A Hybrid FCM-PSO-ANFIS Model for Predicting Student Academic Performance

Victor Osasu Eguavoen [a,1,*], Emmanuel Nwelih [b,2]

[a,b] Department of Computer Science, University of Benin, Benin City, Edo State, Nigeria
[1] eguavoen.osasu@wellspringuniversity.edu.ng, eguavoenvictor@gmail.com[1]*; [2] emmanuel.nwelih@uniben.edu
* eguavoenvictor@gmail.com

ORCIDs:
First AUTHOR: https://orcid.org/0000-0002-3435-1058
Second AUTHOR: https://orcid.org/0000-0003-4439-7225

ABSTRACT

This study introduces a novel hybrid soft-computing model integrating Fuzzy C-Means (FCM) clustering, Particle Swarm Optimization (PSO), and Adaptive Neuro-Fuzzy Inference System (ANFIS) to enhance student academic performance prediction in higher education. We developed a hybrid FCM-PSO-ANFIS model using a comprehensive dataset encompassing pre-admission data, academic constraints, and student performance records. The model's performance was evaluated against standard ANFIS and Genetic Algorithm-optimized ANFIS using multiple error metrics. The proposed PSO-optimized ANFIS model demonstrated superior performance, achieving the lowest error metrics in both training (MSE: 0.16667, RMSE: 0.40826) and testing (MSE: 0.19748, RMSE: 0.44439) phases. Comparative analysis showed that our model outperformed standard ANFIS and GA-optimized ANFIS in terms of prediction accuracy and generalization capability. The hybrid FCM-PSO-ANFIS model offers a robust, adaptive tool for early identification of at-risk students, enabling timely interventions and personalized learning approaches. This research contributes to improving educational outcomes and retention rates in higher education institutions by providing more accurate and reliable predictions of student performance. Future work should focus on enhancing model interpretability, addressing computational complexity, and exploring applications in diverse educational contexts.

Riwayat Artikel
Diterima 20 Juli 2024
Diperbaiki 1 Oktober 2024
Diterbitkan 6 Oktober 2024

KEYWORDS
Student Performance Prediction,
Hybrid Soft-Computing Model,
Fuzzy C-Means Clustering,
Particle Swarm Optimization,
Adaptive Neuro-Fuzzy Inference System

1. Introduction

In an era of rapidly evolving educational landscapes and increasing student diversity, the ability to accurately predict and proactively manage student academic performance has become more critical than ever. Studies from multiple sources indicate that the problem of college students failing to complete their degrees continues to be a widespread concern worldwide. Graduation rates have shown little improvement since 2015, suggesting this challenge remains unresolved [1]. This alarming statistic underscores the urgent need for more effective tools to identify at-risk students and implement timely interventions.

Student achievement encompasses a wide range of academic activities and outcomes, reflecting both quantitative measures and qualitative aspects of learners' growth [2]. Higher education institutions have increasingly focused on anticipating student outcomes to implement proactive measures against dropout rates and underperformance [3]. This approach enables early identification of potential obstacles, allowing for customized support systems [4]. Assessment involves systematically examining a program's efficacy by gathering and analysing information about its characteristics, procedures, and outcomes [5]. Student achievement assessment, crucial for gauging learning effectiveness, includes diverse assignments that challenge students to produce original work showcasing their skills and critical thinking abilities. Early prediction of academic performance can significantly enhance educational outcomes, influencing university policies, teaching strategies, and learning effectiveness assessments [6]. By identifying potentially struggling students early, educators

e-ISSN: 2809-3399, p-ISSN: 2338-5197
Vol. 12., No. 3, Oktober 2024, pp. 91-98

can implement timely interventions. Higher education institutions can leverage advanced information extraction and computational analysis techniques to strategically examine and oversee student achievement [7]. Data-driven computational methods, particularly those rooted in computational intelligence, excel in forecasting outcomes with high precision, enabling proactive support strategies.

1.2 Problem Statement

Current approaches to predicting student academic performance face challenges including complex relationships, data uncertainty, limited adaptability, and optimization difficulties. To address these issues, we propose a novel hybrid soft-computing model integrating Fuzzy C-Means, Particle Swarm Optimization, and Adaptive Neuro-Fuzzy Inference System. This approach aims to provide a more comprehensive, accurate, and adaptable tool for predicting student performance across diverse higher education settings, potentially enhancing early intervention strategies and personalized learning approaches.

2.1 Related Works

2.1.1 Machine Learning Approaches

Recent research has made significant strides in developing models to predict student academic performance. A study by [8] introduced a novel multi-output hybrid ensemble approach, utilizing data from an integrated educational platform. Their model demonstrated high accuracy in forecasting examination results. Building on this progress, [9] unveiled the Learning Ability Self-Adaptive Algorithm, which showed notable improvements in predictive precision compared to existing methods. Complementing these advancements, [10] explored a hybrid methodology that merged optimization strategies with Support Vector Machines, focusing on performance prediction in online educational settings and stressing the importance of adhering to quality assurance criteria. These investigations highlight the growing sophistication of machine learning applications in education, tackling complex issues like data distribution changes and feature space fluctuations over extended periods.

2.1.2 Nature-Inspired Optimization Algorithms

Recent research has explored the integration of nature-inspired optimization algorithms with traditional machine learning models to enhance prediction accuracy. [11] improved Gaussian Process Classification (GPC) by incorporating the Golden Eagle Optimizer (GEO) and Pelican Optimization Algorithm (POA). Their findings showed that GPC+GEO significantly improved predictions for poor grades, while GPC+POA excelled in predicting Acceptable and Excellent grades. Similarly, [7] enhanced the Naive Bayes classifier (NBC) using Jellyfish Search Optimizer (JSO) and Artificial Rabbits Optimization (ARO). Their NBAR model outperformed the NBJS model in predicting certain grade categories, demonstrating the potential of nature-inspired algorithms in academic performance prediction.

2.1.3 Data-Driven Approaches

The literature emphasizes the importance of leveraging extensive datasets for more accurate predictions. [9] utilized data from Tsinghua University to develop their LASA model, while [7] analysed data from 395 students to validate their enhanced NBC models. These studies underscore the value of large-scale data analysis in developing robust prediction models.

2.1.4 Multifactor Consideration

Recent research trends indicate a shift towards considering multiple factors in predicting student performance. [4] highlighted the importance of customized support systems based on early identification of potential obstacles. This approach aligns with the growing recognition that student achievement is influenced by various factors beyond traditional academic metrics.

2.2 Advancing the Field

The proposed model combines the flexibility of FCM in modelling complex relationships between student performance factors, the optimization capabilities of PSO for fine-tuning parameters, and the adaptive learning capabilities of ANFIS to improve prediction accuracy over time. This study proposes a hybrid model that integrates FCM's flexibility in modelling complex relationships between student performance factors, PSO's optimization capabilities, and ANFIS's adaptive learning, building upon recent trends in multifactor consideration to create a practical tool for educators that overcomes individual technique limitations, potentially offering more robust predictions and contributing to data-driven computational methods in education for enhanced early intervention strategies and personalized learning approaches.

3.0 Methodology

The proposed research introduces an innovative hybrid model that integrates Fuzzy C-Means (FCM), Particle Swarm Optimization (PSO), and Adaptive Neuro-Fuzzy Inference System (ANFIS) to address limitations in current predictive and reasoning systems for student performance prediction. This section outlines each component and explains their roles in the hybrid model. At the core of this model lies a comprehensive knowledge repository, housing pre-admission data, academic constraints, and student performance records, along with optimized information. The PSO component is utilized to identify optimal solutions and parameters for ANFIS model training. Fig. 1, shows the framework of the proposed hybrid system.

3.1 Data Collection and Preprocessing

3.1.1 Dataset Acquisition

This study utilized a comprehensive longitudinal dataset obtained from the Computer Science Department at the University of Benin, Nigeria. The dataset encompassed academic records spanning from 2010 to 2013, providing a rich source of information for analysis. The initial dataset comprised eight primary variables: registration number, entry session, faculty, department, admitted course, screening score, Unified Tertiary Matriculation Examination (UTME) score for the first year, and course-specific scores and grades for subsequent academic years.

3.1.2 Data Preprocessing

The raw dataset of 287 first-year students underwent rigorous preprocessing using Python, including data cleaning, transformation, feature engineering, and validation, resulting in a refined dataset of 238 students with complete information, thus providing a robust foundation for subsequent analysis and modelling.

3.1.3 Longitudinal Data Characteristics

The preprocessed dataset revealed a typical attrition pattern in longitudinal educational studies, with student numbers decreasing from 238 in the first year to 130 in the third and fourth years, reflecting factors like academic progression, program transfers, or withdrawals, which were considered during analysis.

3.1.4 Variable Description and Categorization

Table 1 presents the key variables used in the study, their descriptions, categorizations, and variable types. This categorization approach was designed to facilitate subsequent analysis and enhance the interpretability of results.

Table 1: Variable Description and Categorization

Variable	Description	Categorization	Type
UTME Score	Unified Tertiary Matriculation Examination score	3: >250, 2: 200-250, 1: < 200	Categorical

Avg. Pre-Entrance Score	Mean of SSCE results in core subjects	7: ≥70, 6: 60-69, 5: 55-59, 4: 50-54, 3: 45-49, 2: 40-44, 1: < 40	Categorical
SSCE Attempts	Number of SSCE examination sittings	2: One, 1: Two	Categorical
SSCE Exam Board	Examining body for SSCE	3: WAEC, 2: NECO, 1: NABTEB	Categorical
Admission Age	Student's age at admission	3: <18, 2: 18-23, 1: >23	Categorical
Gender	Student's biological sex	2: Male, 1: Female	Categorical

Note: SSCE = Senior School Certificate Examination; WAEC = West African Examinations Council; NECO = National Examinations Council; NABTEB = National Business and Technical Examinations Board.

3.1.5 Data Quality Assurance

To ensure data quality and reliability, we implemented integrity checks, outlier detection, missing data analysis, and detailed documentation, resulting in a refined dataset of 130 complete student records spanning four years, providing a comprehensive basis for investigating factors influencing academic performance in computer science education.

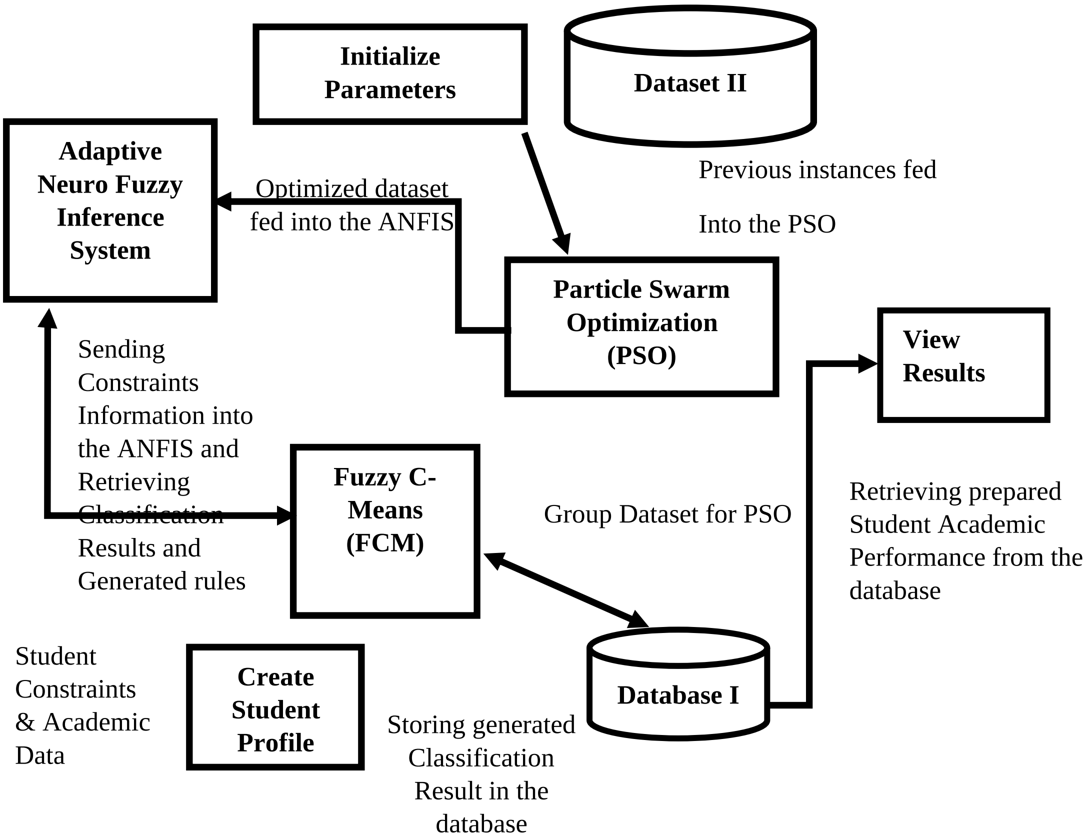

Fig. 1. Framework of the proposed FCM-PSO-ANFIS system

3.2 Components of the Proposed Model

3.2.1 Fuzzy C-Means (FCM)

In our hybrid model, Fuzzy C-Means (FCM) clustering serves as the initial technique, chosen for its capacity to manage ambiguous cluster boundaries - a common characteristic in educational data. FCM operates by iteratively updating membership values and cluster centroids to minimize an objective function. This approach enables the creation of nuanced student profiles based on multiple attributes, establishing a sophisticated foundation for further analysis. By allowing partial membership across clusters, FCM captures the complex nature of student performance data, providing a more realistic representation of student groupings than traditional hard clustering methods.

3.2.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The Adaptive Neuro-Fuzzy Inference System (ANFIS) forms the core of our hybrid model, integrating fuzzy logic interpretability with neural network adaptability. Its six-layer architecture effectively manages educational data complexity while maintaining interpretability through fuzzy rules. This approach enables a predictive model that adapts to nuanced student performance patterns and provides accessible insights for educators.

3.2.3 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is employed to fine-tune ANFIS parameters in our hybrid model. Chosen for its ability to navigate high-dimensional spaces and avoid local optima, PSO iteratively updates particle velocities and positions based on individual and swarm best solutions. This optimization approach aims to enhance ANFIS performance, improving prediction accuracy and generalization by tailoring the model to our specific educational dataset characteristics.

3.3.1 Synergy of Components in the Hybrid Model

The FCM-PSO-ANFIS hybrid model leverages powerful synergies to address key challenges in student performance prediction, combining FCM's ability to handle fuzzy boundaries and manage high-dimensional data, ANFIS's adaptive learning and interpretable rule framework, and PSO's global optimization capabilities. This integration results in a robust, adaptive system that effectively handles data complexity and uncertainty, enhances initial rule generation, balances interpretability with accuracy, adapts to changing educational landscapes, and demonstrates resilience to noise and outliers, ultimately providing accurate, interpretable, and adaptable predictions of student academic performance across diverse educational contexts.

4.0 Results and Discussion

4.1 Results

The performance metrics of the standard ANFIS model, ANFIS with Genetic Algorithm (GA), and ANFIS with Particle Swarm Optimization (PSO) were evaluated during both training and testing phases, as detailed in Table 2 and visually represented in Fig 2.

Table 2: Model Performance Metrics

Model	Phase	MSE	RMSE	Error Mean	Error Std Dev
Standard ANFIS	Training	0.42226	0.64982	-5.08e-16	0.65342
	Testing	0.54399	0.73755	0.0019621	0.74719
ANFIS with GA	Training	0.25297	0.50296	-0.03813	0.50429
	Testing	0.17210	0.41485	-0.14664	0.39314
ANFIS with PSO	Training	0.16667	0.40826	-0.012929	0.41031
	Testing	0.19748	0.44439	0.18933	0.40729

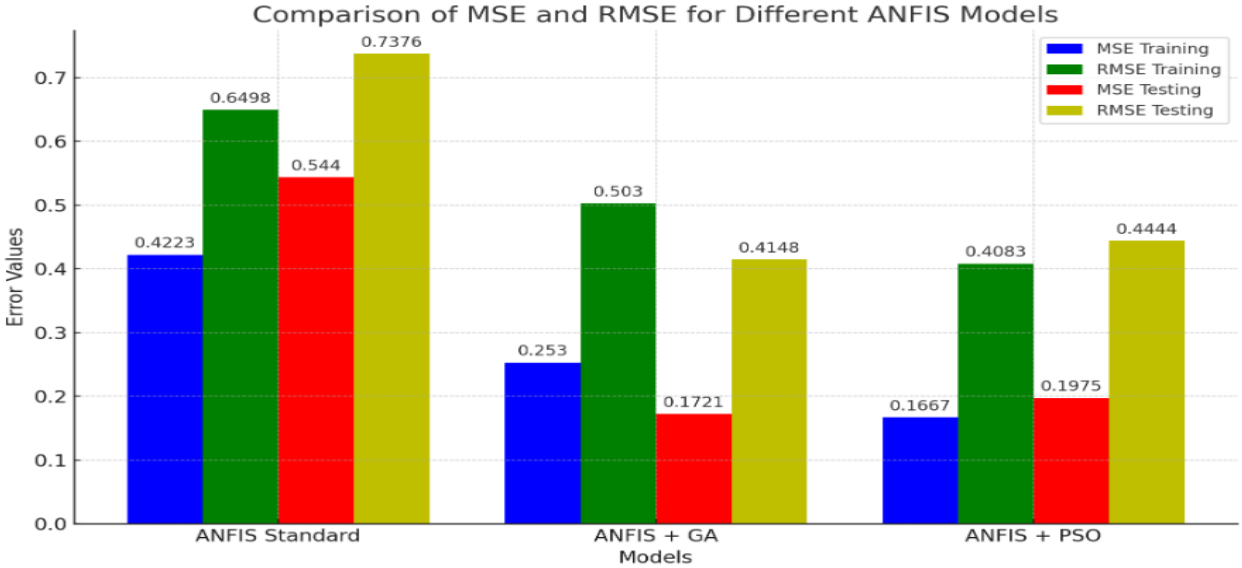

Fig. 2. Comparison of MSE and RMSE for Different ANFIS Models

4.1.1 Standard ANFIS Model

The Standard ANFIS model exhibited an MSE of 0.42226 and an RMSE of 0.64982 during the training phase. However, its performance declined in the testing phase, with an MSE of 0.54399 and an RMSE of 0.73755. The error mean was nearly zero during training but slightly increased to 0.0019621 during testing, with the error standard deviation also increasing from 0.65342 in training to 0.74719 in testing. This indicates a higher variability and less reliable performance during testing.

4.1.2 ANFIS with Genetic Algorithm

The GA-optimized ANFIS model demonstrated substantial improvement over the standard model. During training, it achieved an MSE of 0.25297 and an RMSE of 0.50296. The testing phase results showed further improvement, with an MSE of 0.17210 and an RMSE of 0.41485, suggesting strong generalization capabilities. The error mean was -0.03813 in training and -0.14664 in testing, reflecting a slight negative bias, while the error standard deviation was 0.50429 in training and reduced to 0.39314 in testing, indicating more consistent predictions.

4.1.3 ANFIS with Particle Swarm Optimization

The PSO-optimized ANFIS model exhibited the best overall performance. It recorded the lowest training phase MSE of 0.16667 and an RMSE of 0.40826. The testing phase results were also robust, with an MSE of 0.19748 and an RMSE of 0.44439. The error mean was -0.012929 in training and slightly positive (0.18933) in testing, with the error standard deviation being 0.41031 in training and 0.40729 in testing, indicating highly consistent performance across both phases.

4.2 Discussion

The comparative analysis clearly shows that both GA and PSO optimizations significantly enhance the performance of ANFIS models compared to the standard version. The GA-optimized model exhibited superior generalization to unseen data, as evidenced by the lowest testing phase MSE and RMSE values. The PSO-optimized model, while also showing excellent testing performance, was particularly notable for its lowest training phase error metrics and consistent performance across both phases.

The visualization in Fig. 2, supports these findings, showing a significant reduction in MSE and RMSE values for the GA and PSO models compared to the standard ANFIS. This reduction is more pronounced in the testing phase, highlighting the improved generalization capabilities of the optimized models. These results underscore the effectiveness of incorporating optimization algorithms to enhance the predictive accuracy and reliability of ANFIS models, making them more suitable for practical applications in various fields, including education and data-driven decisionmaking. The integration of GA and PSO optimization techniques with ANFIS models leads to superior performance, characterized by lower error metrics and enhanced generalization capabilities. This highlights the potential of these hybrid models in improving the accuracy and reliability of predictive systems.

4.2 Comparison with State-of-the-Art Methods

Our hybrid ANFIS models, optimized with GA and PSO, demonstrate competitive performance against state-of-the-art methods in student performance prediction, including [9] LASA, [11] GEO-POA enhanced GPC, and [7] JSO-ARO enhanced Naive Bayes, showing low error rates and potentially superior predictive power across all performance levels, while significantly improving upon traditional ANFIS implementations in educational data mining.

5.0 Conclusion

This study introduces a novel FCM-PSO-ANFIS hybrid model for predicting student academic performance in higher education. The model demonstrates superior accuracy and generalization compared to standard and GA-optimized ANFIS models, effectively addressing data imprecision and context adaptability while providing interpretable results. It offers significant potential for early identification of at-risk students and personalized learning strategies. While challenges in computational complexity and interpretability persist, this research advances educational data mining. Future work should focus on real-time applications, diverse educational settings, and ethical considerations. The model represents a valuable tool for enhancing educational outcomes and improving retention rates in higher education.

5.1 Practical Implications and Potential Real-World Applications

Our hybrid ANFIS models, optimized with GA and PSO, demonstrate superior performance with multiple practical applications in education. They enable early intervention systems, personalized learning paths, efficient resource allocation, informed admissions, curriculum enhancement, and real-time performance monitoring. These models' accurate predictions across various factors can potentially improve educational strategies, reduce dropouts, enhance academic outcomes, and optimize student learning experiences.

5.2 Limitations and Future Work

The FCM-PSO-ANFIS hybrid model shows promise but faces limitations in data specificity, temporal constraints, feature selection, interpretability, ethics, and real-time adaptation. Future work should address these through cross-institutional validation, longitudinal studies, feature expansion, efficiency optimization, and interpretable AI integration. Additional focus areas include ethical frameworks, real-time adaptation, LMS integration, comparative studies, multi-objective optimization, transfer learning, and uncertainty quantification. These improvements aim to create a more robust, adaptable, and ethical predictive tool for enhancing educational outcomes across diverse settings.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] H. Sutton, "Six-year completion rates stay mostly flat, though adult learners, community colleges offer bright spots," Recruiting & Retaining Adult Learners, 26(4), pp. 9–9. doi:10.1002/nsr.31119., 2023.

[2] Z. Xu, H. Yuan and Q. Liu, "Student performance prediction based on blended learning," IEEE Transactions on Education, 64(1), pp. 66–73. doi:10.1109/te.2020.3008751., 2021.

[3] L. Vives, I. Cabezas, J. C. Vives, N. G. Reyes, J. Aquino, J. B. Cóndor and S. F. Altamirano, "Prediction of students' academic performance in the programming fundamentals course using long short-term memory neural networks," IEEE Access, 12, pp. 5882–5898. doi:10.1109/access.2024.3350169., 2024.

[4] B. Cheng, Y. Liu and Y. Jia, "Evaluation of students' performance during the academic period using the XG-boost classifier-enhanced AEO hybrid model," Expert Systems With Applications, 238, 122136. https://doi.org/10.1016/j.eswa.2023.122136, 2024.

[5] V. Eguavoen and E. Nwelih, "Hybrid Soft Computing System for Student Performance Evaluation," Studia Universitatis Babeș-Bolyai Engineering, 68(1), pp. 3–17. doi:10.24193/subbeng.2023.1.1., pp. 3-17, 2023.

[6] F. Ofori, E. Maina and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review.," Journal of Information and Technology, 4(1), 33–55., 2020.

[7] X. Zheng and C. LI, "Predicting students' academic performance through Machine Learning Classifiers: A study employing the naive bayes classifier (NBC)," International Journal of Advanced Computer Science and Applications, 15(1), doi:10.14569/ijacsa.2024.0150199, 2024.

[8] H. Xue and Y. Niu, "Multi-output based hybrid integrated models for student performance prediction," Applied Sciences, 13(9), p. 5384. doi:10.3390/app13095384., 2023.

[9] Y. Ren and X. Yu, "Long-term student performance prediction using learning ability self-adaptive algorithm," Complex & Intelligent Systems, doi:10.1007/s40747-024-01476-2, 2024.

[10] M. Nasser Alsubaie, "Predicting student performance using machine learning to enhance the quality assurance of online training via Maharat Platform," Alexandria Engineering Journal, 69. doi:10.1016/j.aej.2023.02.004, pp. 323-339, 2023.

[11] H. Fan, G. Zhu and J. Zhan, "Student performance estimation through innovative classification techniques in Education," International Journal of Advanced Computer Science and Applications, 15(3). doi:10.14569/ijacsa.2024.0150384, 2024.