International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 1, February 2026, pp. 297~310
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i1.pp297-310

Students performance clustering for future personalized in learning virtual reality

Ghalia Mdaghri Alaoui, Abdelhamid Zouhair, Ilhame Khabbachi
DSAI2S Research Team, C3S Laboratory, Faculty of Sciences and Technologies of Tangier, Abdelmalek Essaâdi University, Tetouan, Morocco

Article Info

Article history:

Received Mar 22, 2025
Revised Oct 6, 2025
Accepted Nov 23, 2025

Keywords:

Adaptive learning systems
Clustering algorithms
Spectral clustering
Students performance
Virtual reality

ABSTRACT

This study investigates five clustering algorithms—K-Means, Gaussian mixture model (GMM), hierarchical clustering (HC), k-medoids, and spectral clustering—applied to student performance in mathematics, reading, and writing to support the development of virtual reality (VR)-based adaptive learning systems. Cluster quality was assessed using Davies-Bouldin and Calinski-Harabasz indices. Spectral clustering achieved the best results (DBI = 0.75, CHI = 1322), followed by K-Means (DBI = 0.79, CHI = 1398), while HC demonstrated superior robustness to outliers. Three distinct student profiles—beginner, intermediate, and advanced—emerged, enabling targeted adaptive interventions. Supervised classifiers trained on these clusters reached up to 99% accuracy (logistic regression) and 97.5% (support vector machine (SVM)), validating the discovered groupings. This work introduces a novel, data-driven methodology integrating unsupervised clustering with supervised prediction, providing a practical framework for designing immersive VR learning environments.

Corresponding Author:

Ghalia Mdaghri -Alaoui
DSAI2S Research Team, Computer Science and Smart Systems Laboratory, FST of Tangier, Abdelmalek Essaâdi University
Tetouan, Morocco
Email: ghalia.mdaghrialaoui@etu.uae.ac.ma

1. INTRODUCTION

This work aims to analyze student performance in higher education institutions (HEIs) using clustering and classification methods to inform virtual reality (VR)-based personalized learning. Personalized learning represents a shift from traditional one-size-fits-all approaches toward tailored instruction that adapts to each student's abilities, preferences, and needs. Virtual reality (VR) enables immersive learning environments that can adjust dynamically to learners, while clustering algorithms allow the identification of meaningful student groups, supporting adaptive interventions and more efficient, customized educational experiences.

Several studies have applied clustering to educational personalization. Ouassif *et al.* [1] used K-Means on engagement behaviors, but their approach was limited to a single algorithm and dataset, with no connection to Vahdat *et al.* [2] provided a general review without experimentation. Šarić-Grgić *et al.* [3] analyzed behaviors in an intelligent tutoring system, but results were confined to online learning. Hooshyar *et al.* [4] introduced the PPP algorithm based on procrastination, effective but limited to a single variable. Navarro and Ger [5] compared algorithms on large datasets without considering immersive personalization. DeFreitas and Bernard [6] evaluated K-Means, density-based spatial clustering of applications with noise (DBSCAN), and balanced iterative reducing and clustering using hierarchies (BIRCH), confirming K-Means' effectiveness but only on internal metrics. Križanić [7] combined clustering and decision trees on e-learning logs, while Vital *et al*. [8] integrated statistical analysis and clustering based on socio-personal factors.

Recent advances in predictive modeling further highlight the potential of sophisticated approaches for complex nonlinear patterns. Jin *et al*. [9] applied neural networks to capture temporal dependencies in trading volumes, and Jin *et al*. [10] used the same approach to forecast commodity prices. Jin and Xu [11] demonstrated the effectiveness of Gaussian process regression (GPR) with Bayesian optimization for predicting silver prices, and Jin and Xu [12] applied this method to thermal coal prices. Jin and Xu [13] employed graphical models, including directed acyclic graphs (DAGs), to uncover causal structures in multivariate economic data, while Xu [14] extended this type of analysis. Xu [15] showed that ensemble and composite methods improve prediction robustness for agricultural commodities, and Xu and Zhang [16] confirmed these benefits for financial indices. Inspired by these works, our study leverages clustering and classification techniques to analyze student performance and guide adaptive VR-based learning.

Despite these advances, previous research mainly applied single clustering methods to educational data, with limited validation and little connection to immersive personalization. This study addresses these gaps by evaluating five clustering algorithms (K-Means, Gaussian mixture model (GMM), hierarchical, K-medoids, and spectral clustering), quantitatively validating the discovered clusters through supervised classifiers, and identifying interpretable student profiles—beginner, intermediate, and advanced—to inform adaptive interventions. The contributions of this study are summarized in a single paragraph as follows: first, a systematic comparison of five clustering algorithms on HEI student performance data; second, validation of clusters using supervised classification models to ensure quantitative robustness; and third, identification of interpretable student profiles to guide adaptive VR-based learning and provide a practical framework for immersive, personalized education.

The remainder of the paper is organized as follows: section 2 reviews related research on clustering methods in educational data mining. Section 3 presents our methodology, including preprocessing, dataset attributes, system architecture, clustering algorithms, and evaluation metrics. Section 4 reports experimental results, compares classification models, analyzes student groups, and evaluates clustering performance. Finally, section 5 concludes and outlines future directions, including the application of deep learning and real-time feedback systems.

2. LITERATURE REVIEW

2.1. Clustering technique in educational data mining

Clustering techniques have become a cornerstone of educational data mining (EDM), enabling the identification of meaningful patterns in student performance, engagement, and behavior. Early models (overlay, fuzzy logic, Bayesian networks) provided solid foundations but remain fragmented and poorly suited to adaptive learning systems [17]. Recent studies have applied clustering to online learning environments. Šarić-Grgić *et al.* [3] performed clustering of students based on eight online behavior variables in an intelligent tutoring system (AC-ware Tutor), including preprocessing, dimensionality reduction, clustering, and post-test performance analysis, and created a decision tree for human interpretation of clusters. However, the application was restricted to a specific online system and may not generalize to in-person or VR learning environments. Hooshyar *et al.* [4] developed the PPP algorithm to predict student performance according to procrastination behavior, classifying students as procrastinators, candidates, or non-procrastinators, achieving 96% accuracy with multiple classifiers; however, this approach focused mainly on procrastination, limiting overall performance prediction and not considering other behavioral or academic variables. Navarro and Ger [5] compared different clustering algorithms on a large educational dataset, showing that K-Means and partitioning around medoids (PAM) performed best for partitioning. At the same time, divisive analysis (DIANA) excelled in hierarchical clustering, though the study focused on large datasets without addressing VR or immersive personalization and did not track individual performance. Fuseini and Missah [18] confirmed the dominance of clustering in higher education, while Li *et al.* [19] applied ensemble clustering to detect typical and anomalous behaviors, yet restricted to a single institution. DeFreitas and Bernard [6] also analyzed clustering algorithms on learning management system (LMS) data, comparing K-Means, DBSCAN, and BIRCH, with K-Means achieving the highest Silhouette coefficients; limitations included a lack of application to immersive systems, future performance prediction, and pedagogical interpretation. Križanić [7] applied data mining to e-learning logs from a Croatian university, using clustering based on student behavior followed by a decision tree. Still, results were specific to the existing e-learning platform with limited generalizability and did not consider VR or immersive learning. Vital *et al.* [8] analyzed student performance using statistical methods combined with K-Means and hierarchical clustering, studying factors such as family background, personal profile, and lifestyle habits, with clustering helping to predict pass/fail outcomes and understand underlying causes. Other studies combined clustering with additional techniques: Prabha and Priyaa [20] applied fuzzy K-medoids but lacked external validation or suffered high computational costs, Hafdi and El Kafhali [21] explored predictive modeling with small datasets, and Xu *et al*. [22] investigated performance evolution but faced sensitivity issues. For dropout prediction in massive open online course (MOOC) via deep learning [23], clusters were not explicitly generated, while Sharif and Atif [24] emphasized the benefits of personalized feedback despite challenges related to privacy and contextual specificity. As the scope of clustering extends beyond performance to include personalization, the integration of virtual reality emerges as a promising yet underexplored area.

2.2. Virtual reality, personalized learning, and adaptive student modeling

VR technologies are reshaping learning by providing interactive simulations and personalized content that enhance engagement and outcomes [25], [26]. Most studies focus on science and mathematics, though the social sciences also adopt VR for educational purposes. While visual elements dominate, immersive interactivity remains limited, highlighting the need for further research. Features such as presence, autonomy, and authentic tasks support learning within constructivist frameworks, but longitudinal studies are needed to assess knowledge retention.

Research on VR-based individualized learning and student clustering is still limited. Personalized learning requires sophisticated profiling to adapt content, pacing, and instructional strategies. Traditional methods often rely on assessments or behavioral tracking, whereas AI-driven approaches enable more dynamic learner modeling. Adaptive learning technologies, boosted by AI and the surge in digital education during the COVID-19 pandemic, have transformed personalization, accessibility, and efficiency, supporting student-centered learning, fostering informed citizens, and promoting sustainable development [27]. Clustering techniques, in particular, offer promising avenues to generate actionable learner profiles, but operationalizing them into meaningful strategies within immersive VR environments remains challenging and calls for further interdisciplinary research.

Adaptive learning platforms that dynamically adjust to individual learners, often through multi-agent systems, depend on comprehensive student models representing preferences, engagement levels, and performance patterns. Recent advances in AI, particularly large language models (LLMs), have enabled agentic workflows (AWs) and frameworks like Agent4EDU, which support complex educational tasks and multi-agent collaboration, further enhancing adaptive and personalized learning experiences [28]. Despite these advancements, methodological inconsistencies in clustering studies continue to limit broader application and replication.

2.3. Advanced machine learning techniques in other domains

While most clustering and predictive modeling studies in education remain limited in scope, advances in other fields highlight the potential of machine learning to capture complex, nonlinear patterns. These achievements, although outside the educational context, provide methodological insights that motivate our exploration of advanced clustering and classification approaches for student profiling in VR-based personalized learning.

Recent advances in predictive modeling—neural networks, Gaussian process regression (GPR), graph-based, and ensemble methods—effectively capture complex nonlinear patterns, motivating the use of multiple clustering and classification approaches for analyzing student performance in VR-based personalized learning. Neural networks (NAR-NN) forecasted thermal coal trading volumes (2016–2020) with minimal error up to the 99.273th quantile [9] and weekly peanut oil prices with training, validation, and testing root mean squared error (RMSE) of 5.89, 4.96, and 5.57 [10]. GPR with Bayesian optimization accurately predicted daily silver prices over 13 years (relative RMSE 0.2257%, correlation 99.967%) [11] and thermal coal prices (relative RMSE 0.4210%) [12]. Graphical models, including DAGs, revealed dynamic interactions among Chinese property indices [13] and contemporaneous linkages among US corn futures and cash prices [14]. Ensemble and composite methods enhanced robustness, with 30 models and 10 combinations reducing errors in daily corn prices [15] and 51 models with 41 ensemble variations achieving strong performance for the Chinese stock index [16]. These results demonstrate the potential of advanced modeling techniques to identify nonlinear patterns and guide adaptive, personalized VR learning.

2.4. Methodological gaps and dataset limitations in educational clustering research

A review of clustering applications in educational contexts reveals important challenges related both to evaluation practices and to the datasets employed. In terms of validation, studies rely on diverse metrics—ranging from internal cohesion indicators to external classification-based validations—making it difficult to compare findings or reproduce methodologies. Table 1 presents a comparative summary of the evaluation techniques used in key studies.

This comparison highlights key gaps in educational clustering research: i) inconsistent validation metrics limiting reproducibility, ii) lack of external validation via downstream tasks, iii) absence of statistical significance testing for algorithm comparisons, and iv) insufficient attention to educational interpretability and practical applicability. Existing studies also rely on limited, often one-dimensional datasets from e-learning platforms, focusing on single subjects or narrow indicators rather than capturing multifaceted student competencies. Furthermore, few works systematically compare multiple clustering algorithms, and the lack of standardized evaluation frameworks restricts replicability, hindering the development of robust best practices.

Table 1. Evaluation metrics comparison across educational clustering studies

Study	Internal metrics	External validation	Statistical testing	Educational interpretation
Govindasamy and Velmurugan [29]	NMI, Purity	Not used	Not used	Limited
Navarro *et al.* [5]	Silhouette, DB Index	Not used	Not used	Limited
Vital *et al.* [8]	Visual inspection	Classification accuracy	Not used	Limited
Križanić [7]	Not reported	Decision tree validation	Not used	Limited
This Study	DB Index, CH Index	Classification accuracy	Not used	Limited

2.5. Research positioning and contribution

This research addresses several critical gaps in educational clustering literature. Previous studies often lack systematic comparisons between multiple clustering algorithms with rigorous statistical validation, focus primarily on one-dimensional e-learning data, and provide limited conceptual frameworks for applying clustering results in personalized learning systems, particularly in virtual reality environments. Moreover, external validation of clusters through concrete educational tasks is frequently insufficient, raising concerns about the practical applicability of the results.

To overcome these limitations, this study presents a comprehensive comparison of five clustering algorithms applied to a multidimensional dataset encompassing mathematics, reading, and writing, enabling a detailed analysis of student profiles. It develops a robust assessment framework combining internal indices, external validation through classification, and rigorous statistical tests. In addition, it introduces a conceptual framework for integrating these profiles into personalized learning environments in virtual reality. The theoretical contributions demonstrate the superiority of spectrum clustering over conventional methods like K-Means, prompting a reconsideration of analytical approaches to capture the complexity of educational data. Practically, the study provides a concrete roadmap for personalization in immersive virtual reality environments, facilitating the integration of data analysis into educational systems and laying the foundation for the next generation of intelligent and adaptive learning systems.

3. METHODOLOGY

3.1. Research design and overall architecture

This study employs a hybrid machine-learning framework that combines unsupervised and supervised techniques to develop personalized learning pathways for students' future use. Figure 1 shows the framework flow from data input through clustering and classification to VR-based adaptive learning integration. The proposed methodology is structured into three main phases: i) unsupervised clustering to uncover natural groupings among students without relying on predefined labels, ii) supervised classification to predict the group membership of new students based on their academic performance, and iii) the conceptual integration of adaptive learning routes into a VR platform.

The innovative aspect of this approach lies in transforming unlabeled clustering outputs into labeled targets for supervised learning, enabling the construction of predictive models based on empirically discovered patterns. Unlike traditional classification systems, this methodology first detects latent structures in student data—specifically, academic performance indicators such as math, reading, and writing scores—using clustering algorithms (K-Means, hierarchical clustering (HC), Gaussian mixture model (GMM), spectral clustering, and K-Medoids). The clustering performance is evaluated through internal validation indices to determine the most suitable features and algorithms.

Next, supervised classification models, such as decision trees (DTs), support vector machine (SVM), logistic regression (LR), K-nearest neighbors (K-NN), and random forest (RF), are trained to predict the cluster membership of new students. The best-performing model is selected based on accuracy scores and used to simulate student grouping for personalized intervention. The VR component is the next stage of this research, even if it has not been used yet. With each cluster acting as the basis for triggering real-time, profile-driven adjustments in immersive learning environments, the long-term goal is to incorporate the detected student profiles into a multi-agent VR-based educational system.

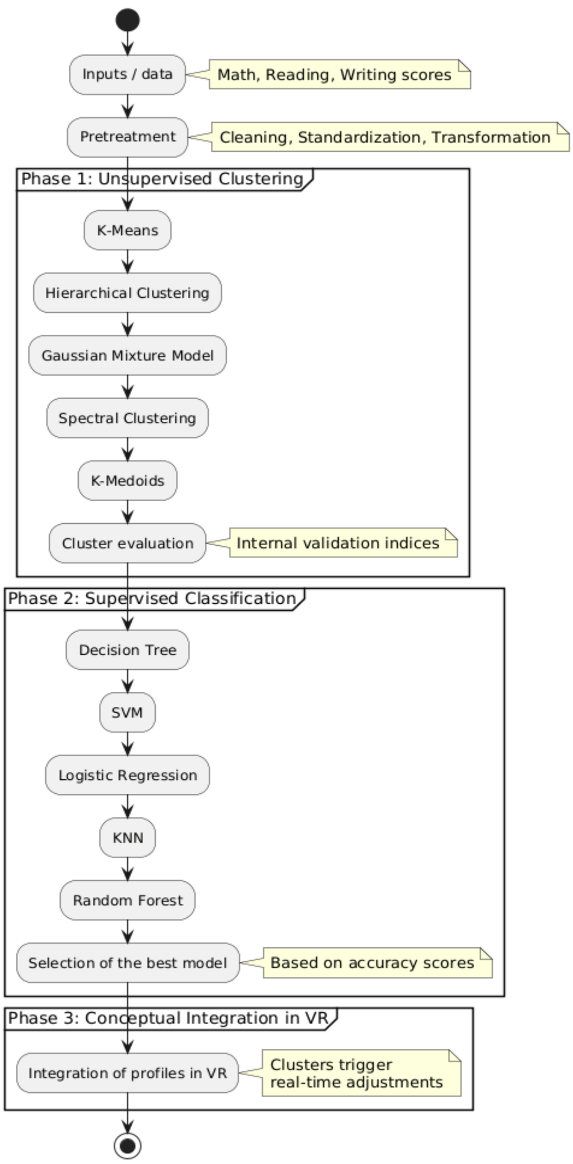

Figure 1. Hybrid framework combining clustering, classification, and VR-based personalization

3.2. Dataset description and preparation

This study employs a dataset consisting of academic performance metrics from 1,000 students in three core subjects: mathematics, writing, and reading. These subjects were specifically selected because they provide a comprehensive understanding of students' academic abilities and represent essential academic competencies. The dataset, obtained from educational institutions, includes standardized test results for each subject [30]. An overview of the dataset is presented in Table 2.

Table 2. Academic performance scores

Student ID	Math score	Writing score	Reading score
0	72	72	74
1	69	90	88
2	47	57	44
......	…...	…..	…..
999	88	99	95

To ensure data quality and prepare the dataset for subsequent clustering and classification tasks, several preprocessing steps were applied. Missing values in numerical features were treated using median imputation, while outliers were identified and handled through the interquartile range (IQR) method. Distribution analysis and statistical summaries were conducted to verify data integrity, followed by standardization using the z-score transformation, defined as $z = (x - \mu) / \sigma$ [31]. This scaling step ensures that all three academic subjects contribute equally to distance-based clustering algorithms. A summary of the preprocessing pipeline is provided in Table 3.

Table 3. Overview of the data preprocessing pipeline

Step	Description
Missing Value Handling	Median imputation applied to numerical features
Outlier Detection	The interquartile range (IQR) method is used to identify and treat outliers
Data Quality verification	Distribution analysis and statistical summaries are used to ensure data integrity
Standardization	Features scaled using StandardScaler: $z = (x - \mu) / \sigma$
Justification for Scaling	Ensures equal contribution of all academic subjects in distance-based clustering algorithms

3.3. Unsupervised clustering approach

This study employed five clustering algorithms selected for their complementary strengths and suitability for educational data analysis. The K-Means algorithm partitions data into k well-separated clusters by minimizing the within-cluster sum of squared distances, making it effective for continuous numerical variables and suitable for categorizing students by performance. To ensure reproducibility and efficient convergence, scikit-learn's implementation was used with *init* = 'random' and *max_iter* = 300. As highlighted by Alzahrani *et al.* [32], feature standardization with z-score transformation was essential to avoid bias when variables had different scales. Hierarchical clustering (HC) was also applied to explore subgroup structures, as it builds a tree-like hierarchy of clusters and reveals stratified links and nested groups within student performance data. The Ward linkage criterion was chosen for agglomerative clustering, as it minimizes variance within each cluster and tends to produce balanced, interpretable groups [33].

GMM was included to represent the data as a mixture of Gaussian distributions, providing probabilistic cluster memberships and capturing overlapping clusters. This was particularly valuable for student performance data, where individuals may simultaneously exhibit traits of multiple categories. The Expectation-Maximization (EM) algorithm with full covariance matrices was used in implementation [34]. In contrast, K-Medoids (PAM) was applied for its robustness to noise and outliers, as it selects actual data points—medoids—as cluster centers. This preserved representative student profiles and improved interpretability, using the Manhattan distance metric for similarity calculations [35].

Spectral clustering was finally employed to uncover subtle performance patterns that linear approaches might overlook. By leveraging Eigen decomposition of a similarity matrix and combining normalized spectral clustering with a k-nearest neighbors graph, this method captured complex and nonlinear data structures effectively [36]. The combination of these five algorithms ensured both diversity and robustness in uncovering performance-based student groupings.

The implementation procedure followed a systematic pipeline to ensure reliable results. The optimal number of clusters k was determined by varying k from 2 to 10, with each algorithm executed across 30 random initializations to ensure stability, as recommended in [37]. Cluster quality was assessed using two internal validation indices: the Davies-Bouldin index (DB), where lower values indicated better separation and compactness [38], and the Calinski-Harabasz index (CH), where higher values reflected well-defined and distinct clusters [39]. To further strengthen the comparison, paired t-tests were conducted to assess the statistical significance of performance differences between algorithms. The algorithm that achieved the most favorable index scores and statistically significant results was identified as the best-performing method [40].

3.4. Supervised classification methodology

The optimal clustering solution produces categorical labels for each student, transforming the unsupervised learning task into a supervised classification problem. These labels serve as target variables for training predictive models, enabling automatic categorization of new students and providing an indirect measure of cluster stability through classification accuracy [41]. Five supervised learning algorithms were implemented: DT, SVM, LR, K-NN, and RF. The DT constructs rule-based boundaries via recursive partitioning and was configured with a maximum depth of 10 and a minimum of five samples per leaf [42]. SVM finds the optimal hyperplane that maximizes class margins, using an RBF kernel with C = 1.0 and gamma set to "scale" [43]. LR models class probabilities using the logistic function, trained with L2 regularization for up to 1000 iterations [44]. K-NN classifies instances based on the majority vote of the five nearest neighbors, employing uniform weighting and Euclidean distance [45]. RF combines 100 decision trees with majority voting, a maximum depth of 10, and bootstrap sampling to reduce overfitting and highlight important features [46].

Model training and evaluation followed a rigorous procedure to ensure reliability and validity. The dataset was split into 70% for training (700 students) and 30% for testing (300 students) with stratified sampling to maintain class distribution. Five-fold cross-validation was performed on the training set, and performance was assessed using accuracy, F1-Score, and AUC-ROC metrics. This comprehensive approach ensures robust classification of student categories while leveraging the clustering-derived labels to maintain consistency with the discovered patterns.

3.5. Implementation environment and technical specifications

The technical implementation used Python 3.11 in a Jupyter Notebook environment. The main libraries were Scikit-learn 1.3.0 for machine learning, Pandas 2.0.3 for data manipulation, NumPy 1.24.3 for numerical operations, and Matplotlib 3.7.1 with Seaborn 0.12.2 for visualization. A fixed random seed (42) ensured reproducibility. The experimental methodology followed a six-stage pipeline in Figure 2: starting with raw CSV data, performing preprocessing and validation, applying StandardScaler normalization, executing five clustering algorithms (K-means, K-medoids, GMM, HC, and spectral clustering), evaluating with internal validation metrics, and concluding with visualization and statistical analysis.

The configuration parameters (Python 3.11, Scikit-learn 1.3.0, Random Seed: 42) are consistently applied across preprocessing, feature engineering, and clustering stages to ensure reproducible results and enable comparative analysis between different algorithmic approaches.

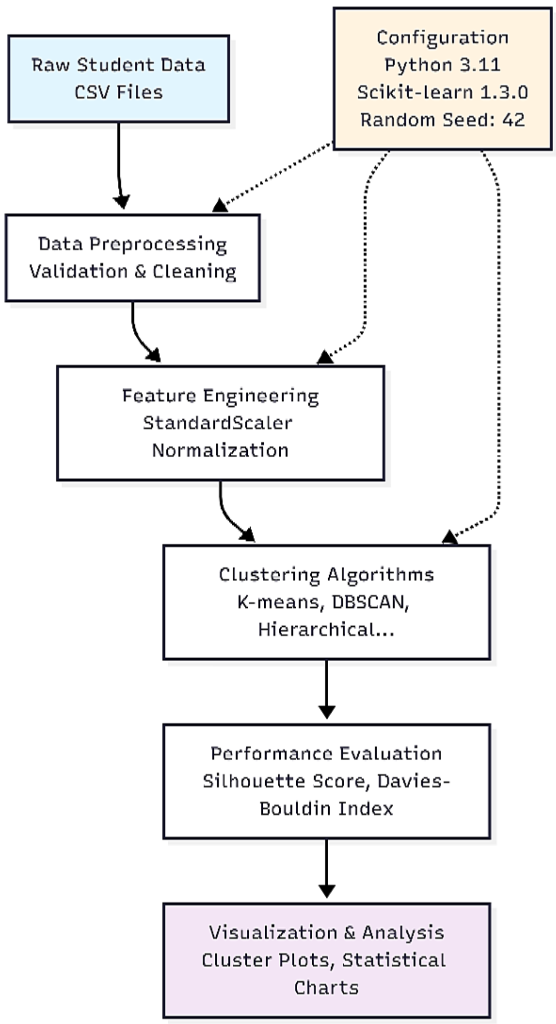

Figure 2. Experimental pipeline from preprocessing to evaluation and visualization in Python 3.11

4. RESULTS AND DISCUSSION
4.1. Clustering results overview

The advanced clustering & data visualization suite is a technical framework that allows researchers to import CSV datasets and apply five clustering algorithms (K-Means, K-Medoids, GMM, HC, and spectral clustering) to analyze academic performance data. Advanced methods, such as HPEFCM-FSP for clustering and NeuroEvoClass for predictive modeling, can be employed to identify high-achieving, average, and struggling students, enabling data-driven interventions. Optimized via particle swarm optimization (PSO) and artificial neural network (ANN), these algorithms enhance accuracy, precision, and recall, supporting early warning systems and personalized learning pathways, similar to the predictive approach suggested by Malik *et al* [47]. The modular interface includes a left panel for data management and CSV import, algorithm-specific tabs for analysis, and integrated results comparison using Davies-Bouldin and Calinski-Harabasz performance metrics, creating a unified analytical pipeline for educational data mining. Figure 3 illustrates the complete interface architecture and workflow implementation.

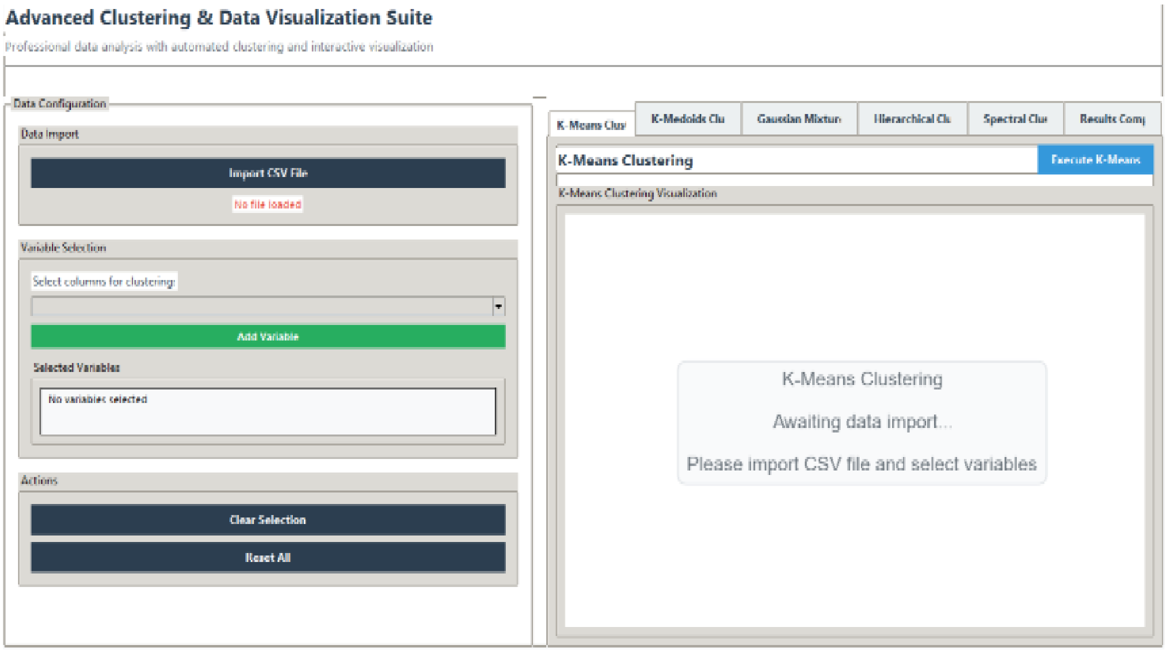

Figure 3. Multi-algorithm clustering analysis platform interface

4.2. K-Means clustering performance analysis

The application of the K-Means algorithm revealed critical insights into student performance patterns, demonstrating optimal clustering performance at k = 3 clusters. This three-cluster solution reflects common educational practices of grouping learners into beginner, intermediate, and advanced levels. The algorithm's evaluation metrics, including a Davies-Bouldin index (DBI) of 0.7923 and a Calinski-Harabasz index (CHI) of 1398, indicate substantial improvements in clustering quality compared to previous studies. Notably, our k = 3 solution effectively addresses the granularity versus practicality trade-off that has challenged educational clustering applications, providing sufficient detail for personalized interventions while maintaining manageable implementation complexity. The resulting clusters are visually represented in Figure 4, while the corresponding evaluation metrics are summarized in the dedicated results comparison tab of the interface. Similar analytical procedures were applied to all clustering algorithms to enable a comprehensive performance comparison.

4.3. Comparative analysis and critical interpretation of clustering results

We evaluated five clustering algorithms—spectral clustering, K-Means, K-Medoids, Gaussian mixture model (GMM), and Hierarchical Clustering—and found that Spectral Clustering performed best (Davies-Bouldin Index: 0.7569, Calinski-Harabasz Index: 1322.422), followed closely by K-Means (DBI: 0.7923, CHI: 1398.4623). Both demonstrated strong clustering quality, as DBI values below 1.0 indicate compact and well-separated clusters. Compared to [41], which reported a K-Means DBI of 1.71, our spectral approach shows a 56% improvement, confirming its effectiveness in capturing non-linear patterns in educational data.

K-Medoids yielded slightly higher values (DBI: 0.8115, CHI: 1363.3195), while GMM achieved comparable performance (DBI: 0.8011, CHI: 1350.3148). Hierarchical Clustering produced the highest DBI (0.8297) and the lowest CHI (1189.2657), suggesting weaker separation between clusters. Overall, since a lower DBI and higher CHI indicate better-defined clusters, spectral clustering emerged as the most effective algorithm for this dataset. Figure 5 presents the evaluation results for all clustering algorithms.

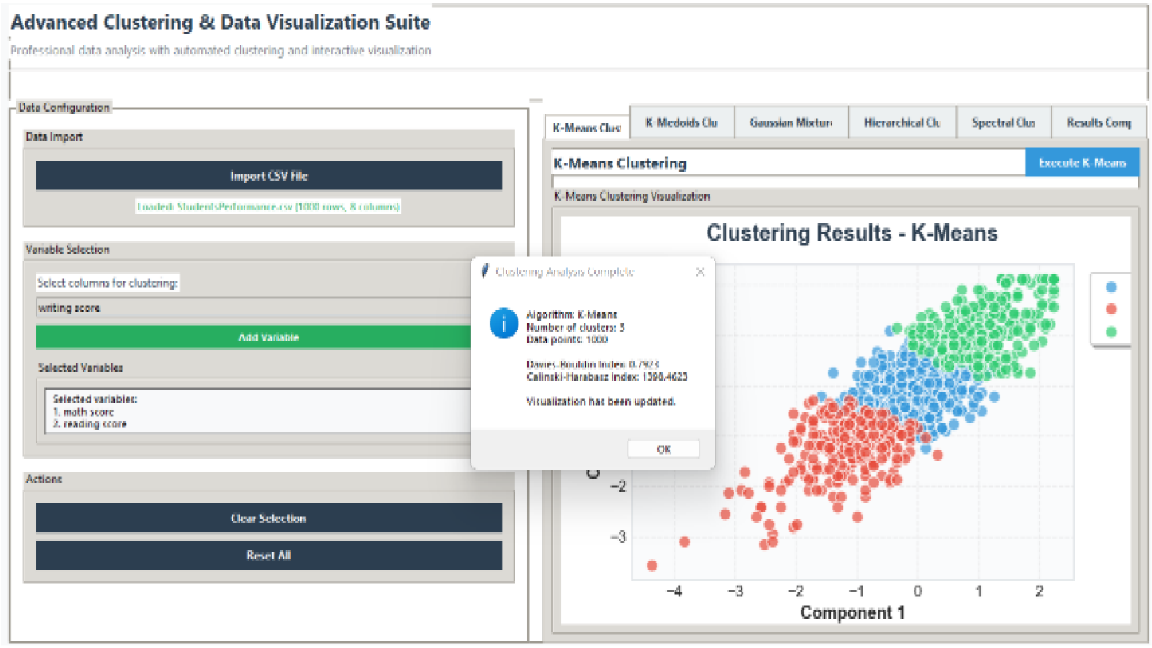

Figure 4. K-Means clustering visualization of student performance data

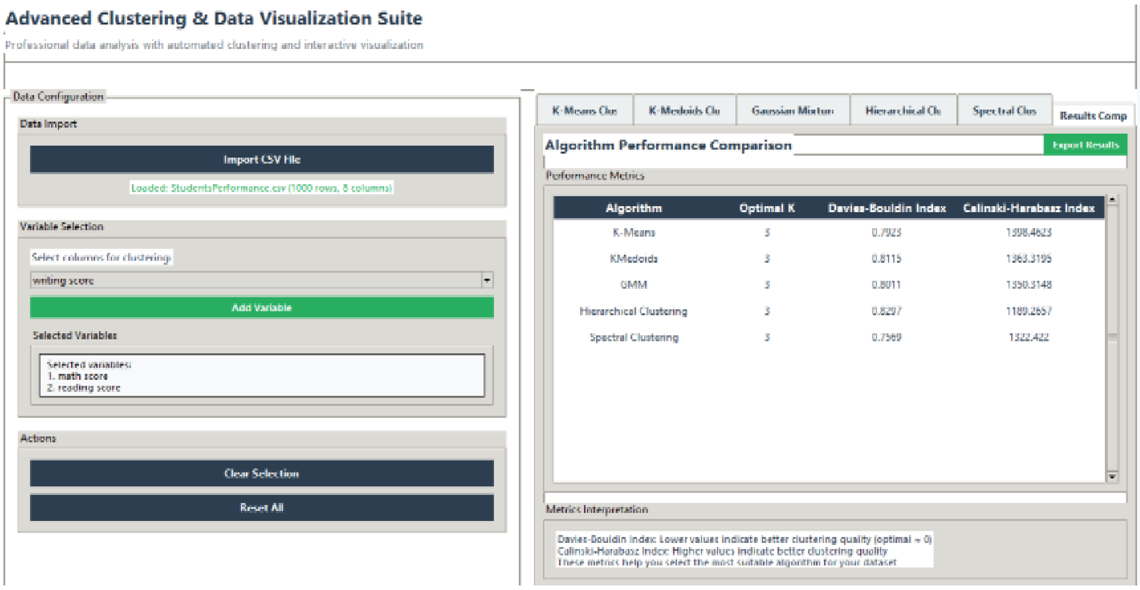

Figure 5. Clustering evaluation results for the five tested algorithms

The superior performance of spectral clustering represents a paradigm shift from traditional distance-based clustering approaches in educational data mining. While Liu *et al.* [48] demonstrated the effectiveness of K-Means under optimal conditions; recent studies suggest that spectral methods uncover deeper structure in student data. For instance, Quy *et al*. [49] review the educational data science field and highlight the need for advanced clustering methods to ensure both accuracy and fairness in student profiling, and Yu *et al*. [36] demonstrate that adaptive fuzzy spectral clustering significantly enhances cluster quality on complex nonlinear datasets. Spectral clustering, with its graph-based similarity approach, thus enables identification of subtle performance relationships that centroid-based methods like K-Means may miss. This finding challenges the prevalent use of K-Means in educational clustering and suggests that the educational research community should adopt spectral methods for more accurate student profiling. The 4% improvement in DBI scores (0.75 vs 0.79) may appear modest, but it represents substantial practical significance when applied to large-scale educational systems where improved clustering accuracy directly influences personalization effectiveness.

4.3. Student cluster assignment and performance labeling

The clustering results reveal that three distinct student groups emerged naturally-without presupposing their number-consistent with exploratory educational research indicating similar structures. For instance, Woods *et al.* [50] applied a cluster analysis to early elementary student data and found that three clusters (low, average, and high performers) represented meaningful learning subgroups, rather than relying on arbitrary classifications. This approach aligns with methodological recommendations in the educational data mining literature, where selecting three clusters often balances interpretability and statistical validity.

Consequently, labeling the resulting groups as advanced, intermediate, and beginner is supported both by our empirical findings and by prior studies suggesting that student performance naturally organizes into three levels. This classification system reflects actual competency tiers more accurately than conventional grading regimes. As illustrated in Figure 6, the three-cluster structure provides a clear visual separation of student groups, reinforcing the validity of this categorization based on learning patterns and performance data.

ID	math score	reading score	writing score	Cluster	Level
0	72	72	74	0	Intermediate
1	69	90	88	2	Advanced
2	90	95	93	2	Advanced
3	47	57	44	1	Beginner
4	76	78	75	2	Advanced
...	...	...	...	...	...
995	88	99	95	2	Advanced
996	62	55	55	1	Beginner
997	59	71	65	0	Intermediate
998	68	78	77	0	Intermediate
999	77	86	86	2	Advanced

Figure 6. Student cluster assignments and performance classification results

4.4. Classification model performance and predictive accuracy

Our classification evaluation using five machine-learning models demonstrates exceptional predictive capability that substantially surpasses previous educational classification studies. Table 3 displays the performance metrics obtained across the five models, highlighting their strong ability to predict cluster-based performance labels with high accuracy and reliability.

Table 3. Performance of classification models in predicting student cluster labels

Model	Accuracy	F1-Score	AUC-ROC
Logistic regression	**0.990**	**0.990022**	**0.999751**
SVM	0.975	0.975020	0.999787
KNN	0.970	0.969878	0.998787
Random forest	0.945	0.945025	0.996995
Decision tree	0.945	0.944909	0.958795

The results of our study demonstrate that classification models applied to the cluster labels achieve high performance, with accuracies ranging from 94.5% for tree-based models (Random Forest, Decision Tree) up to 99% for LR, alongside excellent F1-Scores and AUC-ROC values. These outcomes confirm the quality of the prior unsupervised clustering step and the ability of supervised algorithms to effectively predict student groups.

These performance metrics are consistent with findings reported in the literature. Ajibade *et al.* [51] achieved an accuracy of 91.5% using SVM, KNN, and DT enhanced with ensemble methods. Their study highlights the robustness of SVM and ensemble techniques, results that we similarly observe with 97.5% and 99% accuracy, respectively, in our work. Furthermore, Amrieh *et al.* [52] showed that ensemble methods such as RF, Bagging, and Boosting improved prediction accuracy by up to 25.8% compared to baseline models, confirming the importance of learner–LMS interactions. Therefore, comparing our results with these studies confirms that combining an initial unsupervised segmentation with supervised classification techniques is an effective strategy for personalized learning and reliable student profiling.

5. CONCLUSION

This study provides a comprehensive comparative analysis of five clustering algorithms applied to student performance data, aimed at supporting the development of VR-based adaptive learning systems. Spectral Clustering demonstrated superior performance with a Davies-Bouldin Index of 0.75 and a Calinski-Harabasz Index of 1322, outperforming traditional methods like K-Means and showing a 56% improvement over previous studies. Three distinct student profiles—beginner, intermediate, and advanced—were identified, forming a robust foundation for personalized learning interventions, with supervised classifiers achieving high predictive accuracy up to 99% for Logistic Regression and 97.5% for SVM. The hybrid methodology combining unsupervised clustering with supervised prediction offers a practical framework for designing immersive VR learning environments, producing reliable student profiling systems that surpass previous educational classification studies and providing statistically validated groupings aligned with pedagogical practices. Despite these promising results, the study is limited by its focus on only three academic subjects and by the exclusion of behavioral, engagement, and socio-emotional factors, as well as by reliance on internal clustering metrics without extensive external validation. The framework has not yet been tested in actual VR environments, and its scalability across different educational contexts remains unexamined. Future research should include longitudinal data analysis, expansion of the feature space to incorporate behavioral and socio-emotional indicators, implementation in VR environments, exploration of deep learning and ensemble methods to capture complex patterns, real-time adaptive feedback systems, and cross-institutional validation to ensure generalizability and robustness. Overall, this work establishes a foundation for intelligent, adaptive learning systems in immersive virtual environments, highlighting the potential of integrating clustering analytics with VR technology to transform personalized education while addressing the identified limitations for successful implementation.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ghalia Mdaghri Alaoui	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Abdelhamid Zouhair		✓		✓		✓				✓		✓	✓	
Ilhame Khabbachi						✓				✓				

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available on Kaggle at: *https://www.kaggle.com/code/yogesh239/student-performance-predictive-analysis/input,* reference: Shinde (2025) [30]. *Student performance predictive analysis.*

REFERENCES

[1] K. Ouassif, B. Ziani, J. Herrera-Tapia, and C. A. Kerrache, "Empowering education: Leveraging clustering and recommendations for enhanced student insights," *Education Sciences*, vol. 15, no. 7, p. 819, 2025, doi: 10.3390/educsci15070819.

[2] M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg, "Advances in learning analytics and educational data mining," in *23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings*, 2015, pp. 297–306.

[3] I. Šarić-Grgić, A. Grubišić, L. Šerić, and T. J. Robinson, "Student clustering based on learning behavior data in the intelligent tutoring system," *International Journal of Distance Education Technologies*, vol. 18, no. 2, pp. 73–89, 2020, doi: 10.4018/IJDET.2020040105.

[4] D. Hooshyar, M. Pedaste, and Y. Yang, "Mining educational data to predict students' performance through procrastination behavior," *Entropy*, vol. 22, no. 1, p. 12, 2020, doi: 10.3390/e22010012.

[5] Á. M. Navarro and P. Moreno-Ger, "Comparison of clustering algorithms for learning analytics with educational datasets," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, p. 9, 2018, doi: 10.9781/ijimai.2018.02.003.

[6] K. DeFreitas and M. Bernard, "Comparative performance analysis of clustering techniques in educational data mining," *IADIS International Journal on Computer Science and Information Systems*, vol. 10, no. 2, pp. 65–78, 2015.

[7] S. Križanić, "Educational data mining using cluster analysis and decision tree technique: A case study," *International Journal of Engineering Business Management*, vol. 12, p. 1847979020908675, 2020, doi: 10.1177/1847979020908675.

[8] T. P. Vital, B. G. Lakshmi, H. S. Rekha, and M. DhanaLakshmi, "Student performance analysis with using statistical and cluster studies," in *Advances in Intelligent Systems and Computing*, vol. 758, J. Nayak, A. Abraham, B. M. Krishna, G. T. Chandra Sekhar, and A. K. Das, Eds. Singapore: Springer, 2018, pp. 743–757. doi: 10.1007/978-981-13-0514-6_71.

[9] B. Jin, X. Xu, and Y. Zhang, "Thermal coal futures trading volume predictions through the neural network," *Journal of Modelling in Management*, vol. 20, no. 2, pp. 585–619, 2025, doi: 10.1108/JM2-09-2023-0207.

[10] B. Jin, X. Xu, and Y. Zhang, "Peanut oil price change forecasts through the neural network," *Foresight*, vol. 27, no. 3, pp. 595–612, 2025, doi: 10.1108/FS-01-2023-0016.

[11] B. Jin and X. Xu, "Gaussian process regression based silver price forecasts," *Journal of Uncertain Systems*, vol. 17, no. 3, p. 2450013, 2024, doi: 10.1142/S1752890924500132.

[12] B. Jin and X. Xu, "Forecasts of thermal coal prices through Gaussian process regressions," *Ironmaking & Steelmaking: Processes, Products and Applications*, vol. 51, no. 8, pp. 819–834, 2024, doi: 10.1177/03019233241265194.

[13] B. Jin and X. Xu, "Contemporaneous causal orderings among prices of retail properties: evidence from Chinese cities through vector error-correction modeling and directed acyclic graphs," *Journal of Financial Management of Property and Construction*, 2025, doi: 10.1108/JFMPC-03-2024-0019.

[14] X. Xu, "Contemporaneous and Granger causality among US corn cash and futures prices," *European Review of Agricultural Economics*, vol. 46, no. 4, pp. 663–695, 2019, doi: 10.1093/erae/jby036.

[15] X. Xu, "Corn cash price forecasting," *American Journal of Agricultural Economics*, vol. 102, no. 4, pp. 1297–1320, 2020, doi: 10.1002/ajae.12041.

[16] X. Xu and Y. Zhang, "Individual time series and composite forecasting of the Chinese stock index," *Machine Learning with Applications*, vol. 5, p. 100035, 2021, doi: 10.1016/j.mlwa.2021.100035.

[17] K. Chrysafiadi and M. Virvou, "Student modeling approaches: A literature review for the last decade," *Expert Systems with Applications*, vol. 40, no. 11, pp. 4715–4729, 2013, doi: 10.1016/j.eswa.2013.02.007.

[18] I. Fuseini and Y. M. Missah, "A critical review of data mining in education on the levels and aspects of education," *Quality Education for All*, vol. 1, no. 2, pp. 41–59, 2024, doi: 10.1108/QEA-01-2024-0006.

[19] X. Li, Y. Zhang, H. Cheng, F. Zhou, and B. Yin, "An unsupervised ensemble clustering approach for the analysis of student behavioral patterns," *IEEE Access*, vol. 9, pp. 7076–7091, 2021, doi: 10.1109/ACCESS.2021.3049157.

[20] T. Prabha and D. S. Priyaa, "Knowledge discovery of the students academic performance in higher education using intuitionistic fuzzy based clustering," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 24, pp. 7005–7019, 2017.

[21] Z. S. Hafdi and S. El Kafhali, "Student performance prediction in learning management system using small dataset," in *Lecture Notes on Data Engineering and Communications Technologies*, 2023, vol. 164, pp. 197–205. doi: 10.1007/978-3-031-27762-7_19.

[22] J. Xu, K. H. Moon, and M. Van Der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017, doi: 10.1109/JSTSP.2017.2692560.

[23] W. Xing and D. Du, "Dropout prediction in MOOCs using deep learning for personalized intervention," *Journal of Educational Computing Research*, vol. 57, no. 3, pp. 547–570, 2019, doi: 10.1177/0735633118757015.

[24] H. Sharif and A. Atif, "The evolving classroom: how learning analytics is shaping the future of education and feedback mechanisms," *Education Sciences*, vol. 14, no. 2, p. 176, 2024, doi: 10.3390/educsci14020176.

[25] T. A. Mikropoulos and A. Natsis, "Educational virtual environments: A ten-year review of empirical research (1999-2009)," *Computers and Education*, vol. 56, no. 3, pp. 769–780, 2011, doi: 10.1016/j.compedu.2010.10.020.

[26] I. Paliokas and P. Theodorou, "Virtual/augmented reality applications in education & life long learning," *Electronics (Switzerland)*, vol. 13, no. 3, 2024, doi: 10.3390/electronics13030618.

[27] W. Strielkowski, V. Grebennikova, A. Lisovskiy, G. Rakhimova, and T. Vasileva, "AI-driven adaptive learning for sustainable educational transformation," *Sustainable Development*, vol. 33, no. 2, pp. 1921–1947, 2025, doi: 10.1002/sd.3221.

[28] L. Dai, Y. H. Jiang, Y. Chen, Z. Guo, T. Y. Liu, and X. Shao, "Agent4EDU: Advancing AI for education with agentic workflows," in *Proceedings of 2024 3rd International Conference on Artificial Intelligence and Education, ICAIE 2024*, 2025, pp. 180–185. doi: 10.1145/3722237.3722268.

[29] K. Govindasamy and T. Velmurugan, "Analysis of student academic performance using clustering techniques," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 15, pp. 309–323, 2018.
[30] Y. Shinde, "Student performance predictive analysis," Kaggle, 2022. https://www.kaggle.com/code/yogesh239/student-performance-predictive-analysis/input (accessed Feb. 22, 2025).
[31] S. Malik *et al.*, "Advancing educational data mining for enhanced student performance prediction: a fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution," *Scientific Reports*, vol. 15, p. 8738, Mar. 2025, doi: 10.1038/s41598-025-92324-x.
[32] N. Alzahrani, M. Meccawy, H. Samra, and H. A. El-Sabagh, "Identifying weekly student engagement patterns in e-learning via K-means clustering and label-based validation," *Electronics (Switzerland)*, vol. 14, no. 15, p. 3018, 2025, doi: 10.3390/electronics14153018.
[33] V. Braverman, J. C. Ergun, C. Wang, and S. Zhou, "Learning-augmented hierarchical clustering," in *Proceedings of Machine Learning Research*, 2025, vol. 267, pp. 5454–5507.
[34] M. Yan, T.-J. Hou, Z. Li, K. Mohan, and C.-P. Yuan, "Generalized statistical model for fits to parton distributions," *Physical Review D*, vol. 112, no. 3, 2025, doi: 10.1103/zcsy-34cc.
[35] M. Prasad, M. Jagadeeshwar, and D. D. Shanthi, "Comparative study of K-Medoids and fuzzy K-Means clustering for the selection of optimal cloud service provider," *Educational Administration: Theory and Practice*, vol. 30, no. 4, 2024, doi: 10.53555/kuey.v30i4.220.
[36] Q. Yu *et al.*, "A local adaptive fuzzy spectral clustering method for robust and practical clustering," *Scientific Reports*, vol. 15, no. 1, p. 7833, 2025, doi: 10.1038/s41598-025-91812-4.
[37] M. Gul and M. A. Rehman, "Big data: an optimized approach for cluster initialization," *Journal of Big Data*, vol. 10, no. 1, p. 120, 2023, doi: 10.1186/s40537-023-00798-1.
[38] F. Ros, R. Riad, and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, pp. 178–199, 2023, doi: 10.1016/j.neucom.2023.01.043.
[39] J. M. Contreras, E. Molina Portillo, and J. M. Fernández Luna, "Evaluation of hierarchical clustering methodologies for identifying patterns in timeout requests in EuroLeague basketball," *Mathematics*, vol. 13, no. 15, p. 2414, 2025, doi: 10.3390/math13152414.
[40] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, p. 6086, 2024, doi: 10.1038/s41598-024-56706-x.
[41] A. F. Mohamed Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering analysis for classifying student academic performance in higher education," *Applied Sciences (Switzerland)*, vol. 12, no. 19, p. 9467, 2022, doi: 10.3390/app12199467.
[42] Scikit-learn Documentation, "Decision trees." 2025. Accessed: Sep. 21, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/tree.html
[43] S. Sarwat, M. A. Khan, F. U. Rehman, A. Alharbi, M. S. Hossain, and M. A. Altalbe, "Predicting students' academic performance with deep SVM layers," *Electronics*, vol. 10, no. 14, p. 1673, 2021.
[44] D. Dey *et al.*, "The proper application of logistic regression model in complex survey data: a systematic review," *BMC Medical Research Methodology*, vol. 25, no. 1, p. 15, 2025, doi: 10.1186/s12874-024-02454-5.
[45] L. Xu *et al.*, "Ultrahigh thermal stability and piezoelectricity of lead-free KNN-based texture piezoceramics," *Nature Communications*, vol. 15, no. 1, p. 9alik018, 2024, doi: 10.1038/s41467-024-53437-5.
[46] R. Iranzad and X. Liu, "A review of random forest-based feature selection methods for data science education and applications," *International Journal of Data Science and Analytics*, vol. 20, no. 2, pp. 197–211, 2025, doi: 10.1007/s41060-024-00509-w.
[47] S. Malik *et al.*, "Enhancing education quality with hybrid clustering and evolutionary neural networks in a multi phase framework," *Scientific Reports*, vol. 15, no. 1, p. 21323, 2025, doi: 10.1038/s41598-025-04622-z.
[48] R. Liu, "Data analysis of educational evaluation using K-Means clustering method," *Computational Intelligence and Neuroscience*, vol. 2022, no. 7, pp. 1–10, 2022, doi: 10.1155/2022/3762431.
[49] T. Le Quy, G. Friege, and E. Ntoutsi, "A review of clustering models in educational data science towards fairness-aware learning," *arXiv preprint*, pp. 43–94, 2023, doi: 10.1007/978-981-99-0026-8_2.
[50] A. D. Woods, S. F. Ahmed, B. D. Katz, and F. J. Morrison, "How stable is early academic performance? using cluster analysis to classify low achievement and EF," *Journal of Learning Disabilities*, vol. 53, no. 1, pp. 19–35, 2020, doi: 10.1177/0022219419881632.
[51] S. S. M. Ajibade, N. Bahiah Binti Ahmad, and S. Mariyam Shamsuddin, "Educational data mining: Enhancement of student performance model using ensemble methods," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 551, no. 1. doi: 10.1088/1757-899X/551/1/012061.
[52] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods," *International Journal of Database Theory and Application*, vol. 9, no. 8, pp. 119–136, 2016, doi: 10.14257/ijdta.2016.9.8.13.

BIOGRAPHIES OF AUTHORS

Ghalia Mdaghri Alaoui is a dedicated Ph.D. student passionate about the intersection of educational data mining, machine learning, and personalized learning in virtual reality. Their research focuses on clustering and classification techniques to analyze student performance and to create adaptive learning paths. Ghalia is particularly interested in leveraging AI-driven insights to enhance personalized education. Their work also explores the application of virtual reality to optimize learning experiences. She can be contacted via email at: ghalia.mdaghrialaoui@etu.uae.ac.ma.

Abdelhamid Zouhair is a professor at the University of Abdelmalek Essaâdi, Faculty of Sciences and Techniques of Tangier, Department of Computer Sciences since March 2020, and a Professor at the National School of Applied Sciences of AL Hoceïma, between May 2016 and March 2020. He obtained his Ph.D. in Computer Sciences from the University of Le Havre (France) and the University of Abdelmalek Essaâdi, Faculty of Sciences and Technology of Tangier (Morocco) in 2014. He obtained his HDR in 2020 and is the Director of several Doctoral theses in Computer Sciences. He is an author or co-author of several articles published in International Journals in Computer Science. He can be contacted via email at a.zouhair@uae.ac.ma.

Ilhame Khabbachi is a dedicated Ph.D. student passionate about the intersection of deep learning, facial emotion recognition, and online education security. Her research focuses on developing a cheating detection system for online exams using real-time facial emotion recognition powered by deep learning. She aims to enhance the integrity of e-learning assessments by leveraging AI-driven techniques to detect suspicious behaviors. Her work contributes to the advancement of secure and reliable online examination systems. She can be contacted via email at ilhame.khabbachi@etu.uae.ac.ma.