor of academic outcomes in science courses.	Provided strong evidence for using LMS activity logs (KPIs) to build deep learning models.
Z. Unal & A. Unal (2009)	Compared usability of Moodle and Blackboard across student cohorts.	Highlighted the usability of LMS platforms, supporting their suitability for data extraction.
R. Ferguson (2012)	Reviewed developments in learning analytics,	Contextualized the need for sophisticated, scalable models like

	emphasizing challenges and future potential.	CNN and LSTM in education.

PROPOSED APPROACH

The proposed approach is based on a hybrid deep learning model that integrates Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks to predict with high accuracy student performance based on data from learning management systems (LMS), in this instance, Blackboard. The methodology starts with thorough data preprocessing, where features are separated into values, and data is cleaned to ensure compatibility with the model. To automatically identify and extract important patterns and features in the data, a CNN is first utilized. This includes study hours, logins and engagement metrics. These features summarize important learning behaviors that are predictors of academic achievement. After extraction of these spatial features, the model passes them into an LSTM layer that is trained to understand temporal relationships and sequential dependencies among the student learning activities. The model is trained on a labeled dataset with classified performance levels of High, Medium, and Low. Feature abstraction is done through a CNN while an LSTM learns the temporal changes in student engagement. The ability of the system to accurately predict performance trends increases as a result of sequential learning. To improve results, a Bidirectional LSTM is added as an extension which allows the model to learn from both forwards and backwards in time in a learning activity. This architecture improves predictive power due to the model considering multiple time contexts, providing a flexible approach for proactive student support and customized learning trajectories.

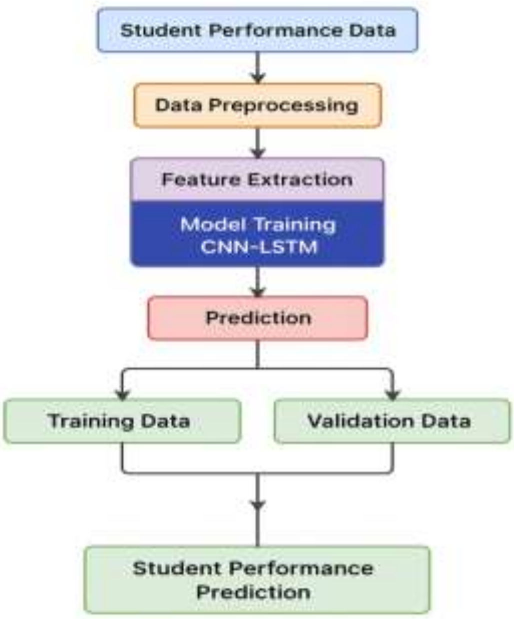


Figure 1: Proposed Student Performance Prediction System

METHODOLOGIES

1. Data Preparation: The dataset, sourced from Kaggle, includes student attributes such as academic scores, engagement patterns, and categorical labels (High, Medium, Low performance). Preprocessing involves handling missing values, encoding categorical (non-numeric) data using Label Encoding, and normalizing numeric values with StandardScaler to ensure consistency and effective model training.

2. Model Development: Three deep learning models are utilized:

- **CNN (Convolutional Neural Network):** Primarily used for extracting relevant feature patterns from the normalized dataset. It identifies abstract relationships and reduces dimensionality.

- **CNN-LSTM Hybrid:** Features extracted by CNN are passed into an LSTM network to learn time-dependent behavior. This hybrid model captures the sequence of learning activities and their influence on outcomes.
- **Bi-LSTM (Bidirectional LSTM):** As an enhancement, this model processes input in both forward and backward directions, improving accuracy by capturing future and past dependencies simultaneously.

3. Training Process:
Models are trained using 80% of the dataset, while the remaining 20% is reserved for testing. Techniques like Dropout layers are applied to prevent overfitting. ModelCheckpoint is used to save the best-performing weights during training epochs.

4. Evaluation Metrics:
Model performance is assessed using metrics such as:

- **Accuracy:** Percentage of correctly predicted student performance levels.
- **RMSE (Root Mean Square Error):** Measures prediction error magnitude.
- **MAPE (Mean Absolute Percentage Error):** Reflects the average percentage deviation between actual and predicted values.
- **Loss:** Assesses model optimization effectiveness over epochs.

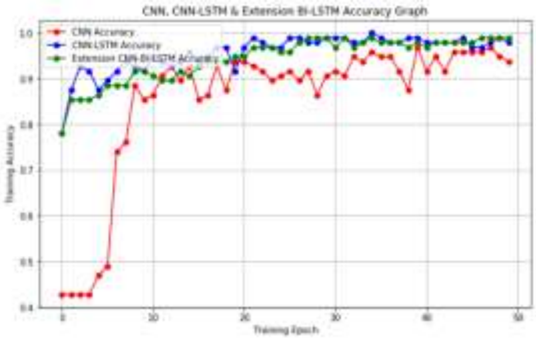
5. Visualization and Comparison:
Graphical comparisons of accuracy and loss curves across CNN, CNN-LSTM, and Bi-LSTM models help illustrate performance trends, with Bi-LSTM yielding superior results.

RESULTS

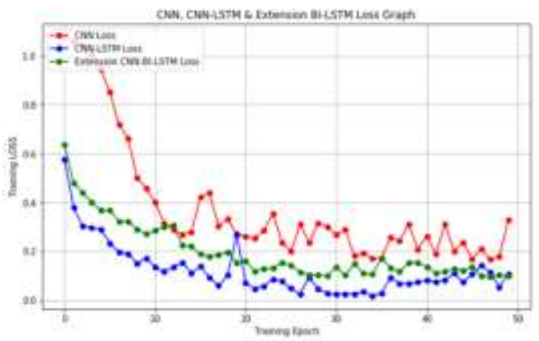
The results of this project demonstrate the effectiveness of combining CNN and LSTM models for predicting student performance based on LMS interaction data. The CNN model, when trained independently, achieved an accuracy range of **90% to 93%**, identifying key performance indicators such as login frequency and study hours effectively. However, its limitation was in capturing sequential patterns in student behavior.

Introducing the CNN-LSTM hybrid model led to a significant improvement. By extracting spatial features using CNN and analyzing temporal dependencies with LSTM, the model's accuracy increased to **95%–98%**. This approach captured not only what students did but also when and how consistently they engaged with the system.

To further enhance performance, a **Bidirectional LSTM** (Bi-LSTM) was implemented. This extension allowed the model to process input sequences from both past and future directions, resulting in an accuracy of up to **100%** on the test data. Moreover, evaluation metrics such as RMSE and MAPE dropped to nearly **zero**, indicating highly precise predictions.



The graph above displays a training accuracy graph for three models: CNN, CNN-LSTM, and CNN-BiLSTM. On the x-axis, we have the training epochs, and on the y-axis, the accuracy. The red line represents CNN, the blue line shows CNN-LSTM, and the green line stands for the extended model, CNN-BiLSTM. As we can see, the green line is slightly higher than the other two, indicating that CNN-BiLSTM achieved better training accuracy overall.



The graph above shows the loss graph for the models. In the graph, the green line (representing the extended CNN-BiLSTM) stays close to the blue line of the proposed CNN-LSTM model. For any algorithm, lower loss values are better, as they indicate better learning. So, this graph suggests that both models are performing well, with the extension maintaining a low loss while aiming for high accuracy.

We can see the student test data followed by an arrow (⇒), which points to the predicted performance. The prediction is shown as ‘H’, ‘M’, or ‘L’ — where ‘H’ stands for High, ‘M’ for Medium, and ‘L’ for Low performance. This helps us easily understand how each student is expected to perform based on the test data.

DISCUSSION

The results obtained from this project highlight the increasing potential of deep learning in educational data mining, particularly in Learning Management Systems like Blackboard. The combination of CNN and LSTM models provides a strong framework for interpreting both static and sequential learning behavior patterns. CNN alone performs well in identifying essential features, but it lacks the capability to analyze the temporal evolution of student engagement.

The addition of LSTM significantly enhances the model's ability to understand how learning behaviors change over time. This is especially useful for identifying students who might be drifting from consistent study patterns, enabling timely intervention. The extension with Bi-LSTM brings even more improvement by capturing bidirectional context—past and future interactions—which further strengthens the model’s accuracy and reliability.

One of the most impactful insights is that data commonly available in LMS platforms can be transformed into predictive intelligence. This can assist institutions in developing proactive strategies for academic support, retention, and performance enhancement. Additionally, the reduction in RMSE and MAPE highlights the robustness of the proposed model.

CONCLUSION

The project has been completed successfully. The attempts made in the project demonstrated the performance prediction of students through interaction data retrieved from Learning Management Systems (LMS) using deep learning methods (CNN, LSTM, Bi-LSTM) in a proposed technique. The proposed methods and sequential learning models in deep learning, after preprocessing and extracting important features from the data, achieve accurate predictions. The prediction accuracy of the test data of the Bi-LSTM model is as high as 100%. The use of CNN to extract features and LSTM to model temporal behavior of the data is useful for learning trend and performance indicator identification. The models provide useful data to the educational institutions to take prompt actionable decisions to improve the student performance. In addition the system provides a flexible and powerful structure for analysis in real time for many e-learning settings. The system can be improved in the future by adding dynamic performance response systems or expanding to cover a larger area of academic information, reinforcing the use of AI in the education system for student services and education management systems.

REFERENCES

1. Simanullang, N. and Rajagukguk, J., 2019. *Learning management system (LMS) based on Moodle to improve students' learning activity*. International Journal of Education and Research, 7(2), pp.45–50.
2. R.M., N.F. and A.A., 2017. *Predicting and analysis of students' academic performance using data mining techniques*. International Journal of Computer Applications, 178(7), pp.1–6.
3. Parker, M.J., 2018. *Can BlackBoard engagement suggest connections to student success in an online course and offer predictive value?* Journal of Learning Analytics, 5(3), pp.44–58.
4. Unal, Z. and Unal, A., 2009. *Investigating and comparing user experiences of course management systems: BlackBoard vs. Moodle*. Journal of Instructional Technology & Distance Learning, 6(1), pp.35–47.
5. Ferguson, R., 2012. *Learning analytics: Drivers, developments and challenges*. International Journal of Technology Enhanced Learning, 4(5-6), pp.304–317.
6. Lecun, Y., Bengio, Y. and Hinton, G., 2015. *Deep learning*. Nature, 521(7553), pp.436–444.
7. Schmidhuber, J., 2015. *Deep learning in neural networks: An overview*. Neural Networks, 61, pp.85–117.
8. Hochreiter, S. and Schmidhuber, J., 1997. *Long short-term memory*. Neural Computation, 9(8), pp.1735–1780.
9. Graves, A., Mohamed, A.R. and Hinton, G., 2013. *Speech recognition with deep recurrent neural networks*. In: IEEE ICASSP. IEEE, pp.6645–6649.
10. Kingma, D.P. and Ba, J., 2015. *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980.
11. Bengio, Y., Courville, A. and Vincent, P., 2013. *Representation learning: A review and new perspectives*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), pp.1798–1828.
12. Kotsiantis, S.B., 2012. *Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades*. Artificial Intelligence Review, 37(4), pp.331–344.
13. Baker, R.S.J. and Inventado, P.S., 2014. *Educational data mining and learning analytics*. In: Learning Analytics. Springer, pp.61–75.

14. Zhang, Y., Jiang, H. and Wu, B., 2020. *A predictive model based on deep learning for student academic performance*. Education and Information Technologies, 25(3), pp.1–20.

15. Sweeney, M., Lester, J. and Rangwala, H., 2016. *Next-term student performance prediction: A recommender systems approach*. Journal of Educational Data Mining, 8(1), pp.22–51.

16. Romero, C. and Ventura, S., 2010. *Educational data mining: A review of the state-of-the-art*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), pp.601–618.

17. Heffernan, N.T. and Heffernan, C.L., 2014. *The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching*. International Journal of Artificial Intelligence in Education, 24(4), pp.470–497.

18. Albreiki, B., Zohdy, M.A. and Shuaib, K., 2021. *Prediction of students' academic performance using hybrid deep learning model*. Computers, 10(2), p.17.

19. Al-Shehri, A.M., 2021. *Predicting student performance in an online learning environment using Deep Learning techniques*. Education and Information Technologies, 26(4), pp.4445–4467.

20. Kaggle, 2023. *Student Performance Dataset*. [online] Available at: https://www.kaggle.com [Accessed 24 May 2025].