Advances in Artificial Intelligence and Machine Learning;Research 6 (4) 6014-6032 Received 16-01-2026; Accepted 10-08-2026; Published online 17-08-2026

Machine Learning Based Evaluation and Prediction of Student Performance in Virtual Learning Environment

Umair Abbasi Umairaabbasi708@gmail.com
Department of Computer Science, Bahria University
E-8, Islamabad, Pakistan

Saba Mahmood smahmood.buic@bahria.edu.pk
Department of Computer Science, Bahria University
E-8, Islamabad, Pakistan

Heba Fasihuddin hafasihuddin@uj.edu.sa
Department of Information Systems and Technology
College of Computer Science and Engineering, University of Jeddah
Jeddah, Saudi Arabia

Safa Habibullah shhabiballah@uj.edu.sa
Department of information Systems and Technology,
College of Computer Science and Engineering,
University of Jeddah, Jeddah,
Saudi Arab.

Manal Linjawi malinjawi@uj.edu.sa
Department of Computer Science and Artificial Intelligence,
College of Computer Science and Engineering,
University of Jeddah, Jeddah, Saudi Arab.

Ali Daud Adaud@ra.ac.ae
Faculty of Resilience, Rabdan Academy,
Abu Dhabi, UAE.

Corresponding Author: Ali Daud and Saba Mahmood

Abstract

In recent years, the rapid expansion of Virtual Learning Environments (VLEs)and online education platforms has significantly transformed higher education.These education setups have introduced new challenges for example increased student attrition and disengagement. Predicting student performance within these digital frameworks is essential to enabling timely interventions and improving academic outcomes. This research addresses the prediction of high risk students by leveraging the Open University Learning Analytics Dataset (OULAD) that comprises demographic, assessment, and detailed engagement data for over 32,000 students. Recognizing gaps in the literature, this study develops a robust ensemble based machine learning pipeline with Particle Swarm Optimization(PSO) as a feature selection strategy. The results demonstrate that PSO reduced the feature space by approximately 40% while maintaining high predictive performance (F1 $\approx$ 0.937, AUC $\approx$ 0.980) in comparison to Gini importance and Factor Analysis of Mixed Data(FAMD). Overall 94% accuracy and AUC scores near 0.981 is achieved that is comparable to advanced deep learning benchmarks. The inclusion of SHAP analysis improves the interpretation of the proposed scheme. This work con-

Citation: Umair Abbasi, et al. Machine Learning Based Evaluation and Prediction of Student Performance in Virtual Learning Environment. **Advances in Artificial Intelligence and Machine Learning**. 2026;6(4):333. https://dx.doi.org/10.54364/AAIML.2026.64333

tributes a balanced, explainable framework for educational early warning systems thereby, bridging methodological gaps in comparative feature selection techniques and offering practical insights for scalable deployment in online learning contexts.

Keywords: Virtual Learning Environments (VLEs), Student Academic, Learning Analytics, Machine Learning in Education, PSO,Explainable AI

1. INTRODUCTION

The world has seen an enormous shift of education to online and virtual environments in the past 5 years in post COVID scenarios. The effectiveness of VLE and MOOCS emerged as game changer and offered opportunities to students and instructors to learn regardless of their geographical locations. These VLE have produced a large amount of data that can be mined to provide insights about the learning capabilities of students and offer various new avenues to the educational institutions, instructors to adapt the courses and personalize the learning experience for the students. Recently researches showed that machine learning models can be effectively utilized to monitor the dropout and delayed graduation causes [1].

Jine et al. (2024) [2], systematically reviewed the use of deep and machine learning based models in forecasting academic outcomes of the students. ML based learning analytics tool mine the student datasets comprising of student demographic details, assessment and behavior pattern of students . The tools can aid the instructors to monitor the student progress and can help in identification of students who may require an early intervention in improvement of the learning experience [3]. Learning analytics dashboards further support this process by helping instructors monitor progress and identify students who may require early support [4].

The early warning systems based on these analytics can help teachers to personalize the learning experience of the students thereby improving the student retention rate and offering equal learning opportunities. Badshah et al. (2023) [5], has shown the effective utilization of IoT with AI technologies in digital classrooms. Neural Network based educational prediction systems can effectively capture large scale student behavior patterns [6].

Open University Learning Analytics Dataset (OULAD) consists of data related to demographics, engagement factors and assessments. If an efficient feature selection method is not applied, this could result in overfitting, higher processing costs, and decreased interpretability. Models like Random Forest and Artificial Neural Networks have been investigated in earlier researches Al-Zawqari et al. (2022) [7]. However, a comprehensive evaluation across various feature selection techniques and predictive algorithms in similar settings not reported. Moreover, while deep learning models can achieve strong predictive accuracy [3], their “black box” nature limits transparency, making them less suitable for educational settings where interpretation and actionable insights are essential. This indicates the need for analysis of different feature selection techniques in combination with machine learning models to evaluate the improvement in the accuracy and interpretation of student outcome predictions in VLEs [7].

We have conducted systematic comparison of Particle Swarm Optimization, Gini Importance and Factor Analysis of mixed data (FAMD). Different machine learning models including Random Forest (RF), Logistic, Gradient Boosting (GB) Regression (LR) and Artificial Neural Network (ANN) coupled with ensemble configurations are proposed. These approaches are evaluated utilizing F1 Score, AUC, Precision, Recall and Accuracy. These results emphasize the application of diverse strategies for feature selection and an ensemble-based method while seeking a compromise among accuracy, interpretability, and scalability.

The results encourage the creation of educational early warning systems that are not only highly predictive but can also be explained and efficiently computed to be applied to real-world conditions.

1.1 Research Contributions

This study has contributed to the field of learning analytics and education data mining, the key points are listed below.

- This work comprehensively compared the competitive feature selection techniques including PSO, Gini Impurity and FAMD thereby fill research gap identified in the recent literature
- This research has proposed feature engineering methodology by integrating data related to demographics, assessments, inspired by the requirements of the learning analytics dashboard
- It evaluates multiple machine learning models, including RF, GB, LR, and ANN, along with ensemble approaches, demonstrating that high predictive performance (F1 $\approx$ 0.938, AUC $\approx$ 0.981) can be achieved while maintaining model interpretability.
- Inclusion of SHAP-based explainability to highlight the contribution of each feature, supporting transparent decision-making in educational environments.

2. LITERATURE REVIEW

Massive online open courses (MOOCs) and virtual learning environments (VLEs) have contributed to the generation of student behavioral and academic data due to the influx of digital transformation in the higher education sector. Interpretation of data using machine learning (ML) and data analytics can be used to improve the identification of risk students at an early stage and to improve the personalization of interventions. Recent studies highlight that learning analytics dashboards can help instructors monitor student progress more effectively and identify learners who may need support [4].

Predictive analytics has become a strategic concern as learning institutions seek to manage an increasing student diversity, as well as massive online enrollments. There is a growing trend in literature around the use of supervised ML models to characterize students according to their probability of success or likelihood of failure, which can be used for instructor and administration intelligence.
This literature review is a systematic review on the state of the art in predicting student performance interest in strategies of feature selection, machine and deep learning models and their application on highly used educational data sets such as OULAD. The literature review discusses the Datasets utilized in educational data mining with special focus on OULAD dataset followed by recent studies in the domain related to various feature selection techniques. Finally various advancements of machine learning and deep learning models utilized in educational analytics and student performance are elaborated. The related work is designed in a way to highlight the research gaps addressed in this study.

2.1 Datasets in Predicting Student Performance

Predictive analytics in education relies heavily on rich multidimensional datasets that capture not only students' demographic profiles but also their behavioral interactions within learning platforms. Among various datasets used in the literature, the Open University Learning Analytics Dataset (OULAD) has emerged as a benchmark, frequently employed in studies due to its comprehensive scope and availability.

2.1.1 The OULAD dataset

T Kuzilek et al. (2017) [8], introduced the OULAD Dataset. It is a large-scale open dataset derived from the virtual learning system of UK Open University. It consists of 32,593 student records in 22 courses. It includes demographic information, assessment results, and the VLE clickstream data.

Recent studies have utilized the OULAD for building and validating learning predictive models:

- Jin et al. (2024) [2], reviewed the ML based education predictive models utilizing the OULAD dataset.
- Torkhani and Rezgui (2025) [9], used machine learning as well as deep. learning paradigms on OULAD MOOC subset and gaining better than 94 per cent accuracies.
- Rasheed et al. (2025) [10], studied scaled optimized data pipelines. analysis with OULAD, and its robustness in the face of heavy query loads.
- Al-Azawei (2022) [7], analyzed tasks based on point-wise category prediction upon OULAD, and gave their highest accuracy of 93.28% in drawing the line between ‘Withdrawn’ ratio(s): out of pass students by random forest. They also achieved 88.95% because of Analysis of Network (ANN), it is used as for ‘Distinction vs Fail’ using ANN. This study serves as a baseline for the present research, which extends to multi-class prediction using advanced feature engineering, feature selection, and ensemble approaches.

These papers highlight the flexibility of OULAD towards algorithm experimentation and deployment issues. There are other datasets such as the UCI Student Performance dataset and other institution-specific LMS exports, which also feature in the literature, though they are usually less widespread and time-granular than OULAD.

Based on such precedents, the OULAD dataset is used in this research; it is a combination of all demographic, assessment, and clickstream tables. A lot of feature engineering was conducted and more than 70 features were first generated, which were then filtered by using PSO, Gini importance and FAMD. This plan will guarantee that the predictive models will be informed by the study of the entire profile and behavior of each student as befitting and beyond previous strategies. In addition, Muhammad Adnan et al. (2021) [11], highlighted the significance of time-based segmentation in student risk modeling according to which predictions performed at different checkpoints in the course lifecycle may demonstrate different intervention results. Their experiment showed that early warning systems using varying course lengths are useful in enhancing agility of institutional response. This research builds upon that notion by aiming to enhance early stage accuracy through comprehensive feature selection and high-resolution behavioral tracking from the OULAD dataset.

2.1.2 Typical features explored

The typical features fall into three categories, including demographic, behavioral, and academic TABLE 1.

Table 1: Typical Feature Categories Across Datasets

Feature Type	Examples
Demographic	Age band, gender, region, disability, highest education level
Behavioral / Interaction	Click counts, resource views, participation frequency, session diversity
Academic / Performance	Past assignment scores, early assessment grades, final exam results

A number of studies have been conducted to determine the predictive capability of demographic and engagement-based features. Bilal et al. (2022) [12], showed that the metrics of learner behavior including the patterns of clicks and the time of submitting assignments were closely related to student achievements. On the same note, Muhammad et al. (2022) [13], processed VLE data using the Random Forest models and emphasized the significance of assessment scores and attendance tracking to the F1-score performance. The application of binary classification to categorize pass/fail results by background characteristics alone, further streamlined by Idrrisu et al. (2023) [14], was used as a consequence of the need to incorporate richer data of behavioral descriptions to perform effective modeling.

The more sophisticated researches tend to form composite measures of engagement and risk. The fact that the highest education level is a variable of the highest importance, along with the importance of the previous qualification, is emphasized in Adekitan and Noma-Osaghae (2019) [15], as the authors state that these variables are important in predicting early academic success. They are useful given their inclusion in the dimensionality reduction step because when they are filtered using either PSO or Gini based methods, they may hold latent predictive information. For instance:

- Fazil et al. (2024) [16], developed weighted measures of engagement based on using a combination of clicks and recency decay.
- This study also presented variables including total_clicks, interaction_diversity, weighted_score, and engineered engagement ratios.

These predetermined features may be highly sensitive to the models in the face of those subtle decorations of conduct that predict the development of academic decline. MOOC forum analysis shows that aspect-based methods can extract meaningful insights from student feedback to help instructors respond to concerns more quickly [17]. Consequently, Deepti Aggarwal et al. (2021) [18], established that the non academic factors such as because motivation, self-discipline, and emotional resilience might have an important role in influencing academic performance, provided they are modeled by the use of ensemble learning. Advanced learning analytics frameworks show that the combination of behavioral patterns with machine learning methods can greatly improve prediction accuracy [19]. They point out the significance of computing behavioral measures on cognitive measures, which is an addition to the multi-dimensional feature space developed in the study.

2.2 Feature Selection Techniques in Predicting Student Performance

In order to minimize model complexity, overfitting, and improve generalization, feature selection is crucial. Zhang et al. (2022) [20], offer a conceptual taxonomy of feature selection methods and divides them into filter, wrapper, and embedded (this study follows the latter approach based on the comparative analysis of Gini importance (embedded), PSO (wrapper), and FAMD (reduction). Guyon & Elisseeff (2003) [21], divides the methods of feature selection into filter, wrapper, and embedded, which has become a foundation of the feature engineering theory. Their taxonomy guides the hybrid strategy adopted in this study in which Particle Swarm Optimization (PSO) is a wrapper-based strategy, Gini importance is an embedded strategy, and FAMD is an unsupervised and filter-based dimensionality reducer.

2.2.1 Importance of feature selection in educational data

Educational data particularly VLE-based data such as that of OULAD contain dozens of features which includes both demographic, behavioral as well as assessment features. Even though high-dimensional datasets provide more valuable information, they are associated with the danger of overfitting, higher computational cost, and less model interpretability.

2.2.2 Traditional and statistical feature selection

Filter-based FS techniques were vastly employed in previous studies thereby taking advantage of statistical measures.

- Ahmed Habeeb (2023) [22], developed Sequential Feature Selection (D-SFS) that outperformed classical Sequential algorithm.They selected features through iterative refinement. Features were reduced to only seven major variables with a prediction accuracy of 92.5% in VLE dropout predictions.
- Navin Kartik et al. (2024) [23], applied correlation-based feature selection (CFS) and information gain. This work emphasized the usefulness of uni variate approaches in the educational settings.
- Gini importance, which is based on the tree-based models similar to Random Forests,serve as a popular metric. It measures the contributions of every feature.

In this research, Gini-based importance was directly tested in the research by taking the top 10 features of a PSO-optimized Random Forest, which was later compared with other FS outputs.

2.2.3 Metaheuristic approaches

Recognizing the limitations of purely statistical filters, several studies have explored metaheuristic algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) to perform wrapper-based feature selection:

- As it was shown by Apriyadi et al. (2022) [24], a combination of PSO and GA was much better combined with SVR predictive accuracy in the case of prediction.
- A cross-border relationship between different nations of the international community are highlighted in the article by Safira Begum & Sunita Padmanna (2022) [25]. They used Binary PSO (BPSO) in feature selection that classified into a CNN feeder. Their hybrid framework has a classification accuracy of 96.6%, highlighting the effectiveness of the technique.

- Kumari & Kumar (2023) [26], advanced this solution by proposing new rules of PSO updates that showed an increase in the accuracy and recall compared to the baseline models.

PSO was also used in combination with Random Forest, which produced a subset of 43-features. The performance results of (F1 $\approx$ 0.938, AUC $\approx$ 0.981)showed the ability of PSO in reduction of the dimensions of the data without compromising on accuracy.

2.2.4 Dimensionality reduction

There are other dimensionality reduction techniques that includes:

- **Factor Analysis of Mixed Data (FAMD)** is a reduction method appropriate for both categorical and quantitative characteristics. FAMD extends principal component analysis (PCA) and multiple correspondence analysis (MCA) into a integrated structure, maintaining variance on mixed data types.
- Within the framework of educational data mining, dimensionality-sensitive methods such as FAMD are inclined towards gaining attraction since they preserve structure and decrease the model complexity [27]. These methods help in early risk modeling and leads to faster convergence during training phase.

2.3 Machine Learning Approaches for Predicting Student Performance

Classical supervised machine learning classifiers are used to estimate student performance, that are easy to interpret providing computationally cost-effective solution. The Random Forest(RF)is oftenly used as it not only captures the interaction but can also be used to identify the feature importance. Kartik et al. (2024) [23], showed that RF is better than J48 and Logistic Model in different academic performance measures. Al-Azawei et al. (2023) [22], used logistic regression(LR)to predict the risk of student dropout. Gradient Boosting (GB) model although computationally expensive, provides the best predictive measures. Gajwani & Chakraborty (2020) [28], results showed improved F1 and AUC scores with boosted trees on blended learning data. RF achieved the highest F1 of 0.938 (PSO-selected features), while GB delivered comparable AUC (0.982)thereby, validating insights from existing literature. The article by Turabieh (2019) [29], examined hybrid frameworks with mix feature selection using ANNs,therby obtaining significant student gains performance prediction on the basis of all-statistical models. Shallow ANNs were also tested by Kartik et al. (2024) [23], along with decision trees, the improvement was mostly observed in the datasets that were augmented with behavioral variables. In this paper, an ANN (MLP Classifier) was applied in testing all subsets, with F1 0.916 on full features, that confirms the ability of ANN to generalize even with limited manual tuning. The increasing scale and Time-based sophistication of pedagogical information has ignited interest in deep learning structures: Masood et al. (2024) [3], suggested a hybrid model training an Evolutionary Convolutional Neural Network. ResNet based Network (ECNN) optimized with Butterfly Optimization with a great result of 95.67% expertise on OULAD. Torkhani and Rezgui (2025) [9], have conducted an experiment on OULAD MOOC data, emphasizing the importance of the sequential and convolutional architectures. Begum and Padmanna (2022) [25], integrated BPSO Online CNN feature selection, demonstrating how feature optimized deep networks may exceed classic models with a 8-10% high precision Nevertheless, deep models are effective in prediction, but their prediction is less precise.A distinct strand in the literature is an amalgamation of many algorithms to utilize their complementary advantages Gajwani & Chakraborty (2020) [28], showed the manner in which methods outperformed single methods. Apriyadi et al. (2023) [30], combined Hyperparameter tuning of SVR models, PSO/GA, cutting down mistakes by a large margin. The second best was proposed by Fan et al. (2023) [31], is a dual PSO optimization of features and parameter optimization in CatBoost that achieved 96.6% accuracy of classification on the secondary education data.

2.3.1 Deep learning extensions & hybrid innovations

Recent literature is strongly motivated by deep learning architectures, frequently along with hybrid feature engineering or optimisation. Torkhani and Rezgui (2025) [9], used several deep and machine learning models on the subset of OULAD MOOC and obtained a classification accuracy of more than 94%. They emphasized that ensemble DL models are able to learn minute patterns of sequential engagement, which is frequently beyond the capability of static ML models.On the other hand, Fazil et al. (2024) [16], introduced a new deep learning model that was fed with finely granular engagement measurements, which has produced significant gains in risky students identification. Their method has shown the benefits of taking advantage of temporal patterns that are usually not fully exploited in simple models. Waheed et al. (2020) [32], also demonstrated the effectiveness of deep learning models to forecast student academic performance based on virtual learning environment (VLE) big data. Their research confirms that ANN based architectures can be implemented on large educational datasets such as OULAD, which justifies the use of neural networks in this research framework. Hernandez-Blanco et al. (2019) [33], conducted a broad review that summarized the tendencies in the use of deep learning in educational settings. They underlined the shift to deep neural nets as the means of processing the more sophisticated data about student behavior. This broad outline underpins the reason why ANN and ensemble models will be present in this study since education data continue to grow larger and more intricate. Deep learning research in areas such as social media analytics demonstrates how neural models can capture complex, large-scale behavior patterns, offering new opportunities for educational prediction [6]. However, these studies also reported trade-offs, especially in interpretability and computational overhead, which strengthens the rationale of why most deployed educational settings continue to be based on classical models with added state-of-the-art feature selection.

Jin et al. (2024) [2], held an extensive systematic review of predictive modeling with the use of OULAD. They distinguished between shallow ML, ensemble learning, and deep learning and indicated gaps including underused metaheuristic feature selection (such as PSO) on a larger scale. Direct comparisons of FS methodologies on the same pipeline are very infrequent.The absence of multi-metric ratings (F1, AUC, precision, recall) that are essential towards subtle academic intervention.

3. METHODOLOGY

The currently used high-risk student predictive methodology has been elaborated in this research. The proposed methodology outlines in an orderly way the dataset, preprocessing, exploratory analysis, feature selection processes, modeling approaches and evaluation. This research approach is intended to explicitly fill research gaps identified in a recent literature, especially the absence of feature selection comparisons across the multi-strategies, and deep multi-metric analyses [2]. FIGURE 1 shows the major modules of the proposed scheme.

3.1 Data Source

The dataset used in this study was processed and prepared in accordance with the standards proposed by Howard (2025) [34], who introduced the ouladFormat R package for standardizing OULAD preprocessing.

Following such best practices enhances reproducibility and ensures consistency in feature engineering and student engagement tracking.

The Open University Learning Analytics Dataset (OULAD) is recognized as benchmark for educational prediction research [2]. OULAD comprises data on approximately 32,000 students enrolled across 22 different courses, with demographic profiles, registration details, assessment scores, and detailed VLE click stream logs.
Various recent researchers have utilized the OULAD dataset for student performance prediction [9, 10].

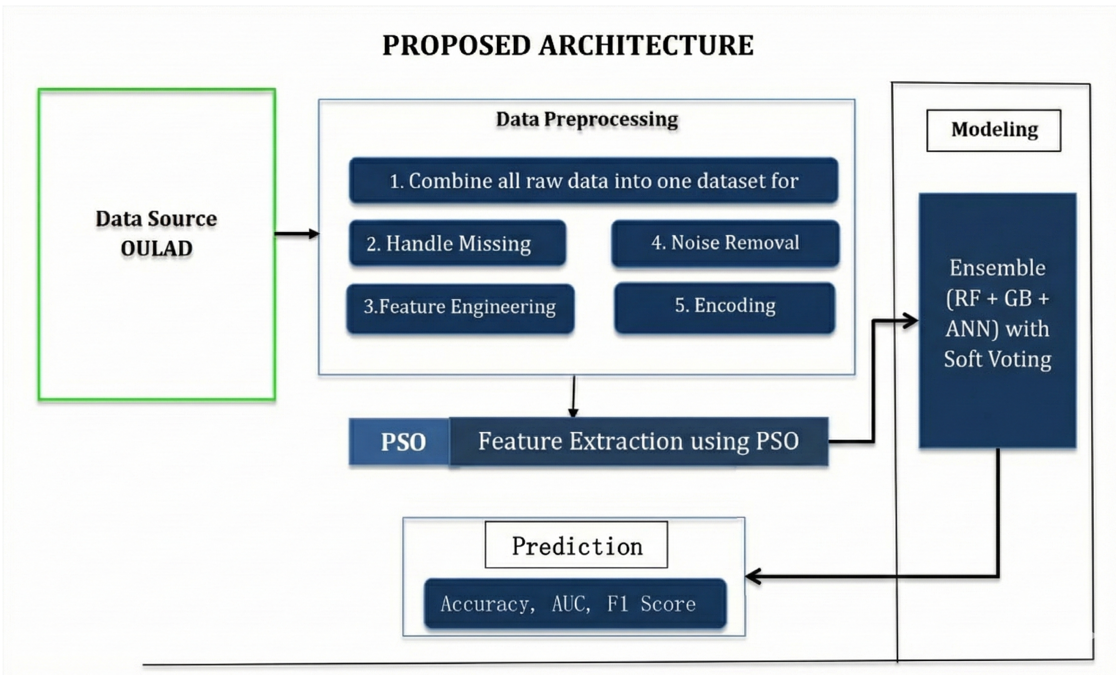

Figure 1: Proposed methodology

3.2 Data Preprocessing and Feature Engineering

3.2.1 Data cleaning

The variety of OULAD tables (studentInfo, assessments, vle, etc.) are firstly combined into an analytical dataset. Categorical variables like disability, region and etc with missing values were filled with the mode.Mean imputation is used for missing numeric features.

3.2.2 Feature engineering

Based on the latest research focused on the importance of having rich behavioral characteristics [28], the following engineered attributes were appended to the dataset:

- **total_clicks:** Aggregate number of VLE interactions, capturing overall engagement intensity.
- **interaction_diversity:** Different kinds of activities accessed, suggestive of depth of involvement.

- **submission_ratio:** Ratio of assessments submitted to task scheduled, which measures academic diligence.
- **weighted_score:** The weighted statistic that takes into consideration the type of assessment and scores.
- **registration_lag:** Days between registering the module and officially starting, which is a measure of preparatory behavior.
- **days_active:** The sum of unique days attended to.

It obtained a predictive feature space of 72 variables, a combination of demographic, assessment, and interactions, which is aligned with the best practices in educational data mining.

Once the creation of the features was done, a number of columns were discovered as being redundant or not useful anymore in the model such as date registration, date unregistration, code module, code presentation and length of module presentation. These were dropped to simplify the dataset, avoid data leakage and enhance the effectiveness of the model.

3.3 Exploratory Data Analysis (EDA)

Before feature selection and modeling, exploratory analyses were performed to know the distributions of data and their important patterns. The independent T-tests indicated numeric aspects such as total clicks, weighted score and days active significantly differentiated between pass and fail groups and Chi-square tests indicated that there is significant relation with categorical other variables like age band and disability (p 1 0.05).The results of the correlation analysis revealed the existence of moderate relationships between engagement features (e.g., total clicks vs. days active), thus the necessity to control the dimensionality. The early SHAP output of an initial XGBoost model pointed to the avg score, submission ratio and interaction diversity as key predictors and informed the further application of feature selection methods to enable reduction redundancy and maximize predictive learning.

3.4 Feature Selection

As feature reduction is known to be the key for stable generalization and reduced computation costs [2], three different strategies of feature selection were systematically implemented:

3.4.1 Particle Swarm Optimization (PSO)

PSO, which is based on social behavior of swarms [35], was adopted as a wrapper feature selection algorithm based on Random Forest as the underlying estimator. Particle Swarm Optimization (PSO) implementation in the proposed study is based on the original contribution of [36], which formalizes the swarm dynamics and convergence behavior in detail. Their theory is the basis of the application of PSO as a feature selection strategy and ensures that the process of evolution in high dimensional educational data is not only interpretable but is also optimized in terms of the F1 score of the classifier.

- Outcome: 72 to 43 most informative variables: The performance has been preserved (F1 $\approx$ 0.937, AUC $\approx$ 0.980).

- Relevance: It is comparable to effective uses of PSO in educational prediction (Begum & Padmanna, 2022; Apriyadi et al., 2022), but further differentiates by making a direct comparison between it and other FS algorithms.

3.5 Predictive Modeling Approaches

3.5.1 Classifiers

Ensemble Modeling is proposed, that combines different machine learning models **Ensemble Voting Classifier**. Rather than relying on just one algorithm, we have combined three models, i.e.Random Forest (RF)is known to be robust and rank features, which is also popular with educational data mining and great at identifying patterns in student backgrounds [23]. Gradient Boosting (GB)Can be used to deal with non-linear interactions that are complex; it usually works better than individual trees and excellent at finding complex, non-linear habits [28]. Artificial Neural Network (ANN) investigated non-linear tree ensemble patterns mimics human-like data processing. By using a **Soft Voting** This approach creates an ensemble of classifiers and uses probability values to predict the final outcome.We have Combined RF, GB, and ANN, that delivered promising results with (F1 $\approx$ 0.938, AUC $\approx$ 0.981). ensemble models uses the predictive power of its classifiers and produces promising results (Masood et al., 2024; Fan et al., 2023).

3.6 Model Evaluation & Setup

To enhance the validation of the results, we used **Stratified 5-Fold Cross-Validation**. This ensures that our Pass, Fail, and Withdrawal groups were evenly tested. We focused on the **F1-Score (0.938)** and **AUC (0.981)** because, in education, it is better to be slightly more cautious than to miss a student who is truly struggling. The TABLE 2 shows the configuration settings of PSO algorithm for feature selection where n_particles are the number of particles in swarm with optimized iterations of 20 the cognitive c1 and social factors c2 are set to 0.3 and 0.5 respectively.

PSO Feature Selection Configuration	
n_particles	15
iterations	20
C1	0.5
C2	0.3
w	0.9
k	5
p	2

Table 2: PSO Parameters

The TABLE 3 shows the hyper parameters of the classification models. The ensemble of these models is evaluated with Stratified KFold ranging from 3 to 5 and random state of 42.

Models	Hyper parameters
Random Forest	n_estimator:100, random_state:42
Gradient Boosting	n_estimator:100, random_state:42
ANN	hidden_layers_size:64,32 max_iter:300 random_state:42

Table 3: Model Hyperparameters

4. ANALYSIS & RESULTS

The evidence of this research describes the predictive accuracy of machine learning models that are trained using the selection of different features. The discussions are made on the results with focus on the prediction strength-feature dimensionality-interpretability balance, which are the major concerns of an educational early warning system. Before the modeling, exploratory analyses were conducted to confirm assumptions regarding feature relevance and to identify some hidden patterns in the dataset. FIGURE 2 illustrates that the dataset had an uneven target distribution with an approximate of 58% as shown in FIGURE 2 ratio of students passing, failing, and withdrawing 27%percent, and 15% Such an imbalance highlights the choice of stratification cross-validation and highlight such metrics as F1 and AUC rather than accuracy. Boxplots

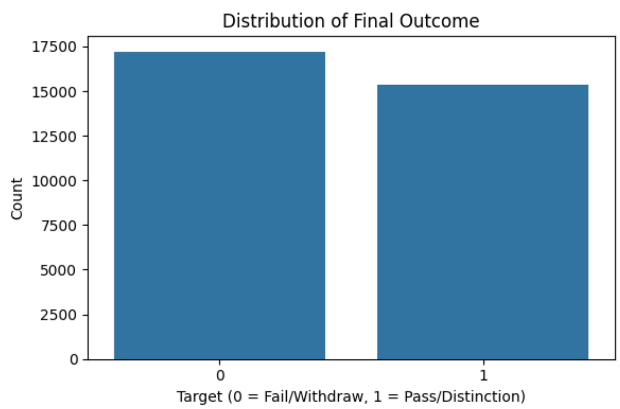

Figure 2: Distribution of students by final outcome

and KDE plots indicated that the failure students had much lower values in major engagement variables including total_clicks, weighted_score, and submission_ratio FIGURE 3.These differences were statistically significant (p ¡ 0.0001) in all the significant numeric features that were independent with T-tests. The chi-square tests revealed significant associations (p -0.05) between categorical characteristics such as age band and disability, as risk patterns are reported by Fan et al. (2023) [31].

Similarly, FIGURE 5 shows distributions for *weighted_score* and *submission_ratio*, highlighting that students with higher interim assessment engagement were significantly more likely to succeed. The core of our research to validate the performance of Particle Swarm Optimization (PSO) and the ensemble.

By applying Particle Swarm Optimization (PSO), the feature set was reduced from 72 to 43 features (approximately 40% reduction), while maintaining nearly identical predictive performance, with accuracy remaining around 94%.The soft voting ensemble model, combining Random Forest (RF), Gradient Boosting (GB), and

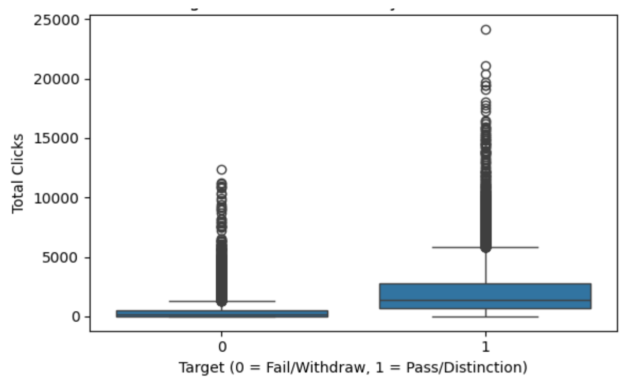

Figure 3: Boxplot of total_clicks across outcome groups

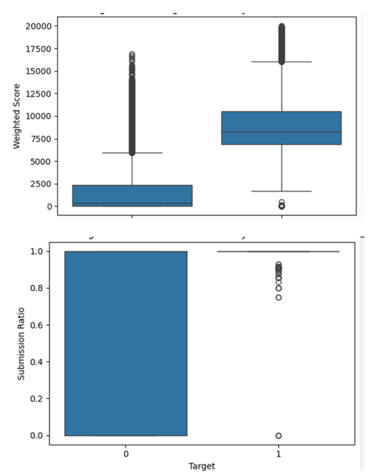

Figure 4: Boxplots for weighted_score and submission_ratio by outcome

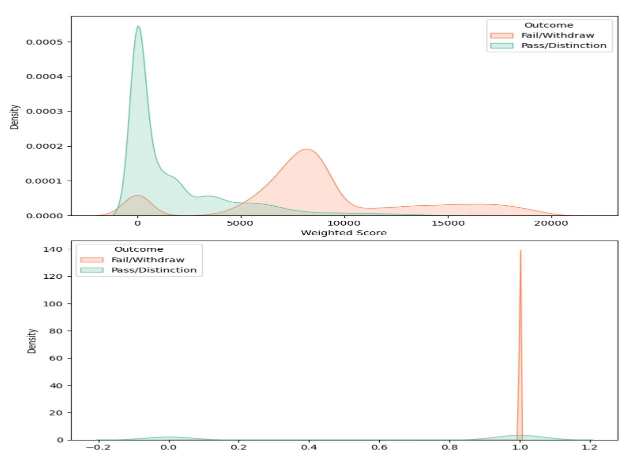

Figure 5: Distributions for *weighted_score* and *submission_ratio*

Table 4: Statistical Significance Tests for Key Features

Feature	Test Type	p-value
total_clicks	T-test	< 0.0001
weighted_score	T-test	< 0.0001
submission_ratio	T-test	< 0.0001
age_band	Chi-square	< 0.05
disability	Chi-square	< 0.05

Table 5: Model Configuration and Performance Comparison

No.	Model Configuration	F1-Score	AUC	Accuracy
1	Ensemble with Full Features	0.938	0.981	0.940
2	Ensemble with PSO (43 Features)	0.937	0.980	0.939

Artificial Neural Networks (ANN), achieved the best overall performance. This level of accuracy is typically associated with complex deep learning black-box models however, the proposed ensemble remains more interpretable and explainable.

4.1 Correlation and SHAP analysis

Correlation heatmaps identified moderate multicollinearity (e.g., total_clicks with days_active), motivating the rigorous feature selection stages that followed.

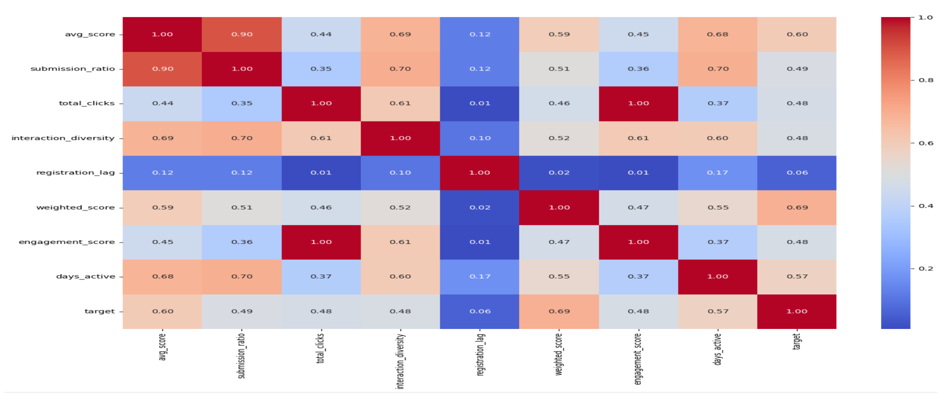

Figure 6: Correlation Heatmap

To analyze numeric interdependencies, a correlation heatmap was created. There were moderate correlations between *total_clicks* and *days_active*, as well as between *avg_score* and *weighted_score* as indicated in FIGURE 6.

In order to enhance model interpretability, SHAP values were included, which is based on the unified explainability framework introduced by Lundberg and Lee (2017) [37]. Their methodology allows trans-

parency at the feature level, where the contribution of each variable to a model prediction is measured, which is especially appropriate with high stakes educational decision-making.
An initial SHAP analysis FIGURE 7 has highlighted avg_score, submission_ratio, total_clicks, and interaction_diversity as the most predictive features, which validates the feasibility of the feature engineering strategy.

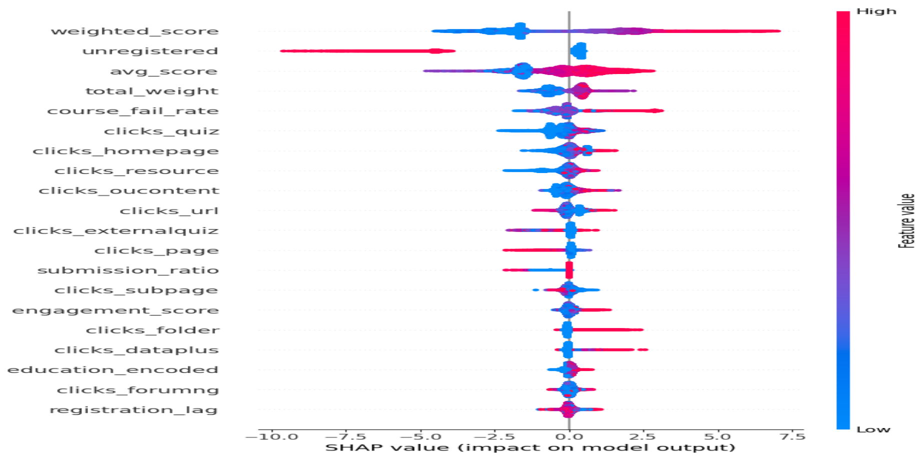

Figure 7: SHAP summary plot

4.2 Model Performance Across Feature Selection Strategies

The results of TABLE 6 show the performance measures of different models trained on different feature sets. Random Forest and Gradient Boosting has recorded the best results, and the use of ensemble also performed a little better than single classifiers.

Table 6: Performance of Models with Different Feature Selection Strategies

Model	Feature Set	F1	AUC	Accuracy
RF	Baseline (72)	0.938	0.981	0.940
RF	PSO (43)	0.937	0.980	0.939
RF	Gini (10)	0.935	0.979	0.937
RF	FAMD (20)	0.865	0.935	0.869
ANN	Baseline (72)	0.916	0.969	0.916
ANN	PSO (43)	0.90	0.96	0.916
ANN	Gini (10)	0.913	0.961	0.915
ANN	FAMD (20)	0.87	0.94	0.88
GB	Baseline (72)	0.936	0.982	0.938
GB	PSO (43)	0.936	0.982	0.938
GB	Gini (10)	0.935	0.982	0.937
GB	FAMD (20)	0.852	0.923	0.846
Ensemble (RF+GB+ANN)	Full	0.938	0.981	0.940

4.2.1 Impact of Feature Selection

- PSO decreased the dimension size by a factor of 40% with virtually no reduction in F1 or AUC, which shows its power to retain important information.
- Gini-based top 10 features provided very competitive results, which implies that the predictive power was very concentrated on a few general features.
- It should be noted that FAMD, though a graceful way of working with mixed data types, has led to a significant decrease in metrics,

4.3 Comparative Analysis with Existing Literature

The pipeline, despite employing classical ML models, achieved metrics on par with more computationally demanding deep learning architectures. Interestingly, in this research multi-strategy FS approach, a trade-off between interpretability was balanced along with predictive power.

Table 7: Accuracy Recent Studies

Study	**Method**	**Dataset**	**Best Accuracy / F1**
Masood et al. (2024) [3]	ECNN + ResNet	OULAD	95.67% acc
Al-Zawqari et al. (2022) [7]	RF / ANN (No Feature Eng.)	OULAD	92.91% acc
Torkhani & Rezgui (2025) [9]	DL Ensemble	OULAD MOOC	~94% acc
Kalita et al. (2025) [38]	LSTM + SHAP	OULAD (disabled learners)	92.8% acc
Present Study	RF, GB, ANN, Ensembles + PSO/Gini/FAMD	OULAD	~94% acc

Overall, the proposed ensemble with PSO based feature selection achieves results comparable to advanced deep architectures while maintaining interpretability.

4.4 Discussion of Key Findings

Following key findings from the research includes,

- PSO proved highly effective, cutting feature space by 40% with negligible performance compromise, corroborating success seen in Begum & Padmanna (2022).
- Gini-based selection demonstrated that most predictive power lay within a tight core of 10 features, enabling streamlined, scalable implementations.
- FAMD, while mathematically appealing, highlighted the risks of excessive compression when class-boundary nuances are critical.

The PSO ensemble methodology used in this research has reached competitive accuracy and interpretability.

5. CONCLUSION

The research has proposed a machine learning framework that is accurate and tansparent in predicting the student performance in the Virtual Learning Environment(VLEs). The Experiments performed on the OULAD dataset indicated that combination of PSO inspired feature selection and ensemble machine learning framework produced reliable and accurate results. We have highlighted the key conclusive points from our study that are listed below

- **Dimensionality Reduction:** The integration of PSO successfully reduced the feature space from 72 to 43 variables while maintaining a 94% accuracy rate.
- **Ensemble Architectures:** The proposed soft voting ensemble showed improved performance in comparison to the individual classifiers, achieving an F1-score of ≈ 0.938 and an AUC of ≈ 0.981.
- **Explainability**The SHAP based explainability provides in depth insights to the decision makers regarding the features that contribute towards the student performance.

This work thus contributes a balanced, transparent, and reproducible predictive framework, addressing notable gaps in existing literature that rarely compare multiple feature selection strategies under consistent experimental conditions [2].
In future we intend to include the temporal based prediction that can utilize data from initial weeks of the course. The temporal models can help in identifying the high risk students well before time .

References

[1] Bisri A, Supardi S, Heryatun Y, Hunainah H, Navira A. Educational Data Mining Model Using Support Vector Machine for Student Academic Performance Evaluation. J Educ Learn. 2025;19:478-486.

[2] Jin L, Wang Y, Song H, So HJ. Predictive Modelling With the Oulad: A Systematic Literature Review. In International Conference on Artificial Intelligence in Education. Cham: Springer Nature Switzerland. 2024:477-484.

[3] Masood JA, Kalyan Chakravarthy NS, Asirvatham D, Marjani M, Abdulkareem Shafiq D, et al. A Hybrid Deep Learning Model to Predict High-Risk Students in Virtual Learning Environments. IEEE Access. 2024;12:103687-103703.

[4] Aljohani NR, Daud A, Abbasi RA, Alowibdi JS, Basheri M, et al. An Integrated Framework for Course Adapted Student Learning Analytics Dashboard. Comput. Hum. Behav. 2019;92:679-690.

[5] Badshah A, Ghani A, Daud A, Jalal A, Bilal M, Crowcroft J. Towards Smart Education Through Internet of Things: A Survey. ACM Comput. Surv. 2023;56:1-33.

[6] Hayat MK, Daud A, Alshdadi AA, Banjar A, Abbasi RA, et al. Towards Deep Learning Prospects: Insights for Social Media Analytics. IEEE Access. 2019;7:36958-36979.

[7] Al-Zawqari A, Peumans D, Vandersteen G. A Flexible Feature Selection Approach for Predicting Students' Academic Performance in Online Courses. Comput. Educ.: Artif. Intell. 2022;3:100103.

[8] Kuzilek J, Hlosta M, Zdrahal Z. Open University Learning Analytics Dataset. Sci. Data. 2017;4:170171.

[9] Torkhani W, Rezgui K. OULAD MOOC Student Performance Prediction Using Machine and Deep Learning Techniques. In Proceedings of International Conference on Decision Aid and Artificial Intelligence (ICODAI 2024). Atlantis Press. 2025;12:228.

[10] Rasheed SS, Arif M, Nawaz H, Zeb MD. Optimizing query processing for scalable educational data analytics: a comprehensive case study with the OULAD dataset. Ann Methodol Arch Res Rev. 2025;3:495-510.

[11] Adnan M, Habib A, Ashraf J, Mussadiq S, Raza AA, et al. Predicting At-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. IEEE Access. 2021;9:7519-7539.

[12] Bilal M, Omar M, Anwar W, Bokhari RH, Choi GS. The Role of Demographic and Academic Features in Student Performance Prediction. Sci Rep. 2022;12:12508.

[13] Jawad K, Shah MA, Tahir M. Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest With Data Balancing. Sustainability. 2022;14:14795.

[14] Iddrisu I, Appiahene P, Appiah O, Fuseini I. Exploring the Impact of Students' Demographic Attributes on Performance Prediction Through Binary Classification in the KDP Model. Knowl Eng Data Sci. 2023;6:24-40.

[15] Adekitan AI, Noma-Osaghae E. Data Mining Approach to Predicting the Performance of First Year Student in a University Using the Admission Requirements. Educ Inf Technol. 2019;24:1527-1543.

[16] Fazil M, Rísquez A, Halpin C. A Novel Deep Learning Model for Student Performance Prediction Using Engagement Data. J Learn Anal. 2024;11:23-41.

[17] Amjad T, Shaheen Z, Daud A. Advanced Learning Analytics: Aspect Based Course Feedback Analysis of MOOC Forums to Facilitate Instructors. IEEE Trans. Comput Soc Syst. 2024;11:4698-4706.

[18] Aggarwal D, Mittal S, Bali V. Significance of Non-Academic Parameters for Predicting Student Performance Using Ensemble Learning Techniques. Int J Syst Dyn Appl. 2021;10:38-49.

[19] Daud A, Aljohani NR, Abbasi RA, Lytras MD, Abbas F, et al. Predicting Student Performance Using Advanced Learning Analytics. In Proceedings of the 26th international conference on world wide web companion. 2017:415-421.

[20] Zhang Y, Hu Y, Gao X, Gong D, Guo Y, et al. An Embedded Vertical-Federated Feature Selection Algorithm Based on Particle Swarm Optimisation. CAAI Trans Intell Technol. 2022;8:734-754.

[21] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3:1157-1182.

[22] Al-Masoudy MA, Al-Azawei A. Proposing a Feature Selection Approach to Predict Learn Ers' Performance in Virtual Learning Environments (Vles). Int J Emerg Technol Learn. 2023;18:110-131.

[23] Kartik N, Mahalakshmi R, Venkatesh KA. Predicting Students' Performance Using Feature Selection-Based Machine Learning Technique. In International Conference on Data Analytics & Management. Singapore: Springer Nature Singapore. 2024:389-397.

[24] Apriyadi MR. Ermatita, Rini DP, Yulnelly: Implementation of Feature Selection Based on Particle Swarm Optimization and Genetic Algorithm on Support Vector Regression Algorithm to Predict Student Performance. In2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS). IEEE. 2022:395-400.

[25] Begum S, Padmanna SS. Student Performance Prediction With BPSO Feature Selection and CNN Classifier. Int J Adv Appl Sci. 2022;9:84-92.

[26] Kumari R, Kumar S. Student's Performance Prediction Using Modified PSO. In 2023 IEEE World Conference on Applied Intelligence and Computing (AIC). IEEE. 2023:1015-1020.

[27] López-Meneses E, López-Catalán L, Pelícano-Piris N, Mellado-Moreno PC. Artificial Intelligence in Educational Data Mining and Human-In-The-Loop Machine Learning and Machine Teaching: Analysis of Scientific Knowledge. Appl Sci. 2025;15:772.

[28] Gajwani J, Chakraborty P. Students' Performance Prediction Using Feature Selection and Supervised Machine Learning Algorithms. In International Conference on Innovative Computing and Communications: Proceedings of ICICC. Singapore: Springer Singapore. 2020:347-354.

[29] Turabieh H. Hybrid Machine Learning Classifiers to Predict Student Performance. In 2019 2nd international conference on new trends in computing sciences (ICTCS). IEEE. 2019:1-6.

[30] Apriyadi MR, Rini DP. Hyperparameter Optimization of Support Vector Regression Algorithm Using Metaheuristic Algorithm for Student Performance Prediction. Int J Adv Comput Sci Appl. 2023;14:144-150.

[31] Fan Z, Gou J, Wang C. Predicting Secondary School Student Performance Using a Double Particle Swarm Optimization-Based Categorical Boosting Model. Eng Appl Artif Intell. 2023;124:106649.

[32] Waheed H, Hassan S, Aljohani NR, Hardman J, Alelyani S, et al. Predicting Academic Performance of Students From VLE Big Data Using Deep Learning Models. Comput Hum Behav. 2020;104:106189.

[33] Hernández-Blanco A, Herrera-Flores B, Tomás D, Navarro-Colorado B. A Systematic Review of Deep Learning Approaches to Educational Data Mining. Complexity. 2019;2019:1-22.

[34] Howard E. Ouladformat R Package: Preparing the Open University Learning Analytics Dataset for Analysis. 2025. arXiv preprint: https://arxiv.org/pdf/2501.08366.

[35] Kennedy J, Eberhart R. Particle Swarm Optimization. In Proceedings of ICNN'95-international conference on neural networks. IEEE. 1995;4:1942-1948.

[36] Poli R, Kennedy J, Blackwell T. Particle Swarm Optimization: An Overview. Swarm Intell. 2007;1:33-57.

[37] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems. NeurIPS. 2017;30:4768-4777.

[38] Kalita E, El Aouifi H, Kukkar A, Hussain S, Ali T, et al. Lstm-Shap Based Academic per Formance Prediction for Disabled Learners in Virtual Learning Environments: A Statistical Analysis Approach. Soc Netw Anal Min. 2025;15:65.