

Predicting High-Risk Students in Virtual Learning Environments

P.Kavya¹, B.Manideep², A.Anand³ and R.Mahesh⁴

UG Student, Dept. of IT, Malla Reddy Engineering College Secunderabad, Telangana, India.

Corresponding Author: hodaim1439@gmail.com

Abstract

Online learning has accelerated with the development of the Internet and communication technology. The widely accessible open online courses are delivered using digital environments that allow students to participate at speed and location. Virtual learning environments (VLEs) have developed quickly in recent years, giving students access to high-quality digital resources. Online learning environments have numerous benefits but drawbacks, including poor engagement, high dropout rates, low engagement, and self-regulated behavior, making students define their aims. Forecasting failed students in a VLE can help organizations and teachers improve their pedagogical practices and make data-driven decisions. This work proposes a Hybrid Deep Learning (HDL) approach to predict students' performance utilizing ECNN (Enhanced Convolution Neural Networks) Resnet model-based classification algorithms. The HDL approach is evaluated using the OULAD (Open University Learning Analytics Dataset), which provides a comprehensive and reliable assessment of the model's performance. The hybrid DLT approaches, demonstrating superiority, exhibited greater prediction accuracy than the existing classifiers.

Keywords: virtual learning environments (VLE), hybrid deep learning framework, enhanced convolutional neural network (ECNN).

I. INTRODUCTION

Student accomplishments are vital in higher education as they are quality measures of a university's academic success record. Many higher education institutions have established that high-quality education can change students' mental abilities, awareness, and knowledge levels. Teachers seek strategies to increase student accomplishments and enhance teaching process effectiveness continuously. Recent technological advancements and DMTs (Data Mining Techniques) allow instructors to examine and analyze online databases for patterns representing student behaviors and learning. Despite the importance of student performance to the learning process, it is a complicated phenomenon impacted by various elements, including the teaching environment and personal study habits. There are several definitions of student performance. And analyses of student successes in their co-curricular activities for learning evaluations. However, most studies assert graduation as a gauge of students' development.

Traditional DMTs' applications to address issues in education are called EDM (Educational Data Mining), and educational data includes student information, academic performances, and test scores. Reference class participation and student query frequencies are analyzed. Recently, EDM has become a powerful technique for forecasting academic successes, uncovering hidden patterns in educational data, and improving learning and teaching environments. EDM applications have given learning analytics a new perspective, including aspects of student data acquisition, understanding learning environments by inspections and analyses, and determining the ideal performances of students and teachers. Data about students and their environments, including institutions' employment of novel strategies, are gathered, assessed, and

reported using learning analytics to comprehend educational procedural changes better. Some research studies have employed educational data mining to forecast academic achievement; nevertheless, most of these studies have concentrated on traditional high schools and universities with little attention to college education. However, the CGPA (cumulative grade point average) is the primary emphasis of the dataset utilized in earlier studies, along with additional demographic information, ultimately making CGPA feature with maximum information gains.

GPA (Grade Point Aggregates) of the courses that students took must be considered in the study since the CGPA combines elements of numerous GPAs across semesters and levels. This work intends to uncover markers for high performances amongst teachers-in-training and predict future academic successes using academic datasets. Hence, further studies are required to improve elementary education, which helps to stop bad trends. Since most teachers at primary schools are trained exclusively in colleges of education, which primarily produce instructors meant for the education industry, therefore may need help understanding the behavior of students and the exact reasons for failing their courses. Using DLT (deep learning approaches-based mining, focused studies enable the management and evaluation of educational data obtained from multiple sources.

II. LITERATURE REVIEW

A. Predicting youth at high risk of aging out of foster care using machine learning methods

Youth who age out, or emancipate, from foster care without permanency by age 18 are at increased risk of experiencing difficulties during their transition to adulthood. Although general support programs have been implemented, it is unknown whether an earlier, more targeted and proactive identification of youth with a heightened risk of exiting care

without permanency would improve both the services provided and the outcomes achieved. In the current study, we explored whether algorithmic approaches, using machine learning methods applied to historical child protection and welfare records, might assist child protection agencies to better identify and serve youth at high risk of not having established legal permanency by age 18. Importantly, we also describe how metrics of fairness and bias from the computer science literature relate to this specific use case.

B. Innovative approaches to identifying and responding to the needs of high-risk youth

Following a critical review of key issues facing the delivery of effective, cost-attractive services to high-risk youth, and research addressing these experiences, we identify some innovative approaches to identify and respond to the multiple needs of these youth. The importance of providing family services with an ecological focus is stressed. Further, some exciting developments occurring in juvenile assessment centers, involving screening and in-depth assessment, as well as intervention strategies are presented. These innovative developments include for the Tampa Juvenile Assessment Center: 1) a family empowerment intervention service for arrested youth; 2) a family-focused early intervention, intensive case management service for youth entering a diversion program; and 3) for the Miami-Dade Juvenile Assessment Center, the comprehensive program of research and program development occurring in the context of the National Demonstration Project. We conclude with a discussion of major issues facing the field and the continuing need for a national commitment to help the many troubled youths entering the juvenile justice system.

C. Exploring online activities to predict the final grade of student

Student success rate is a significant indicator of the quality of the educational services offered at higher education institutions (HEIs). It allows students to make their plans to achieve the set goals and helps teachers to identify the at-risk students and make timely interventions. University decision-makers need reliable data on student success rates to formulate specific and coherent decisions to improve students' academic performance. In recent years, EDM has become an effective tool for exploring data from student activities to predict their final grades. This study presents a case study for predicting the students' final grades based on their activities in Moodle Learning Management System (LMS) and attendance in online lectures conducted via Zoom by applying statistical and machine learning techniques. The data set consists of the final grades for 105 students who study Object-Oriented Programming at the University of Plovdiv during the 2021–2022 year, data for their activities in the online course (7057 records), and attendance to lectures (738). The predictions are based on 46 attributes. The Chi-square test is utilized to assess the association between students' final grades and event context (lectures, source code, exercise, and assignment) and the relationships between attendance at lectures and final results. The logistic regression model is utilized to assess the actual impact of event context on "Fail" students in a multivariate setup. Four machine learning algorithms (Random Forest, XGBoost, KNN, and SVM) are applied using 70% of training data and 30% of test data to predict the

students' final grades. Five-fold cross validation was also utilized.

III. PROPOSED METODOLOGY

The overview of our proposed system is shown in the below figure.

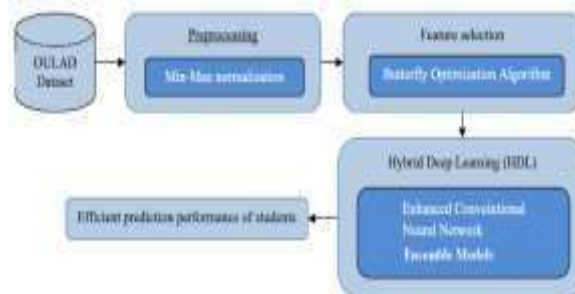


Fig. 1: System Overview

Implementation Modules

a. Service Provider Module

In this module, service provider login to the system using valid username and password. After login successful, he can perform the following operations like train and test dataset, view trained and tested accuracy and view remote users.

b. Train and Test Model

In this module, the service provider split the Used dataset into train and test data of ratio 70 % and 30 % respectively. The 70% of the data is consider as train data which is used to train the model and 30% of the data is consider as test which is used to test the model

c. Remote User

In this module, the remote user register to the system, and login to the system valid username, and password. After login successful, he can perform view profile, predicting risk student status.

d. Graphical Analysis

In this module, display the graphs like accuracy and predicted ratio of the system. Various factors take into consideration for the graph analysis. In this phase plot the charts like bar chart and so others.

This study proposes a Hybrid Deep Learning (HDL) framework for predicting high-risk students in Virtual Learning Environments (VLEs). The methodology consists of multiple stages, including data collection, preprocessing, feature selection, model development, training, and evaluation. The dataset used in this research is the Open University Learning Analytics Dataset (OULAD), which contains comprehensive information about student demographics, assessment scores, and their interactions within virtual learning platforms. This dataset provides a reliable foundation for analyzing student behavior and predicting academic performance.

In the preprocessing stage, the dataset is cleaned by handling missing and inconsistent values. To ensure uniformity in feature representation, Min-Max normalization is applied to scale all input features within a fixed range. This step improves the efficiency and convergence of the deep learning model. Feature selection is carried out using the Butterfly Optimization Algorithm (BOA), which identifies the most relevant features contributing to student performance prediction. This process reduces dimensionality and enhances model accuracy by eliminating redundant and irrelevant attributes.

The core of the proposed system is a Hybrid Deep Learning model that combines Enhanced Convolutional Neural Networks (ECNN) and Residual Networks (ResNet). The ECNN component is responsible for extracting meaningful features from the input data, while the ResNet architecture facilitates deep feature learning and prevents vanishing gradient problems through skip connections. This hybrid architecture enables the model to capture both shallow and deep patterns in the dataset. The dataset is divided into training and testing sets in a ratio of 70% and 30%, respectively. The training data is used to build the model, while the testing data is used to evaluate its performance. The effectiveness of the model is assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score.

Additionally, the system includes modules for service providers and users. The service provider module handles model training, testing, and performance visualization, while the user module allows prediction of student risk status. Graphical analysis is also incorporated to visualize model accuracy and prediction outcomes.

V. RESULTS AND DISCUSSIONS

The proposed Hybrid Deep Learning model was evaluated using the OULAD dataset, and its performance was compared with existing models. The experimental results demonstrate that the HDL model achieved an accuracy of 95.67%, which is significantly higher than traditional models such as Deep Feed Forward Neural Networks (DFNN) with 93.9% accuracy and Multi-Layer Perceptron (MLP) with 71.41% accuracy.

The improved performance of the proposed model can be attributed to the effective combination of ECNN and ResNet architectures. The ECNN enhances feature extraction capabilities, while the ResNet ensures efficient deep learning by mitigating issues such as vanishing gradients. Furthermore, the use of Butterfly Optimization for feature selection plays a crucial role in improving prediction accuracy by focusing on the most relevant attributes. The results indicate that the HDL model can effectively identify students at risk of poor academic performance in virtual learning environments. This capability is particularly important in modern education systems, where early intervention can significantly improve student outcomes. The graphical analysis of the model further supports its effectiveness, showing consistent accuracy and reliable prediction ratios across different test cases. Overall, the results confirm that the proposed model outperforms existing approaches in terms of accuracy and robustness.

IV. CONCLUSION

In this study, a Hybrid Deep Learning framework was proposed to predict high-risk students in Virtual Learning Environments. The integration of Enhanced Convolutional Neural Networks and ResNet architectures, along with optimized feature selection using the Butterfly Optimization Algorithm, resulted in a highly accurate prediction model. The experimental evaluation demonstrates that the proposed model achieves superior performance compared to existing methods. By accurately identifying students at risk, the system enables educators and institutions to take timely and effective measures such as personalized guidance, counseling, and academic support. Thus, the proposed HDL model serves as an efficient and reliable solution for improving student performance and reducing dropout rates in online learning environments. Although the proposed model shows promising results, there are several directions for future research. The model can be enhanced by incorporating real-time student activity data to enable dynamic and continuous prediction. Additionally, integrating behavioral and psychological factors could further improve prediction accuracy.

Future work can also focus on applying Explainable Artificial Intelligence (XAI) techniques to make the model's predictions more interpretable and transparent for educators. The system can be extended to integrate with Learning Management Systems (LMS)

such as Moodle for real-time monitoring and intervention. Moreover, the use of larger and more diverse datasets can improve the generalization capability of the model. Advanced deep learning techniques, such as transformer-based models and graph neural networks, can also be explored to further enhance prediction performance. Finally, developing a user-friendly web or mobile-based application for real-time student risk analysis can make the system more practical and accessible for educational institutions.

REFERENCES

- [1]. E. Ahn, Y. Gil, and E. Putnam-Hornstein, "Predicting youth at high risk of aging out of foster care using machine learning methods," *Child Abuse Neglect*, vol. 117, Jul. 2021, Art. no. 105059, doi: 10.1016/j.chiabu.2021.105059.
- [2]. R. Dembo and W. Walters, "Innovative approaches to identifying and responding to the needs of high risk ."
- [3]. M. A. Shaik, A. Fatima, M. Parveen, A. Soumya Rani, A. Mohammad and A. Rahim, "Dual-Model Approach for Lung Disease Classification Using Convolutional Neural Networks and Support Vector Machines," 2024
- [4]. A. Marwaha and A. Singla, "A study of factors to predict at-risk students based on machine learning techniques," in *Advances in Intelligent Systems and Computing*. Singapore: Springer, 2020, pp. 133–141.
- [5]. O. Embarak, "Apply machine learning algorithms to predict at-risk students to admission period," in *Proc. 7th Int. Conf. Inf. Technol. Trends (ITT)*.