IJERST International Journal of Engineering Research and Science & Technology

A Hybrid Deep Learning Model To Predict High-Risk Students In Virtual Learning Environment

[1] Ms. Jahnavi Deepika Gubbala, [2]Garipelly Sahruthi, [3]Anampally Vaishnavi, [4]Pokala Varshini,[5] K. Pravalika

[1]Assistant Professor, Department of Computer Science & Engineering (AI & ML), Malla Reddy Engineering College for Women

[1]Email: jahnavideepika@gmail.com

[2,3,4,5,6]B. Tech Students, Department of Computer Science & Engineering (AI & ML), Malla Reddy Engineering College for Women

[2]Email: sahgrply@gmail.com, [3]Email: anampallyvaishnavi9@gmail.com, [4] Email: pokalavarshini14@gmail.com,

[5] Email: pravalikakotha1199@gmail.com,

ABSTRACT

The rapid growth of virtual learning environments (VLEs) has increased the need for intelligent systems that can identify students who are at risk of poor academic performance or disengagement. This study proposes a hybrid deep learning–based predictive framework for detecting high-risk students by analyzing behavioral and engagement data collected from online learning platforms. The system is implemented as a web-based application using a Flask framework, where student interaction data such as time spent weekly, quiz score average, forum participation, video watching percentage, assignment submission rate, login frequency, session duration, device type, course difficulty, and region are utilized as predictive features. Data preprocessing includes missing value removal, label encoding for categorical attributes, and feature normalization using a standard scaler. The dataset is then divided into training and testing subsets to evaluate model performance. Multiple machine learning and deep learning models—including Convolutional Neural Networks (CNN), XGBoost, Random Forest, and Decision Tree classifiers—are trained and compared to determine the most effective approach for predicting student engagement risk levels. The system classifies students into low, medium, or high engagement risk categories, enabling early identification of learners who may require academic support. Experimental results demonstrate that integrating deep learning with traditional machine learning models enhances prediction accuracy and provides a scalable solution for improving student retention and personalized learning interventions in virtual education systems.

Keywords: Virtual Learning Environments, Student Performance Prediction, Engagement Analysis, Hybrid Deep Learning, CNN, XGBoost, Random Forest, Decision Tree, Educational Data Mining, Flask Web Application.

I. INTRODUCTION

The rapid advancement of digital technologies has significantly transformed the education sector, leading to the widespread adoption of Virtual Learning Environments (VLEs). These platforms provide flexible learning opportunities by enabling students to access course materials, participate in discussions, submit assignments, and interact with instructors remotely. However, despite the advantages of online education, many institutions face challenges in identifying students who are at risk of disengagement or poor academic performance. Unlike traditional classrooms, instructors in virtual environments often lack direct observation of student behavior, making it difficult to detect early signs of learning difficulties or reduced participation.

In recent years, educational data mining and learning analytics have emerged as effective approaches to analyze large volumes of learner interaction data generated within online platforms. By examining behavioral indicators such as time spent on learning activities, quiz performance, login frequency, and participation in discussion forums, it is possible to gain valuable insights into students' learning patterns. Machine learning and deep learning techniques have shown great potential in predicting student performance and identifying those who may require additional support. Early detection of at-risk students enables educators to implement timely interventions, improve student engagement, and increase course completion rates.

This project proposes a Hybrid Deep Learning Model to Predict High-Risk Students in Virtual Learning Environments by leveraging multiple predictive algorithms and behavioral learning features. The system analyzes student engagement metrics including weekly study time, quiz score averages, forum activity, video watching percentage, assignment submissions, login frequency, session duration, device type, course difficulty, and regional information. Data preprocessing techniques such as encoding and normalization are applied to improve model performance. Various models including Convolutional Neural Networks (CNN), XGBoost, Random Forest, and Decision Tree classifiers are trained and evaluated to determine the most accurate approach for predicting student engagement levels.

The developed system is implemented as a web-based application using the Flask framework, allowing administrators to upload datasets, train predictive models, and compare their performance. Based on the trained model, the system can classify students into low, medium, or high-risk categories, enabling educators to identify struggling learners early and provide targeted academic support. By integrating hybrid deep learning and machine learning techniques, the proposed approach aims to enhance predictive accuracy and contribute to more effective student monitoring and personalized learning strategies in virtual education systems

II. LITERATURE SURVEY

1. Title: Predicting Student Performance in Online Learning Using Machine Learning

Author: K. Jayaprakash, E. Moody, E. Lauría, J. Regan, and J. Baron

Abstract:

This study explores the use of machine learning techniques to predict student performance in online learning environments. The authors analyze behavioral and engagement data obtained from learning management systems to identify patterns associated with student success or failure. Several predictive models are developed to evaluate the effectiveness of data-driven approaches in educational analytics. The research demonstrates that machine learning models can accurately identify students who are at risk of poor academic performance. The study highlights the importance of early intervention strategies to improve student retention and learning outcomes in digital education platforms.

2. Title: Educational Data Mining for Student Performance Prediction

Author: C. Romero and S. Ventura

Abstract:

Educational Data Mining (EDM) has emerged as an important research area for analyzing student learning behavior and improving educational outcomes. This study investigates the application of data mining techniques to predict student performance using historical academic records and interaction data from online learning systems. The authors evaluate several classification algorithms and demonstrate their effectiveness in identifying at-risk students. The results indicate that data mining approaches can significantly enhance decision-making processes in educational institutions and support personalized learning strategies.

3. Title: Early Prediction of Student Success Using Machine Learning Techniques

Author: M. Hussain, W. Zhu, W. Zhang, and S. M. Raza

Abstract:

This research focuses on early prediction of student success by applying machine learning techniques to analyze educational datasets. The study examines various behavioral and academic indicators such as attendance, participation, and assessment scores. Several classification models are implemented and compared to determine their predictive capabilities. The findings suggest that machine learning algorithms can effectively predict student outcomes and assist instructors in identifying students who require additional support. Early prediction helps improve learning engagement and academic performance.

III. EXISTING SYSTEM

In current virtual learning environments, student monitoring and performance evaluation are generally performed using traditional statistical analysis methods or simple rule-based systems. Educational platforms typically rely on basic indicators such as assignment submission status, quiz scores, and login records to assess student performance. These approaches often involve manual observation by instructors or simple analytical tools that provide limited insights into student learning behavior. Although some learning management systems provide analytics dashboards, they usually offer descriptive statistics rather than predictive insights. As a result, these systems are not capable of effectively identifying students who are at risk of disengagement or poor academic performance at an early stage. Furthermore, traditional methods often analyze limited features and fail to capture complex relationships among multiple behavioral factors such as study time, forum participation, video interaction, and login patterns. Therefore, existing systems lack the capability to provide accurate and timely predictions for identifying high-risk students in virtual learning environments.

IV. PROPOSED SYSTEM

The proposed system introduces a Hybrid Deep Learning Model to Predict High-Risk Students in Virtual Learning Environments by utilizing advanced machine learning and deep learning techniques to analyze student engagement data. The system collects and processes multiple behavioral features such as time spent weekly, quiz score average, forum participation, video watching percentage, assignments submitted, login frequency, session duration, device type, course difficulty, and region. Data preprocessing techniques including missing value removal, label encoding for categorical attributes, and feature scaling using a standard scaler are applied to improve data quality and model performance. The dataset is then divided into training and testing sets for model development. Several algorithms including Convolutional Neural Network (CNN), XGBoost, Random Forest, and Decision Tree are implemented to train predictive models and evaluate their performance. The system is developed as a web-based application using the Flask framework, where administrators can upload datasets, preprocess data, train models, and compare their accuracy through graphical visualization. Based on the trained model, the system predicts the engagement level of students and classifies them into low, medium, or high-risk categories, enabling educators to identify struggling learners early and provide timely academic support.

V. SYSTEM ARCHITECTURE

The system architecture presented in the diagram illustrates the workflow of the hybrid deep learning model designed to predict high-risk students in virtual learning environments. The architecture begins with the data collection stage, where student interaction data is gathered from online learning platforms. This data includes important behavioral features such as time spent on learning activities, quiz score averages, forum participation, video watching percentage, assignment submissions, login frequency, session duration, device type, course difficulty, and region. In addition to real-time interaction data, datasets uploaded by administrators are also used for model training and analysis.

After collecting the data, the system performs data preprocessing, which is a critical step for improving the quality of the dataset and preparing it for machine learning algorithms. In this stage, missing values are handled to ensure that incomplete records do not affect the prediction results. Categorical features such as device type, course difficulty, and region are converted into numerical values using label encoding so that machine learning models can process them effectively. Furthermore, feature scaling is applied to normalize numerical attributes, ensuring that all features contribute equally during model training and preventing bias caused by differences in data ranges.

Once the data is preprocessed, it is passed to the hybrid predictive modeling layer, where multiple machine learning and deep learning algorithms are applied. The architecture integrates several models including Convolutional Neural Network (CNN), XGBoost, Random Forest, and Decision Tree classifiers. These models analyze patterns in student behavior and engagement data to learn relationships between input features and student risk levels. The hybrid approach allows the system to leverage the strengths of different algorithms, improving the overall prediction performance and robustness of the system.

Following model training, the system performs model evaluation and comparison. In this stage, the trained models are tested using unseen data to measure their prediction accuracy. The performance of each algorithm is compared through evaluation metrics and graphical visualization, allowing administrators to identify the most accurate model for predicting student engagement risk. This comparison helps in selecting the best-performing model for deployment within the system.

The next stage is risk classification, where the selected predictive model analyzes new student data and determines their engagement level. Based on the prediction results, students are categorized into three groups: Low Risk, Medium Risk, and High Risk. This classification helps educators understand which students are actively engaged and which students may require additional academic support.Finally, the system generates early intervention alerts based on the predicted risk levels. If a student is identified as medium or high risk, the system can notify instructors or administrators so that timely intervention strategies can be implemented. These interventions may include personalized guidance, additional learning resources, or academic counseling..

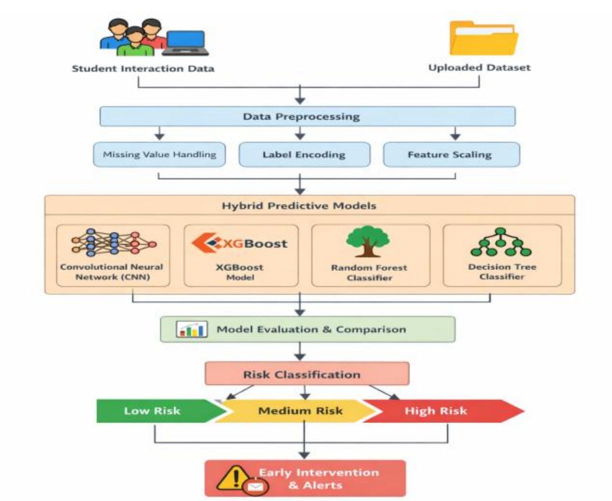

Fig 5.1: System Architecture

VI. IMPLEMENTATION

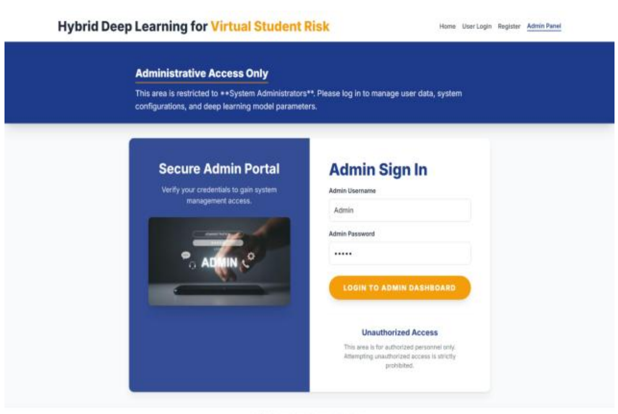

Fig 6.1: Home Page

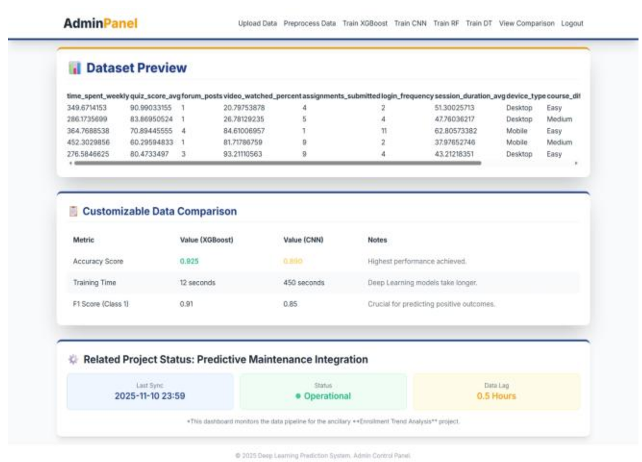

Fig 6.2: Dataset

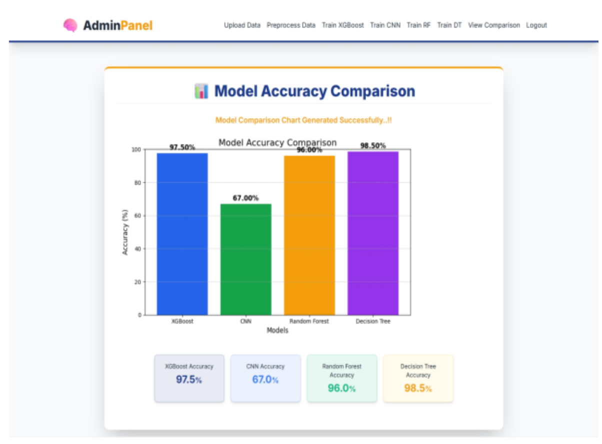

Fig 6.3: Algorithms

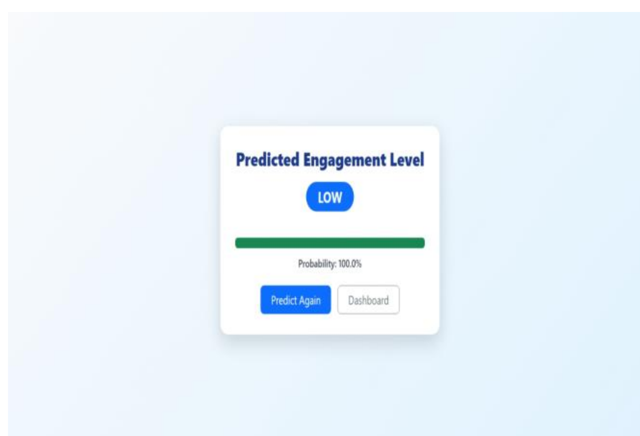

Fig 6.4: Final Output

VII. CONCLUSION

This project presented a Hybrid Deep Learning Model for Predicting High-Risk Students in Virtual Learning Environments by analyzing student engagement and behavioral data. The system utilizes various features such as time spent weekly, quiz score average, forum participation, video watching percentage, assignments submitted, login frequency, session duration, device type, course difficulty, and region to evaluate student learning patterns. Data preprocessing techniques including missing value removal, label encoding, and feature scaling were applied to ensure data quality and improve model performance.

Multiple machine learning and deep learning algorithms including Convolutional Neural Network (CNN), XGBoost, Random Forest, and Decision Tree were implemented and evaluated to determine the most effective model for predicting student engagement risk levels. The system compares the accuracy of these models and selects the best-performing model for predicting whether a student falls under low, medium, or high-risk categories. The developed web-based application allows administrators to upload datasets, train models, and visualize performance comparisons, while users can input student activity data to obtain risk predictions.

Overall, the proposed system demonstrates the effectiveness of hybrid machine learning and deep learning approaches in identifying at-risk students in virtual learning environments. By enabling early detection of disengaged learners, the system can help educators provide timely interventions, improve student engagement, and enhance academic performance in online education platforms.

VIII. FUTURE SCOPE

The proposed system can be further enhanced by incorporating several advanced features and technologies to improve its performance and applicability in modern educational environments. In the future, the system can be integrated with real-time data collection from Learning Management Systems (LMS) such as Moodle, Google Classroom, or other online learning platforms. This integration would enable continuous monitoring of student engagement and provide real-time predictions for identifying students who may be at risk.

The prediction accuracy of the system can also be improved by incorporating additional deep learning techniques such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), or Transformer-based models, which are effective in analyzing sequential and time-series educational data. These models can better capture the temporal patterns of student learning behavior and provide more accurate predictions.

Another possible improvement is the inclusion of more diverse features such as attendance records, assignment deadlines, learning pace, and emotional or sentiment analysis derived from student feedback. By considering these additional parameters, the system can generate more comprehensive insights into student performance and engagement levels.

The system can also be extended to include an automated recommendation system that suggests personalized learning resources, study plans, or instructor interventions for students identified as medium or high risk. This would help educators provide targeted support and improve the overall learning experience.

Furthermore, the application can be deployed as a cloud-based scalable platform capable of handling large educational datasets from multiple institutions. With enhanced security mechanisms and data privacy protection, the system could serve as an intelligent educational analytics tool for universities, online learning platforms, and educational organizations to improve student retention and academic success.

IX. REFERENCES

[1] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, Nov. 2010.

[2] R. S. J. d. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.

[3] K. Jayaprakash, E. Moody, E. Lauría, J. Regan, and J. Baron, "Early alert of academically at-risk students: An open source analytics initiative," Journal of Learning Analytics, vol. 1, no. 1, pp. 6–47, 2014.

[4] M. Hussain, W. Zhu, W. Zhang, S. M. Raza, and A. H. Khan, "Student engagement predictions in an e-learning system and their impact on student course assessment scores," Computational Intelligence and Neuroscience, vol. 2018, pp. 1–21, 2018.

[5] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," EDUCAUSE Review, vol. 46, no. 5, pp. 30–40, 2011.

[6] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004.

[7] A. A. Kardan, H. Sadeghi, S. S. Ghidary, and M. F. Sani, "Prediction of student course selection in online higher education institutes using neural network," Computers & Education, vol. 65, pp. 1–11, 2013.

[8] T. M. Thai-Nghe, L. Drumond, T. Horváth, A. Krohn-Grimberghe, and L. Schmidt-Thieme, "Factorization techniques for predicting student performance," in Proc. Educational Data Mining Conference, 2011, pp. 11–20.

[9] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," Telematics and Informatics, vol. 37, pp. 13–49, 2019.

[10] X. Xing, M. Guo, S. Petakovic, and G. Goggins, "Participation-based student final performance prediction model through interpretable machine learning," IEEE Access, vol. 8, pp. 137261–137272, 2020.