Predicting at-Risk Students in Virtual Learning Environments Using A Hybrid Deep Learning Model

[1]Ms. CH. Vasavi, [2]Ms. N. Siva Nagamani,
[1]Assistant Professor, [2]Associate Professor, Department of Computer Science & Engineering,
Geethanjali Institute Of Science And Technology, Nellore, India

Abstract
With the rise of digital education platforms, virtual learning environments (VLEs) have become increasingly popular, offering flexible access to educational content. Despite their advantages, VLEs often suffer from issues such as low engagement, high dropout rates, and a lack of self-regulated learning among students. Early prediction of high-risk students in these environments is essential for enabling timely interventions and enhancing learning outcomes. This study presents a Hybrid Deep Learning (HDL) framework that utilizes an Enhanced Convolutional Neural Network (ECNN) integrated with a ResNet-based classification model to predict student performance in VLEs. The proposed model is evaluated using the Open University Learning Analytics Dataset (OULAD), which provides a robust benchmark for academic performance prediction. Our HDL approach outperforms traditional models, achieving a prediction accuracy of 95.67%, compared to 93.9% for Deep Feedforward Neural Networks (DFFNN) and 71.41% for Multi-Layer Perceptrons (MLP). These results demonstrate the model's effectiveness in identifying at-risk students and supporting data-driven pedagogical decision-making.
Keywords: Academic Performance Prediction, Virtual Learning Environments (VLE), Hybrid Deep Learning, Enhanced Convolutional Neural Network (ECNN), ResNet, Min-Max Normalization, Butterfly Optimization, Learning Analytics

Introduction

The project presents a Hybrid Deep Learning (HDL) model designed to predict high-risk students in virtual learning environments (VLEs). With the rapid growth of online education, institutions face challenges such as low engagement and high dropout rates. The HDL model aims to address these issues by utilizing Enhanced Convolutional Neural Networks (ECNN) and ResNet classification algorithms to enhance prediction accuracy. The research leverages the Open University Learning Analytics Dataset (OULAD), which provides a comprehensive dataset for evaluating student performance . The HDL approach incorporates advanced preprocessing techniques, including min-max normalization and Butterfly optimization-based feature selection. The results indicate that the HDL model achieves an accuracy rate of 95.67%, significantly outperforming existing models such as the Deep Feedforward Neural Network (DFFNN) and Multi-Layer Perceptron (MLP), which recorded accuracies of 93.9% and 71.41%, respectively . Furthermore, the project emphasizes the importance of early identification of at-risk students, allowing educational institutions to implement timely interventions such as additional coaching and counseling. This proactive approach can lead to improved academic outcomes and better retention rates in VLEs . The findings suggest that integrating advanced predictive models like HDL can transform educational practices, enabling educators to make data-driven decisions that enhance student support and engagement

Motivation

The motivation for developing a Hybrid Deep Learning (HDL) model to predict high-risk students in virtual learning environments (VLEs) stems from the increasing reliance on online education and the challenges it presents. As educational institutions shift towards digital platforms, they encounter issues such as low student engagement, high dropout rates, and the need for self-regulated learning. These challenges highlight the necessity for effective predictive models that can identify students at risk of underperformance early on, allowing for timely interventions to enhance their academic success .

Objective

The primary objective of this project is to develop a Hybrid Deep Learning (HDL) model that effectively predicts high-risk students in virtual learning environments (VLEs). This objective is driven by the need to address the challenges faced in online education, such as low engagement and high dropout rates. By utilizing advanced machine learning techniques, specifically Enhanced Convolutional Neural Networks (ECNN) and ResNet models, the project aims to create a predictive tool that can identify students who may struggle academically, allowing educators to intervene proactively and support these students in their learning journey. Another key objective is to enhance the accuracy of student performance predictions compared to existing models. The HDL approach is designed to leverage the strengths of various classification algorithms, aiming for a prediction accuracy that surpasses traditional methods. The research indicates that the HDL model achieves an impressive accuracy of 95.67%, which is significantly higher than other models like the DFFNN and MLP. This improvement in predictive accuracy is crucial for educational institutions seeking to implement data-driven strategies to enhance student outcomes and optimize teaching practices .

Lastly, the project seeks to contribute to the broader field of educational analytics by employing innovative feature selection techniques, such as the Butterfly Optimization Algorithm (BOA). This objective focuses on identifying the most relevant features from the dataset related to student performance, ensuring that the model is not only accurate but also interpretable. By integrating BOA into the feature selection process, the project aims to refine the model further, making it a valuable resource for educators and administrators in understanding and addressing the factors that influence student success in VLEs .

Literature Review

The literature on predicting student performance in virtual learning environments (VLEs) has evolved significantly, focusing on various machine learning techniques and deep learning models to enhance educational outcomes. Hybrid Methods for Student Performance Prediction “Ayienda et al” This study combines multiple machine learning techniques (MLTs) and deep learning techniques (DLTs) to create a hybrid method for predicting student performance. The authors utilized a weighted voting classifier along with classifiers like SVM, MLPs, LRs, KNNs, and NBs, achieving a notable accuracy of 97.6% in their predictions . Neural Networks in Academic Predictions “Kalyani et al” This research highlights the effectiveness of neural networks in predicting student success. It emphasizes the importance of various factors, such as study hours and academic engagement, in determining student performance, suggesting that these variables are critical for accurate predictions .

Sequence-Based Performance Classifier “ Wang et al” Wang et al. introduced a two-stage classification system called the Sequence-based Performance Classifier (SPC), which integrates sequence encoders with traditional data mining classifiers. Their approach demonstrated improved recall and accuracy in predicting student performance by focusing on sequential features of student behavior . **Ensemble Classifier-Based Models “Begum and Padmannavar”** This study proposed an ensemble classifier-based model for predicting student performance. They emphasized data preprocessing to eliminate redundancies and enhance the correlation between features, utilizing techniques like Boosting, Bagging, and Random subspace classifiers to improve prediction accuracy . **Deep Neural Networks for**

Performance Assessment "Neha et al" Neha et al. developed a model using Deep Neural Networks (DNNs) to evaluate predictive variables for student performance. Their findings indicated that their model significantly outperformed traditional models, showcasing the potential of DNNs in educational assessments. **Genetic Algorithms in Prediction Models "Damuluri et al"** This research explored the application of Genetic Algorithms (GA) to enhance the accuracy of prediction models. By optimizing classifiers, they demonstrated a significant improvement in predicting students' final scores and identifying those at risk of failing .

Behavioral Characteristics and Academic Success "Amrieh et al" Amrieh et al. focused on the behavioral characteristics of students in relation to their interactions with e-learning management systems. Their study found substantial links between student behaviors and academic performance, suggesting that understanding these behaviors can lead to better predictive models . In summary, the literature reveals a diverse range of approaches and methodologies aimed at predicting student performance in VLEs, highlighting the importance of integrating various techniques to enhance accuracy and effectiveness in educational settings.

Proposed Model

The proposed model is designed to predict high-risk students in virtual learning environments using a Hybrid Deep Learning (HDL) approach. It utilizes the Open University Learning Analytics Dataset (OULAD), which contains detailed student demographic and interaction data. The process begins with data preprocessing using Min-Max normalization to standardize feature ranges. To enhance model performance, the Butterfly Optimization Algorithm (BOA) is employed for selecting the most relevant features. The core of the system is a hybrid architecture that integrates a Convolutional Neural Network (CNN), an Artificial Neural Network (ANN), and ResNet V2. The CNN is used to extract spatial features from student activity patterns, while the ANN captures complex, nonlinear relationships within the data. ResNet V2, a deep residual network, adds depth and robustness to the learning process. This combination allows the system to effectively learn from both structured and unstructured data.

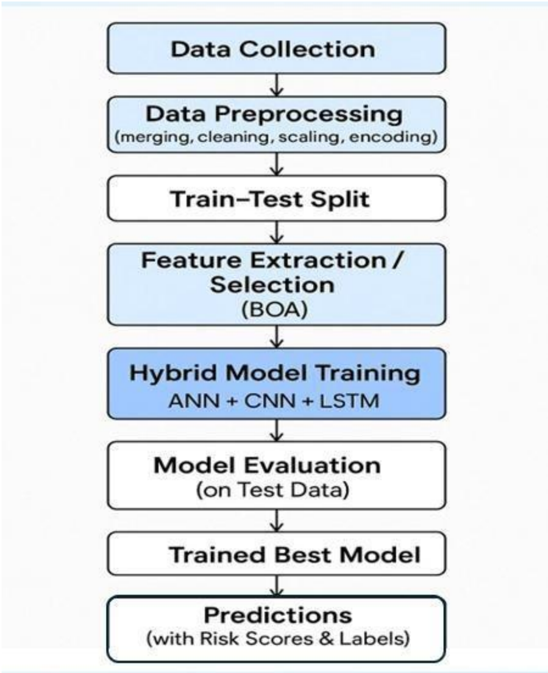

Fig.1. The Process of Student Performance Prediction Using HDL

ADVANTAGES OF PROPOSED SYSTEM • Enhanced Prediction Accuracy • Incorporation of Advanced Deep Learning Methods • Optimization of feature selection with Butterfly Optimization • Efficient Data preprocessing with Min-Max Normalization

The proposed system aims to predict high-risk students in virtual learning environments using a Hybrid Deep Learning (HDL) model. It uses the OULAD dataset, which includes student demographics, assessments, and activity data. The data is preprocessed with Min-Max normalization, and relevant features are selected using the Butterfly Optimization Algorithm (BOA). The core model combines CNN, ANN, and ResNet V2 to effectively learn from both structured and behavioral data. This hybrid approach improves prediction accuracy and enables early identification of at-risk students, supporting timely educational interventions.

OULAD Dataset: The **Open University Learning Analytics Dataset (OULAD)** is a publicly available dataset provided by The Open University (UK) to support research in educational data mining and learning analytics. It contains detailed information about **32,593 students** enrolled in various online courses. The dataset includes multiple files with information such as: • **Student demographics** (age, gender, education, disability status)

Course details (module codes, presentation dates) • **Assessment records** (scores, dates, types)

VLE interaction data (click counts, resource types, activity dates) **Enrollment and final results** OULAD is ideal for building predictive models like the one in your project, as it offers both **behavioral and performance-related data**, enabling accurate identification of at-risk students.

Preprocessing: Preprocessing is a crucial step in preparing the OULAD dataset for model training. It involves cleaning, transforming, and organizing the raw data to improve model accuracy and efficiency. The following steps are performed:

Handling Missing Values: Records with missing or incomplete data are either filled or removed to ensure consistency.

Encoding Categorical Data: Non-numeric features such as gender, education level, and course codes are converted into numerical form using encoding techniques like one-hot or label encoding.

Merging Tables: Relevant tables from OULAD (e.g., studentInfo, studentAssessment, studentVLE) are joined to create a comprehensive dataset for each student.

Feature Engineering: New features may be derived from existing ones (e.g., total clicks, average score) to improve model input quality. These steps ensure the dataset is clean, consistent, and ready for the hybrid deep learning model to process effectively.

Min-Max normalization: Min-Max Normalization is a data scaling technique used to transform features to a fixed range, typically between **0 and 1**. It ensures that all input features contribute equally to the learning process, especially important for deep learning models.

Formula: $X_{normalized} = (X - X_{min}) / (X_{max} - X_{min}))$ • **X**: Original value

X_min: Minimum value in the feature

X_max: Maximum value in the feature

X_normalized: Scaled value between 0 and 1

Feature selection: Feature selection is the process of identifying and selecting the most relevant input features from a dataset that contribute significantly to the prediction outcome. In this project, it is a critical step to improve the performance and efficiency of the Hybrid Deep Learning (HDL) model.

Butterfly Optimization Algorithm (BOA): The **Butterfly Optimization Algorithm (BOA)** is a **nature-inspired metaheuristic** algorithm developed by mimicking the foraging and mating behavior of butterflies. It is used in this project for **feature selection**, helping identify the most relevant attributes from the dataset to improve model performance.

Working of Butterfly Optimization Algorithm in Hybrid Deep Learning Model In the proposed system, the **Butterfly Optimization Algorithm (BOA)** plays a crucial role in the **feature selection**

phase before feeding data into the **Hybrid Deep Learning (HDL) model**, which combines **CNN, ANN, and ResNet V2**. BOA ensures that only the most relevant features from the OULAD dataset are used, improving both performance and efficiency.

SYSTEM IMPLEMENTATION

SYSTEM MODULES

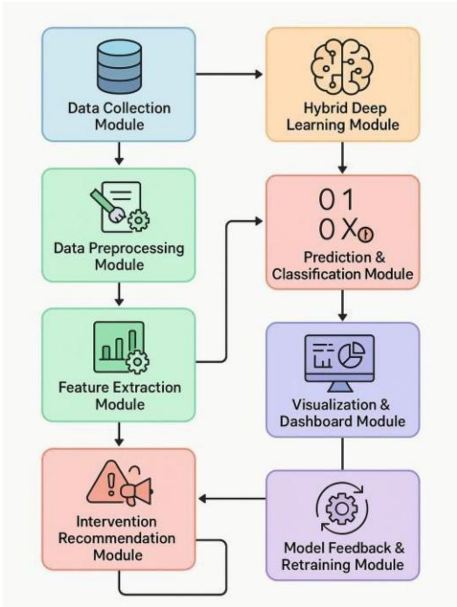

Fig.2. System Modules

OULAD Dataset: • Full Form: Open University Learning Analytics Dataset. Purpose: Serves as the input data for the system. Contents: ✓ Student demographics (age, gender, education level) ✓ Course information ✓ click stream data ✓ Assessment results and performance logs This real world data set enables the model to learn from genuine student behaviour in online courses.

Results & Analysis

The execution of the project will be explained clearly with the help of continuous screen shots

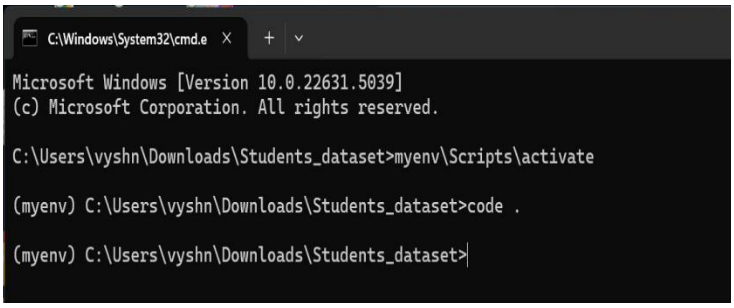

Fig.3. Commands to activate environment

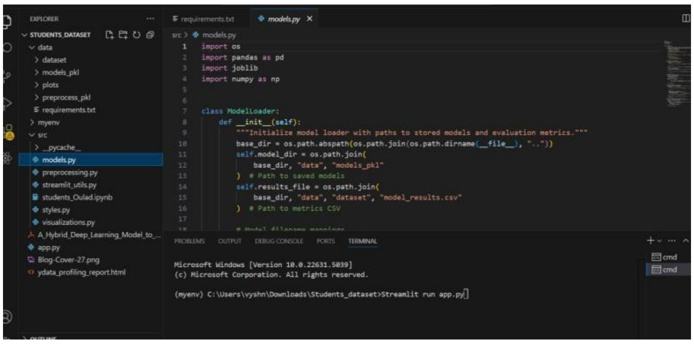

Fig.4. Command to run Application

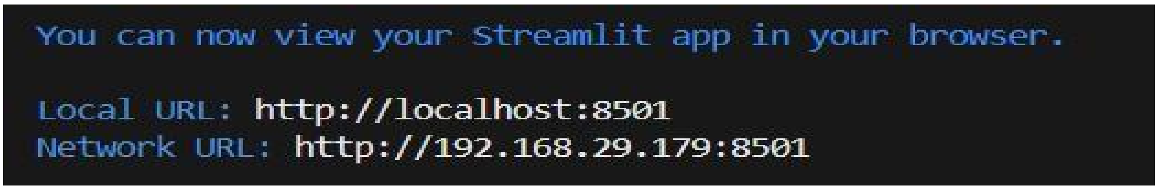

Fig.5. After execute the main file, it gives the Local URL (website)

Fig.6. Problem statement

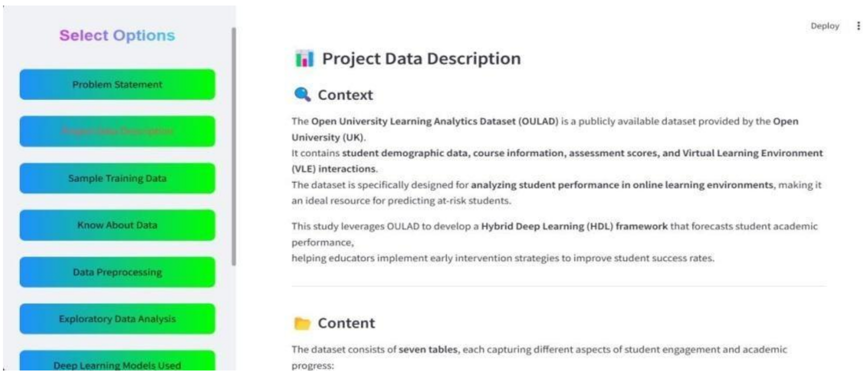

Fig.7. Project Data Description

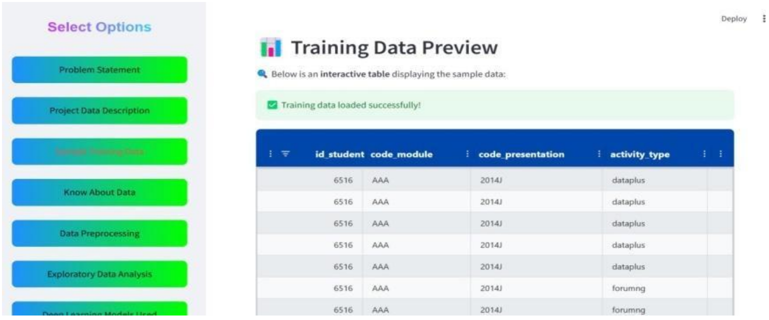

Fig.8. Sample Training Data

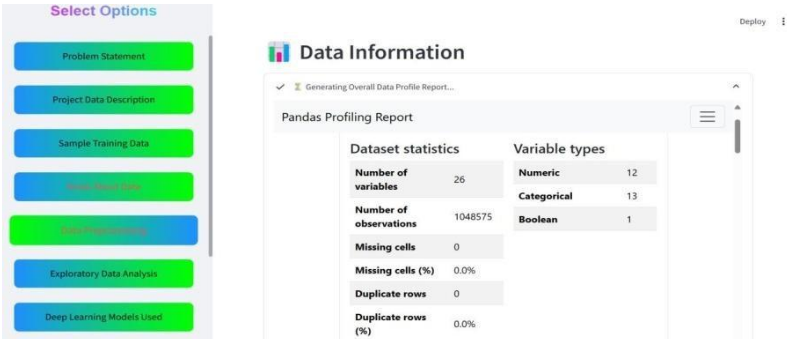

Fig.9. Information about data

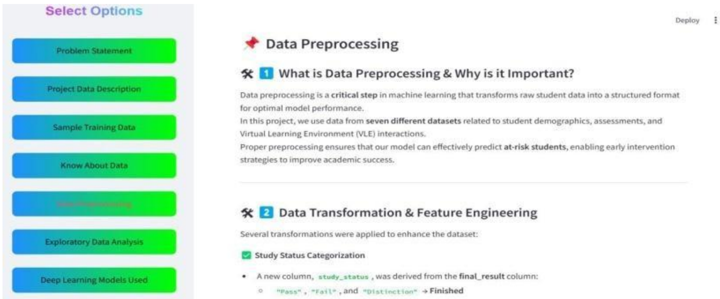

Fig.10. Methodologies used

Fig.11. Data Visualization using Heat Map

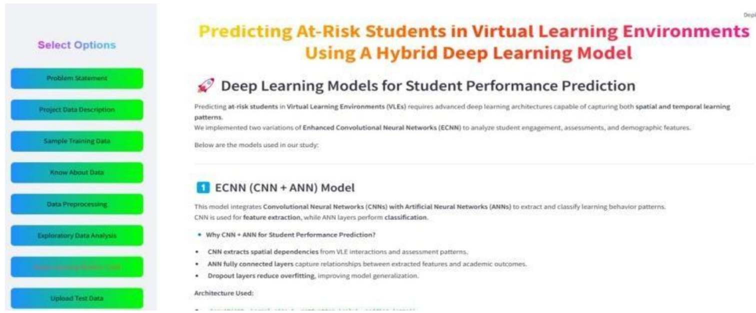

Fig.12. Deep Learning Models used

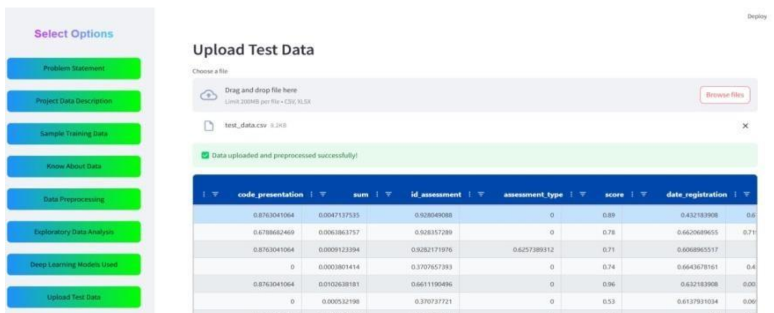

Fig.13. Uploading dataset

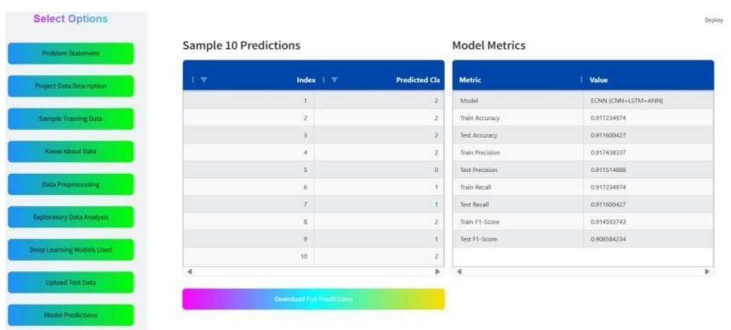

Fig.14. Model Predictions

Conclusion

Academic achievement of students at any professional institution has emerged as management's main concern. Early identification of pupils at risk for underperformance enables management to move quickly to boost those students' performance through additional coaching and counseling. To find the best-performing predictive model, this project proposes an HDL framework to forecast students' performance utilizing ECNN and Resnet model-based classification algorithms. Eventually, the HDL is created to effectively forecast high-risk pupils in a model based on a VLE. The OULAD is used to assess the proposed model, and the effectiveness of various classifiers is included in this project for comparative evaluations using the metrics of precision, recall, and F-score values. From the experimental results, the proposed HDL model has an accuracy rate of 95.67%, which is higher when compared with the other existing works like the DFFNN model (93.9%) and the MLP model (71.41%). Statistics demonstrate that the proposed FDL methodology produces better accuracy than current classification methods.

Future Scope

This project lays the groundwork for enhancing predictive models aimed at identifying high-risk students in virtual learning environments. Future research can focus on incorporating a broader range of data, including emotional, behavioral, and socio-economic indicators, to improve model accuracy and comprehensiveness. The use of real-time analytics can also be explored, enabling timely interventions by educators. Moreover, integrating explainable AI techniques would make the model's predictions more transparent and trustworthy, allowing instructors to understand the reasoning behind each risk classification. Expanding the system for scalability across different educational platforms and regions, along with optimizing it for real-time deployment, would make the solution more practical and impactful in real-world academic settings. These advancements would contribute significantly to creating a more proactive and supportive virtual learning environment

References

1. R. Ayienda, R. Rimiru, and W. Cheruiyot, ''Predicting students academic performance using a hybrid of machine learning algorithms,'' in Proc. IEEE AFRICON, Sep. 2021, pp. 1–6, doi: 10.1109/ AFRICON51333.2021.9571012.
2. B. S. Kalyani, D. Harisha, V. RamyaKrishna, and S. Manne, ''Evaluation of students performance using neural networks,'' in Advances in Intelligent Systems and Computing. Cham, Switzerland: Springer, 2020, pp. 499–505.
3. N. Naicker, T. Adeliyi, and J. Wing, ''Linear support vector machines for prediction of student performance in school-based education,'' Math. Problems Eng., vol. 2020, pp. 1–7, Oct. 2020, doi: 10.1155/2020/4761468.
4. X. Wang, X. Yu, L. Guo, F. Liu, and L. Xu, ''Student performance prediction with short-term sequential campus behaviors,'' Information, vol. 11, no. 4, p. 201, Apr. 2020, doi: 10.3390/info11040201.
5. K. Neha, J. Sidiq, and M. Zaman, ''Deep neural network model for identification of predictive variables and evaluation of student's academic performance,'' Revue d'Intelligence Artificielle, vol. 35, no. 5, pp. 409–415, Oct. 2021, doi: 10.18280/ria.350507.
6. S. Begum and S. S. Padmannavar, ''Genetically optimized ensemble classifiers for multiclass student performance prediction,'' Int. J. Intell. Eng. Syst., vol. 15, no. 2, pp. 316–328, Apr. 2022, doi: 10.22266/ijies2022.0430.29.
7. 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.
8. M. Yağcı, ''Educational data mining: Prediction of students' academic performance using machine learning algorithms,'' Smart Learn. Environ., vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
9. T. O. Olaleye and O. R. Vincent, ''A predictive model for students' performance and risk level indicators using machine learning,'' in Proc. Int. Conf. Math., Comput. Eng. Comput. Sci. (ICMCECS), Mar. 2020, pp. 1–7, doi: 10.1109/ICMCECS47690.2020.240897899X/1051/1/012005.
10. L. Ismail, H. Materwala, and A. Hennebelle, ''Comparative analysis of machine learning models for students' performance prediction,'' in Advances in Digital Science. Cham, Switzerland: Springer, 2021, pp. 149–160.

11. S. Gaftandzhieva, A. Talukder, N. Gohain, S. Hussain, P. Theodorou, Y. K. Salal, and R. Doneva, ‘‘Exploring online activities to predict the final grade of student,’’ Mathematics, vol. 10, no. 20, p. 3758, Oct. 2022, doi: 10.3390/math10203758.
12. A. Siddique, A. Jan, F. Majeed, A. I. Qahmash, N. N. Quadri, and M. O. A. Wahab, ‘‘Predicting academic performance using an efficient model based on fusion of classifiers,’’ Appl. Sci., vol. 11, no. 24, p. 11845, Dec. 2021, doi: 10.3390/app112411845.
13. M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, ‘‘Predicting at-risk students at different percentages of course length for early intervention using machine learning models,’’ IEEE Access, vol. 9, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.