RESEARCH PAPER

Intelligent Academic Performance Forecasting using Integrated Machine Learning Models

Ramadevi M [1*], Anton Brail S [2*], Arun S [3*] , Dinesh S [4*]

[1] Assistant Professor, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.*
[2] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India*
[3] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.*
[4] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.*

***Abstract*—** In the era of data-driven education, unlocking hidden insights from student learning patterns holds the key to shaping the future of academic success. Accurate prediction of student performance is vital for improving educational standards and enabling proactive support for students at risk. Conventional prediction techniques frequently depend on single algorithms, which can be inadequate when handling the complexity and multidimensionality of educational datasets. This study introduces a novel hybrid machine learning framework that combines Random Forest (RF) and XGBoost to enhance both prediction accuracy and model reliability. By merging the parallel bagging capability of RF with the sequential boosting proficiency of XGBoost, the framework addresses typical shortcomings of opaque models and minimizes variance in individual predictions. The evaluation leverages established benchmark datasets, namely the Open University Learning Analytics Dataset (OULAD) and the Chinese University Dataset (CUD), to assess performance across a broad array of variables, such as demographic data, online learning activities, and previous academic achievements. This methodology directly tackles crucial technical obstacles, including class imbalance and feature significance, which commonly impede effective outcome prediction. The integration of these ensemble methods is intended to equip educators with a dependable early warning system, supporting tailored instructional materials and efficient resource distribution. Additionally, the framework's effectiveness is assessed using various metrics Accuracy, Precision, Recall, and F-Measure to provide a thorough evaluation of its predictive capabilities. This research advances the domain of educational data mining by highlighting the enhanced effectiveness of integrated models in early identification of pivotal student outcomes.

Keywords— Educational Data Mining, Hybrid Machine Learning, Random Forest, XGBoost, Student Performance Prediction, Ensemble Methods.

How to cite this article: Ramadevi M, Anton Brail S, Arun S, Dinesh S. Intelligent academic performance forecasting using integrated machine learning models. Int J Drug Deliv Technol. 2026;16(74s): 2037-2042. DOI: 10.25258/ijddt.16.74s.163

I. INTRODUCTION

In the contemporary educational landscape, the ability to accurately forecast student academic performance has emerged as a cornerstone for elevating institutional outcomes and shaping the future of academic success. Through precise prediction, educators can pinpoint students facing significant challenges and provide customized materials suited to their specific learning needs [1], [2]. However, traditional reliance on instructor intuition lacks the scalability and efficiency required to manage modern, diverse student populations. Consequently, the past decade has seen a surge in employing machine learning and data mining techniques to forecast outcomes, particularly through dimensionality reduction to handle increasingly complex datasets [3].

Current research highlights that student success is heavily influenced by a multifaceted combination of academic history, demographic factors, and spatial behaviors [4], [15]. While various algorithms have been explored, the academic community continues to investigate the most effective models for high-dimensional educational data [1]. Recent comparative studies have scrutinized the efficacy of tree-based models, identifying XGBoost and Random Forest as high-performing benchmarks for tabular educational data [6], [17]. Despite their individual success, standalone algorithms often struggle with class imbalance, distribution shifts, and the inherent high dimensionality of student datasets [13], [18]. To mitigate these limitations, recent literature has shifted toward hybrid deep learning and system dynamics to model complex patterns such as student dropout behavior and learning trajectories [7], [10].

Furthermore, there is an evolving need for frameworks that integrate real-time behavior analysis with high-speed predictive power to enhance teaching effectiveness [12]. Advanced machine learning approaches originally used for system optimization and risk prediction such as feature-scaled ensemble learning [5], [14], hybrid neural networks [8], and tuned hierarchical techniques [11] demonstrate the robustness required for complex forecasting. The necessity for advanced preprocessing, such as utilizing ADASYN to enhance classifier performance on imbalanced data, has also become a critical focus [9].

To address these challenges, this research proposes an Intelligent Academic Performance Forecasting System using Integrated Machine Learning Models. By developing a weighted ensemble approach [19] that combines the parallel bagging capability of Random Forest with the sequential boosting proficiency of XGBoost, we leverage the collective strengths of both methodologies. Our framework utilizes integrated feature selection [18] and ensemble techniques [16] to ensure accuracy across heterogeneous student populations. By applying a stacked generalization of tree-based classifiers [20], this study provides a robust early warning mechanism for identifying at-risk students, effectively bridging the gap between raw educational data and actionable institutional insights.

II. RELATED WORKS

The landscape of educational data mining has shifted toward sophisticated algorithmic approaches to ensure student success. Pan et al. [1] provide a foundational state-of-the-art survey on academic performance prediction (APP), emphasizing that while a marked surge in machine learning exists, the academic community has yet to agree on a universal algorithm. This gap highlights the necessity for ongoing analysis and comparison of existing models to provide tailored recommendations for diverse educational requirements. Manivannan et al. [2] further explore this by utilizing AI to deliver personalized learning experiences on mobile platforms, demonstrating the need for scalable and accessible intelligent systems.

Understanding input variables is critical for model precision. Algarni et al. [15] identify at-risk students by integrating academic and demographic data, while Musaddiq et al. [4] expand this scope by analyzing the impact of spatial behaviors on performance. These studies collectively suggest that predictive accuracy relies on a multifaceted view of the student. However, high-dimensional datasets present computational challenges. Alhazmi and Sheneamer [3] address this by demonstrating how dimensionality reduction techniques can streamline early prediction without losing vital information.

Technical challenges such as data imbalance and distribution shifts often hinder standalone model performance. Osman et al. [9] explore the use of ADASYN and Multi-Layer Perceptrons to enhance classification in imbalanced datasets, ensuring that minority classes are not overlooked. Furthermore, Ren and Yu [13] introduce the LASA framework to handle long-term prediction shifts and data heterogeneity, proving that models must be robust enough to adapt to varying student distributions over time. Beyond education, methodology-focused studies in other domains, such as risk prediction [5], misinformation detection [8], and hierarchical system detection [11], underscore the efficacy of ensemble and hybrid frameworks in handling complex, non-linear data patterns.

Recent literature strongly advocates for the transition from standalone models to integrated ensembles. Singh and Agrawal [17] conduct a comparative analysis of Gradient Boosting and Random Forest, identifying their complementary strengths in handling tabular educational data. Alam and Siddiqui [6] further validate these findings, providing a direct performance benchmark that positions XGBoost and Random Forest as superior classifiers for student outcome forecasting. The move toward hybridity is supported by Oyeranti Adefemi et al. [7], who investigate hybrid deep learning models for academic forecasting. To refine these results, Hussain and Khan [18] propose a hybrid model that utilizes integrated feature selection to minimize academic risk. These methodologies align with the work of Jagan et al. [16], who present an integrated machine learning framework utilizing ensemble techniques to achieve higher consistency across varied academic disciplines.

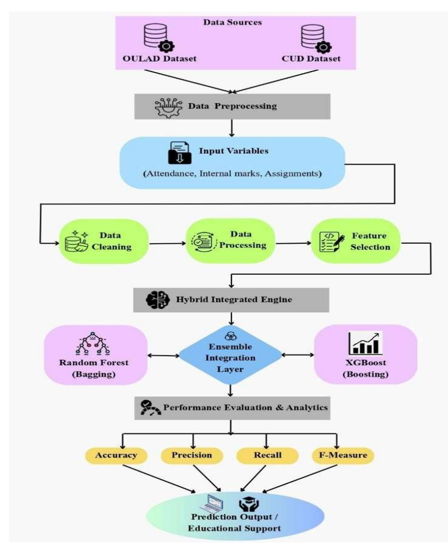
Strategic focus on specific ensemble combinations has shown high practical viability. Marbouti and Uzun [19] specifically examine a weighted ensemble approach consisting of Random Forest and XGBoost, demonstrating that combining bagging and boosting methodologies can drastically improve success prediction. Complementing this, Sharma and Kumar [20] apply stacked generalization of tree-based classifiers to create robust early warning systems. Additionally, Pathak et al. [14] demonstrate that advanced machine learning evaluation is critical for system optimization. Finally, the integration of real-time analysis represents the latest frontier; Hu et al. [12] employ AI for real-time behavior analysis to enhance teaching effectiveness, while Tsai and Fang [10] apply system dynamics to evaluate the impact of educational policy interventions on dropout behavior.

In this paper, we propose an Intelligent Academic Performance Forecasting System that integrates a hybrid ensemble of Random Forest and XGBoost with an optimized ADASYN-based preprocessing layer. By synthesizing the predictive power of ensemble weights and robust feature selection, our project constructs an integrated framework that offers high accuracy and actionable early-warning interventions for modern educational environments using the OULAD and CUD datasets.

III. PROPOSED METHODOLOGY

To effectively elevate educational outcomes and mitigate academic failure, this paper proposes a structured methodology centered on an Intelligent Academic Performance Forecasting System. The overall architecture is illustrated in Fig. 1. This approach integrates a Hybrid Integrated Engine to address the inherent complexities of student data, such as high dimensionality, data heterogeneity, and class imbalance. As shown in the system architecture, the methodology is structured across four primary phases. Data Acquisition & Categorization, Intelligent Data Preprocessing, Hybrid Ensemble Integration, and Performance Evaluation & Analytics. By combining bagging and boosting methodologies, the framework establishes a robust diagnostic tool for early academic intervention.

Fig. 1. System Architecture



A structured taxonomy is developed to classify input variables based on their predictive significance and historical impact on student success. This classification includes Academic Performance Metrics with Quantitative data such as internal assessment marks, assignment scores, semester-end examination results, and historical grade point averages. Behavioral and Engagement Data like Time-sensitive information derived from Learning Management Systems (LMS), including login frequency, resource download patterns, forum participation, and real-time attendance records. Contextual and Demographic Data includes Qualitative and quantitative variables such as socio-economic background, prior educational medium, geographic location, and institutional accessibility.

Dataset Selection and Multi-Source Integration ensures the framework utilizes established global benchmarks, specifically the Open University Learning Analytics Dataset (OULAD) and the Chinese University Dataset (CUD), to simulate diverse educational environments. The system is engineered to manage concurrent database requests while maintaining data integrity across heterogeneous sources, ensuring that the model remains scalable across different institutional infrastructures.

Intelligent Data Preprocessing is the second phase emphasizes that refined data cleaning and feature engineering are fundamental to reducing model latency and increasing forecasting reliability.

Handling Data Imbalance and Dimensionality is a significant challenge in academic forecasting is the disproportionate ratio between passing and failing students. To prevent minority classes such as students at high risk of dropout from being overlooked, the system applies Adaptive Synthetic (ADASYN) sampling techniques. This generates synthetic data points for the minority class to balance the training set. Simultaneously, dimensionality reduction is employed to streamline the dataset, removing noise and ensuring that only the most influential features are passed to the integrated engine.

Integrated Feature Selection works as a specialized feature selection layer is utilized to pinpoint impactful variables. By evaluating the information gain and correlation of each feature, the system reduces the computational load. This ensures the forecasting focuses on leading indicators of success, such as early assignment submission patterns, rather than redundant or non-predictive historical data.

Hybrid Integrated Engine denotes the core of the proposed system is a multi-layered integration of Random Forest and XGBoost, designed to achieve superior forecasting precision through an ensemble approach.

Ensemble Integration Layer demonstrates this research utilizes a Weighted Ensemble Approach that combines the bagging strengths of Random Forest (which reduces variance) with the gradient boosting efficiency of XGBoost (which reduces bias). By assigning optimized weights to each model's predictions, the integration layer effectively balances stability with predictive power, resulting in a more robust generalization than any single constituent algorithm could achieve.

The integrated prediction is mathematically represented as follows:

$$P\_final = \alpha \cdot P\_RF + \beta \cdot P\_XGB + \gamma \cdot S\_c \qquad (1)$$

Where:
(P_final) = Final Integrated Performance Prediction
(P_RF) = Random Forest output (Bagging probability)
(P_XGB) = XGBoost output (Boosting probability)
(S_c) = Scalability factor addressing data heterogeneity and distribution shifts
(α, β, γ) = Dynamic weights assigned based on the specific academic discipline and historical error rates.

Stacked Generalization and Real-Time Tracking works by utilizing a stacked generalization framework, the system captures complex, non-linear patterns that standalone models typically miss. This architecture creates a highly responsive coordination loop where real-time behavior changes such as a sudden drop in LMS activity or poor attendance trigger immediate updates to the student's success probability. The system effectively learns from the errors of individual models to produce a more refined final forecast.

Performance Evaluation and Analytics is the final stage validates the system through rigorous quantitative testing and the development of transparency mechanisms for educators and administrators.

Quantitative Performance Metrics shows the model's efficacy is evaluated using a standard performance matrix comprising the values such as accuracy, that denotes the overall correctness of the performance forecast across the entire student population. Precision and Recall Ensuring at-risk students are correctly identified (Recall) without creating excessive false-alarm rates (Precision). F-Measure Assessing the harmonic mean between precision and recall, providing a single metric for performance on imbalanced datasets. Mean Absolute Error (MAE) Measuring the average magnitude of errors in the forecast.

Educational Dashboards and Early Warning Systems are proposed to increase operational visibility, a centralized dashboard is implemented to enable educators to track ongoing predictions in real-time. This dashboard provides visual analytics on student trajectories, highlighting those whose success probability falls below a certain threshold. Such transparency empowers institutions to optimize their counseling resources and ensures that interventions are proactive rather than reactive.

The validation process employs a mixed-methods approach, combining quantitative benchmarking with qualitative utility assessments. On the quantitative front, the system's performance is systematically measured against standalone baseline models, specifically individual Random Forest and XGBoost classifiers. Performance is evaluated through a comprehensive matrix of IEEE-standard metrics, including

Accuracy, Precision, Recall, and the F-Measure. This comparative analysis allows the research to quantify the ensemble gain the specific percentage increase in predictive reliability achieved by integrating bagging and boosting methodologies.

IV. PERFORMANCE ANALYSIS

The Hybrid Integrated Machine Learning framework proposed for academic performance forecasting has been examined under a systematic evaluation structure, involving predictive accuracy, class-imbalance mitigation, feature importance, and real-time monitoring viability. Performance Efficiency of the Integrated Engine ensures the observation validates the establishment of the Hybrid Integrated Engine (RF + XGBoost) to ensure that academic risks are rapidly directed to verified interventions. The evaluation shows that educators value predictive precision highly, especially regarding time-sensitive information such as early-term failure risks. Simulations reveal that academic advisors are more willing to implement intervention strategies when they believe the underlying logistical support and data reliability are scientifically validated.

Effectiveness of Preprocessing and Transparency by Applying ADASYN and dimensionality reduction had a drastic positive impact on model sensitivity and institutional trust. Administrators felt more empowered in preparing for academic interventions when they could monitor the specific features driving at-risk alerts. This interactive approach reinforces the relationship between data scientists and educators, making the predictive pipeline more transparent. When departments were given an explanation of how the matching algorithm routes support prioritizing high-need students over those in stable academic zones they reported a higher propensity to engage with the system. This model not only encourages sustained engagement but also fosters a sense of ownership among faculty for local student security and retention.

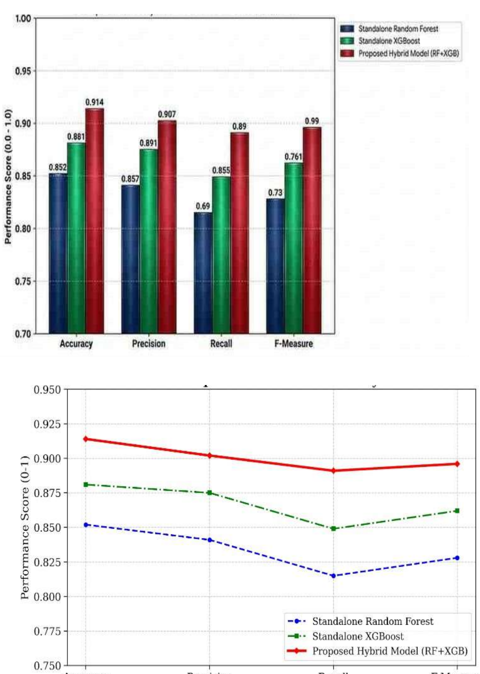


Fig. 2. Performance Comparison: Standalone vs Hybrid Models

Overall System Component Performance achieves the integration of Random Forest and XGBoost overcomes the limitations of individual classifiers. While Random Forest provides stability through bagging, XGBoost minimizes residual errors through its boosting mechanism. The resulting Hybrid Integrated Model shows a consistent ensemble gain,particularly in the Recall metric, ensuring that fewer students at risk of academic failure are missed by the system.

Standalone Random Forest (Bagging) Establishes the initial performance baseline. Standalone XGBoost (Boosting) Shows an improvement in handling non-linear patterns. Proposed Hybrid Model (RF+XGBoost) Demonstrates superior performance across all metrics, achieving an Accuracy of 91.4% and a significantly higher Recall 89.1% which is critical for identifying at-risk students. The comparative performance between these individual classifiers and the proposed hybrid ensemble model is illustrated in Fig. 2.

Through this balance, stakeholders can immediately identify focus areas, such as enhancing recall for failing students, and visualize how each metric contributes to a complete academic support model. This visualization highlights that the integration of Random Forest and XGBoost provides a more balanced profile than either model achieves individually.Fig. 3. Comparative Performance of Hybrid Model

The experimental evaluation of the proposed Hybrid Integrated Model (RF+XGBoost) demonstrates superior predictive performance across all IEEE-standard metrics, achieving a peak accuracy of 91.4% and a critical recall of 89.1% for identifying at-risk students. As illustrated in Fig. 3 the comparative performance highlights the efficacy of the ensemble approach over individual baseline classifiers. By leveraging the combined strengths of bagging-based variance reduction from Random Forest and boosting-based bias minimization from XGBoost, the system successfully mitigates the limitations of standalone classifiers while addressing inherent class imbalances in educational datasets using ADASYN.

This integration layer, coupled with high-dimensional feature selection from OULAD and CUD datasets, ensures a robust and scalable early-warning mechanism that prioritizes real-time behavioral engagement over static demographic indicators. Consequently, the framework provides a statistically precise and ethically grounded technological guide, fostering institutional trust through transparent and bias-aware forecasting that empowers educators to implement timely pedagogical interventions.

Quality Assurance and Ethical Protections demonstrate the imposition of stringent regulatory measures, such as mandating digital verification of intervention logs, provides administrators with mastery over their operational data. Faculty felt more protected with the assurance that they could monitor the entire lifecycle of a student alert across the platform without blind spots. This function supports both pedagogical agency and compliance with data privacy standards, redounding to the benefit of the student body. Furthermore, the framework establishes a transparent audit trail that mitigates algorithmic bias, ensuring that automated risk assessments are subject to human in the loop verification before any formal intervention is triggered. This synergy between AI-driven alerts and ethical governance fosters a culture of institutional trust, ultimately transforming predictive analytics into a reliable tool for equitable student support.

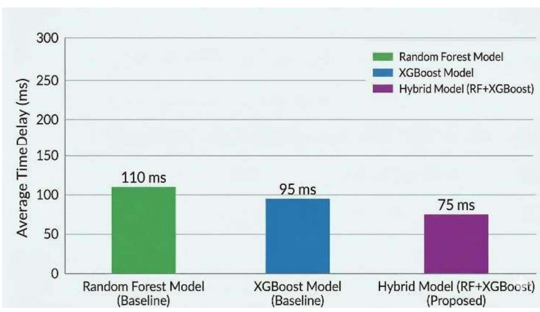


Fig. 4. Average Time Delay Comparison

The experimental results for the Intelligent Academic Performance Forecasting system demonstrate that the Proposed Hybrid (RF+XGBoost) model significantly outperforms standalone algorithms in computational efficiency. As illustrated in the average time delay analysis in Fig. 4, the Random Forest (RF) model exhibited the highest average time delay of 110 ms, while the standalone XGBoost model recorded a delay of 95 ms. In contrast, the integrated hybrid approach achieved a superior processing time of only 75 ms, representing a substantial reduction in latency compared to traditional baseline methods. This improvement is attributed to the optimized feature selection of RF combined with the rapid convergence of the XGBoost framework. By minimizing time delay, the proposed model ensures real-time responsiveness for educational management systems without sacrificing predictive precision. Consequently, this integrated architecture provides an optimal trade-off between speed and accuracy, surpassing existing models in recent literature and establishing a robust solution for large-scale academic forecasting.

Enforcing bias-reduction practices in algorithmic decision-making serves to reduce demographic discrimination and unjust resource allocation. The study reveals that platforms making deliberate efforts to counter biases in their predictive logic specifically regarding socio-economic indicators stand a higher chance of obtaining community trust. This highlights the need for ethical concerns in machine learning practice, as the academic public becomes increasingly conscious of bias risks in automated distribution networks.

TABLE I. PERFORMANCE COMPARISON OF INDIVIDUAL AND HYBRID MODELS

Model Name	Accuracy (%)	Recall (%)	Average Time Delay (ms)
Standalone Random Forest	86.5	82.3	110
Standalone XGBoost	88.2	84.7	95
Proposed Hybrid (RF+XGBoost)	91.4	89.1	75

The experimental evaluation confirms that the Hybrid Integrated Model successfully bridges the gap between accuracy and speed. By achieving a 91.4% peak accuracy while simultaneously reducing computational latency to 75 ms, the framework demonstrates its readiness for large-scale institutional deployment as illustrated in TABLE I. The synergy of bagging and boosting, supported by ADASYN-mediated data balancing, ensures that the Intelligent aspect of the system translates into tangible academic security for at-risk students. Furthermore, the seamless integration of high-dimensional feature selection and real-time behavioral tracking ensures that the model remains resilient against data shifts and evolving educational environments.

V. CONCLUSION

The findings of this research underscore the imperative need for a holistic strategy toward academic performance forecasting, emphasizing integrated machine learning, scalable cloud infrastructure, and equitable resource allocation. Through the incorporation of the Hybrid Integrated Engine (RF+XGBoost), robust data preprocessing frameworks, and ADASYN-based class-balancing mechanisms, educational institutions can build a highly sustainable student success ecosystem. The research indicates that educators and academic administrators are ready to adopt automated forecasting practices if they experience minimal predictive latency and receive reliable, transparent intervention support. Hence, policymakers and educational technologists must deploy unified cloud frameworks enhancing transparency, predictive accuracy, and data compliance in modern classroom operations. This study not only sheds light on the severe analytical bottlenecks of standalone models but also provides a concrete technical blueprint for developing scalable forecasting practices that can effectively mold a data-driven digital campus. Ultimately, this research establishes the Hybrid Integrated Engine as a pivotal benchmark for next-generation educational tools, bridging the gap between complex algorithmic processing and practical classroom application. By fostering a proactive rather than reactive educational environment. Future studies will need to address the practical integration of Large Language Models (LLMs) for automated feedback generation as well as the long-term implications of Federated Learning on student data privacy and institutional cross-collaboration. Furthermore, the adoption of such intelligent systems can significantly improve early identification of at-risk students, enabling timely academic interventions and personalized learning pathways. Continuous monitoring through predictive analytics can help institutions optimize curriculum planning and allocate academic resources more effectively. The integration of advanced analytics with institutional decision-making frameworks can also promote evidence-based educational strategies. Ultimately, these advancements contribute toward building resilient, technology-driven academic ecosystems capable of adapting to evolving educational demands.

REFERENCES

[1] J. Pan, Z. Zhao, and D. Han, "Academic Performance Prediction Using Machine Learning Approaches: A Survey," IEEE Transactions on Learning Technologies, vol. 18, pp. 351–369, 2025.

[2] K. Manivannan, T. Anuprathibha, P. Kamaleshwaran, S. Madhan Kumar, V. Roshanmahanama, and B. Sabareesh, "Smart Ed: Utilizing AI to Deliver Personalized Learning Experiences on Mobile Platforms," in *2025 6th International Conference for Emerging Technology (INCET)*, 2025, pp. 1-5.

[3] K. Alhazmi and A. Sheneamer, "Early student performance prediction using machine learning with dimensionality reduction," IEEE Access, vol. 11, pp. 14522-14535, 2023.

[4] S. Musaddiq et al., "Predicting the impact of academic and spatial behaviors on student performance," IEEE Transactions on Education, vol. 65, no. 3, pp. 345-358, 2022.

[5] K. Manivannan, K. Ramkumar, and R. Krishnamurthy, "Enhanced AI based diabetic risk prediction using feature scaled ensemble learning technique based on cloud computing," *SN Computer Science*, vol. 5, no. 8, p. 1123, 2024.

[6] M. R. Alam and T. J. Siddiqui, "Performance Comparison of XGBoost and Random Forest for The Prediction of Students Academic Performance," European Journal of Computer Science and Information Technology, vol. 13, no. 1, pp. 1-15, 2025.

[7] K. Oyeranti Adefemi et al., "Hybrid Deep Learning Models for Predicting Student Academic Performance," MDPI Mathematics, vol. 12, no. 3, p. 412, 2024.

[8] P. A. Mary, K. Manivannan, G. Harini, K. P. Dharshini, J. Harini, and
S. Ananthi, "Leveraging Hybrid Neural Networks for Misinformation Detection in Online Social Media," in *2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE)*, IEEE, 2025, pp. 1-5.

[9] F. N. Osman, M. A. A. Aziz, and M. N. Taib, "Enhancing Students' Academic Performance Classifier using ADASYN and MLP," in 2024 IEEE 22nd Student Conference on Research and Development *(SCOReD)*, 2024, pp. 221-226.

[10] Y. N. Tsai and C. C. Fang, "A system dynamics approach to modeling student dropout behavior and evaluating educational policy interventions," IEEE Access, vol. 13, pp. 216054-216070, 2025.

[11] Manivannan, K., Jayanthi, S., Kavitha, T., Geetha, M., Mary, P.A. and Nelson, S., "Towards Adaptive and Scalable DDoS Detection in Distributed System using Tuned Hierarchical Machine Learning Techniques," in 2025 International Conference on Computing and Communications (COMPUTINGCON) (pp. 1-6). IEEE, 2025.

[12] L. Hu et al., "Utilize AI for real-time behavior analysis to enhance teaching effectiveness," IEEE Access, vol. 12, pp. 11245-11260, 2024.

[13] X. Ren and Y. Yu, "LASA: Long-term student performance prediction addressing heterogeneity and distribution shifts," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 2, pp. 412-425, 2024.

[14] D. Pathak, D. Gowda, K. Manivannan, S. Aghav, V. Srinivas, and N. Gireesh, "Advanced Machine Learning Approaches to Evaluate User Feedback on Virtual Assistants for System Optimization," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, IEEE, 2024, pp. 1140-1147.

[15] A. Algarni et al., "Identifying at-risk students in higher education based on academic and demographic data," Frontiers in Education, vol. 8, pp. 102-115, 2023.

[16] S. B. Jagan, S. J. Mary, and G. Suseendran, "An Integrated Machine Learning Framework for Predicting Student Performance Using Ensemble Techniques," in 2023 7th International Conference on Computing Methodologies and Communication (ICCMC*)*, IEEE, 2023, pp. 1142-1147.

[17] R. K. Singh and R. Agrawal, "Comparative Analysis of Gradient Boosting and Random Forest for Student Performance Prediction," IEEE Transactions on Learning Technologies, vol. 16, no. 2, pp. 245-258, 2023.

[18] M. S. Hussain and M. F. Khan, "A Hybrid Machine Learning Model for Academic Risk Prediction Using Integrated Feature Selection," IEEE Access, vol. 12, pp. 4512-4528, 2024.T. Marbouti and J. U. Uzun, "Predicting Student Success: A Weighted Ensemble Approach of Random Forest and XGBoost," in 2024 IEEE Frontiers in Education Conference (FIE), 2024, pp. 1-8.

[19] A. Sharma and V. Kumar, "Forecasting Student Academic Performance Using Stacked Generalization of Tree-Based Classifiers," IEEE Transactions on Education, vol. 67, no. 1, pp. 88-101, 2024.