

AN EXPLAINABLE FUZZY CLUSTERING FRAMEWORK FOR MODELING LEARNING TRAJECTORIES IN OUTCOME-BASED EDUCATION

Rustam ^{1*}, Koredianto Usman ²

^{1,2}Telecommunication Engineering Study Program, School of Electrical Engineering, Telkom University

^{1,2}Center of Excellence for Strategic Advancement for Key digital Transformation Indonesia, Research Institute for Intelligent Business and Sustainable Economy, Telkom University
Jln. Telekomunikasi no. 1, Bandung, Jawa Barat 40257, Indonesia

Corresponding author's e-mail: * rustamtelu@telkomuniversity.ac.id

Article Info

Article History:

Received: 11th July 2025

Revised: 20th December 2025

Accepted: 16th March 2026

Available online: 8th April 2026

Keywords:

Conceptual mastery;

Educational data mining;

Explainable AI;

Fuzzy clustering;

Learning trajectories;

Outcome-based education.

ABSTRACT

Although various studies have applied clustering and categorization techniques in educational assessment, most rely on deterministic thresholds or heuristic-based partitions that fail to uncover latent conceptual structures. Existing fuzzy clustering applications in education also seldom incorporate rigorous validity-index-driven model selection or explainability-focused interpretation, leaving a gap in modeling the gradual and overlapping nature of learning progression. Outcome-Based Education (OBE) emphasizes measurable learning outcomes as the cornerstone of curriculum design and assessment. However, traditional methods for classifying student performance—typically based on fixed score thresholds—fail to reflect the inherent complexity of conceptual learning. This study proposes an Explainable Fuzzy Clustering Framework to model student learning trajectories in an OBE environment. Using final scores derived from multiple Course Learning Outcomes (CLOs), the Fuzzy C-Means (FCM) algorithm is applied to cluster students into conceptual performance levels with soft membership assignments. The optimal number of clusters is determined using the Tang–Sun–Sun (TSS) and Xie–Beni (XB) validity indices. The resulting fuzzy clusters are then compared with three conventional manual classification schemes—fixed thresholding, quantile partitioning, and mean–standard deviation banding—using cross-tabulation, heatmaps, and quantitative agreement metrics such as Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI). Visualization techniques including stacked membership plots and cluster-size bar charts are employed to enhance interpretability. Results show that fuzzy clustering moderately aligns with manual schemes while revealing latent transitions and overlapping boundaries that rigid methods overlook. Quantitatively, the fuzzy clusters formed a natural distribution of 17.9% Low, 36.9% Moderate, and 45.2% High performers, with agreement scores of $ARI = 0.405–0.462$ and $NMI = 0.550–0.629$. These findings confirm the robustness and interpretability of the proposed model. The framework provides a principled, explainable, and adaptive approach to formative assessment, contributing to the advancement of interpretable learning analytics in higher education.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) (<https://creativecommons.org/licenses/by-sa/4.0/>).

How to cite this article:

Rustam and K. Usman., “AN EXPLAINABLE FUZZY CLUSTERING FRAMEWORK FOR MODELING LEARNING TRAJECTORIES IN OUTCOME-BASED EDUCATION”, *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 1949-1966, Sep, 2026.

Copyright © 2026 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: barekeng.math@yahoo.com; barekengjournal@mail.unpatti.ac.id

Research Article · Open Access

1. INTRODUCTION

Outcome-Based Education (OBE) is a widely adopted paradigm in higher education that emphasizes clearly defined learning outcomes as the foundation for curriculum design, instructional delivery, and student assessment [1], [2]. In STEM education, this approach requires instructors to ensure that students achieve measurable competencies aligned with course-level and program-level goals [3]. However, one persistent challenge in implementing OBE lies in assessing students' conceptual understanding in a manner that is both formative and meaningful [4]. Traditional approaches often rely on fixed grading thresholds to classify student performance into discrete categories such as Low, Moderate, and High [5]. While these categories are convenient, they can be misleading due to their arbitrary nature and lack of alignment with the underlying distribution of learning outcomes [6]. These conventional classification schemes fail to capture the complexity, variability, and transitional nature of student learning [7]. Learning is a gradual process that often involves overlapping concepts, partial understanding, and nonlinear progressions [8]. As such, deterministic labels may obscure latent conceptual structures and hinder meaningful formative feedback [9]. In response to these limitations, fuzzy clustering has emerged as a promising alternative for modeling learning profiles [10].

Fuzzy clustering, particularly through the Fuzzy C-Means (FCM) algorithm, allows each data point—i.e., student performance—to belong to multiple clusters with varying degrees of membership [11]. This characteristic mirrors the nature of cognitive development, which is continuous rather than discrete [12]. Unlike hard partitions, fuzzy clustering enables the representation of ambiguous or transitional learners who may not fit neatly into one category [13]. This provides a richer understanding of where a student stands conceptually and allows educators to design more responsive interventions [14]. A core challenge in applying fuzzy clustering is determining the optimal number of clusters, which directly affects the interpretability and utility of the results [15]. While some studies arbitrarily fix the number of clusters to match grading schemes, such practice undermines the advantages of data-driven modeling [16]. To address this, clustering validity indices are used to evaluate the quality of clustering results based on internal criteria such as compactness and separation [17]. Among these, the Xie–Beni (XB) index is widely used due to its ability to penalize overlapping clusters and reward compact structures [18]. However, the Tang–Sun–Sun (TSS) index has been shown to offer more stability and sensitivity in the presence of fuzzy boundaries and is particularly suited for educational data [19].

Several prior studies have applied fuzzy clustering in education for various purposes, including student segmentation, learning style identification, and evaluating the effects of instructional interventions [20]. Some approaches combine fuzzy clustering with transition matrices to model changes over time [21]. However, many of these studies are limited to pretest-posttest designs and fail to consider the complexity of learning trajectories across multiple Course Learning Outcomes (CLOs) [22]. In formal OBE implementation, each course defines a set of CLOs that represent progressive levels of conceptual mastery [1]. These CLOs are aligned with Bloom's taxonomy and provide scaffolding from lower-order to higher-order thinking skills [23]. CLO-based assessments thus offer a valuable lens to observe the evolution of student understanding over time [4].

Recent research beyond the education domain has also demonstrated the value of explainable clustering and classification methods in high-dimensional and complex data scenarios. For example, fuzzy and neural-network approaches have been successfully applied for identifying the geographical origin of clove buds using metabolite compositions [29], and hard c-means clustering has been employed for grouping Indonesian plantation commodities based on their metabolic profiles [30]. In medical imaging, decision tree classifiers combined with visual feature extraction techniques like Histogram of Oriented Gradients have been utilized for anemia detection through conjunctival analysis [31]. Moreover, studies in metabolomics have shown how proper pretreatment methods can significantly enhance machine learning model performance in high-dimensional datasets [32]. These findings reinforce the importance of explainability and preprocessing across diverse domains, offering methodological insights that are transferable to educational analytics.

Despite this, there remains a gap in modeling student progress across CLOs using mathematically grounded and explainable techniques [24]. This study proposes an Explainable Fuzzy Clustering Framework to address this gap. The framework combines fuzzy clustering with a systematic model selection procedure based on the TSS and XB indices. It then compares the resulting clusters with manual classification schemes—threshold-based, quantile-based, and mean–standard deviation-based—to assess consistency, interpretability, and granularity. By incorporating explainable visualization techniques, the model also

supports the identification of transitional learners and latent performance patterns. Unlike previous research that uses fuzzy clustering solely for exploratory segmentation, this study emphasizes explainability and educational relevance. The results are not only evaluated statistically but also interpreted pedagogically to support formative assessment. Furthermore, this research contributes a methodological foundation for extending fuzzy modeling to dynamic analysis, such as using fuzzy transition matrices across CLO stages. Ultimately, the proposed framework provides a novel, interpretable approach to analyzing student performance in OBE environments. It facilitates the discovery of hidden conceptual clusterings, supports adaptive instruction, and offers actionable insights for educators. In doing so, it contributes to the advancement of learning analytics and the development of intelligent, learner-centered education systems.

2. RESEARCH METHODS

This section presents the methodological framework developed to support the construction and evaluation of an explainable fuzzy clustering approach for modeling conceptual learning trajectories within an Outcome-Based Education (OBE) environment. The framework integrates conventional deterministic classification schemes with a mathematically grounded fuzzy clustering model, allowing for both comparative evaluation and deeper pedagogical interpretation of student performance patterns. The methodology is organized into five key components. First, we describe the dataset, which consists of total scores from multiple assessment components—such as quizzes, assignments, midterms, and final exams—each mapped to specific Course Learning Outcomes (CLOs). This structure provides a multidimensional perspective on student achievement that reflects the hierarchical and cumulative nature of learning in an OBE framework. Second, we construct three baseline classification models based on deterministic rules: fixed thresholding, quantile-based partitioning, and mean–standard deviation banding. These manual schemes represent standard practices in institutional assessment and serve as reference anchors for evaluating the fuzzy clustering output. Third, we formally define the fuzzy clustering model using the Fuzzy C-Means (FCM) algorithm, highlighting its capacity to represent partial membership and uncover latent conceptual structures. This includes the mathematical formulation of the objective function, the iterative optimization process, and the interpretability advantages of soft clustering in educational contexts. Fourth, we detail the procedure for determining the optimal number of clusters using internal clustering validity indices. Specifically, we employ the Tang–Sun–Sun (TSS) index as the primary selection criterion due to its robustness in fuzzy environments, and the Xie–Beni (XB) index as a complementary benchmark for model validation. This data-driven approach ensures that the clustering structure reflects the intrinsic characteristics of the student performance distribution rather than arbitrary assumptions.

Finally, we outline the evaluation and comparison framework, which involves: cross-tabulation analysis to examine alignment with manual labels, visualization techniques—including heatmaps, bar charts, and fuzzy membership plots—to enhance explainability, and quantitative agreement metrics such as the Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) to measure clustering similarity. These complementary components enable a multi-angle assessment that considers both statistical validity and pedagogical significance.

2.1 Dataset

The dataset used in this study was obtained from the Calculus 2 course offered in the Telecommunication Engineering undergraduate program at Telkom University. This course is implemented under the Outcome-Based Education (OBE) framework and is designed to develop students' competence in solving mathematical problems relevant to engineering contexts. The course includes assessments that are directly aligned with four Course Learning Outcomes (CLOs), each focusing on a distinct conceptual area.

The data were collected during the even semester of the 2024/2025 academic year, involving two parallel class sections. The first class consisted of 40 students, and the second of 44 students, resulting in a total of $N = 84$ student records. Each student's total score was computed from a weighted combination of evaluations mapped to the following CLOs:

CLO 1: Techniques of integration

CLO 2: Infinite sequences and series

CLO 3: Vector-valued and multivariable functions

CLO 4: Double and triple integrals

Let $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}$ represent the set of total scores, where each $x_i \in [0, 100]$ corresponds to the aggregated performance of the i -th student. These scores serve as a one-dimensional numerical representation of students' conceptual mastery across the course.

The primary objective of this study is to model and analyze the underlying conceptual structure of student performance using a fuzzy clustering approach, and to evaluate its explainability and alignment with conventional deterministic classification schemes. This analysis is intended to support a more nuanced interpretation of learning trajectories within an OBE setting.

2.2 Manual Classification as Deterministic Partitioning

To establish comparative baselines for evaluating the proposed fuzzy clustering framework, this study employs three deterministic (manual) classification schemes commonly used in educational data analysis: fixed thresholding, quantile-based partitioning, and mean-standard deviation-based classification. Each of these approaches assigns students to mutually exclusive conceptual performance levels: **Low** (C_L), **Moderate** (C_M), and **High** (C_H).

2.2.1 Fixed Thresholding

This approach defines class boundaries based on predetermined score intervals. Let $\alpha = 60$ and $\beta = 80$ be the fixed cutoff points, then the data domain X is partitioned into:

$$C_L = \{x_i \in X \mid x_i \leq \alpha\}, \quad (1)$$

$$C_M = \{x_i \in X \mid \alpha < x_i \leq \beta\}, \quad (2)$$

$$C_H = \{x_i \in X \mid x_i > \beta\}. \quad (3)$$

The three deterministic classification schemes described above serve as baseline clusterings against which the proposed fuzzy clustering framework will be compared. Among these, the fixed-threshold approach aligns with the official grading policy implemented at Telkom University, in which final course scores are mapped to letter grades based on institutional criteria. According to the university's assessment guidelines, scores of 60 and below correspond to grades BC and lower, while scores above 80 are associated with excellent performance levels, mapped to grades AB and A. This policy-driven mapping lends institutional validity to the use of 60 and 80 as critical cut-off points in the Low–Moderate–High classification model. This institutional alignment lends pedagogical and contextual justification to the chosen thresholds, making this classification approach both practically relevant and academically grounded.

By incorporating these institutional standards and augmenting them with data-driven partitioning methods (quantile and mean-standard deviation based), this study provides a robust foundation for comparing conventional evaluation strategies with the fuzzy clustering approach. The next section introduces the mathematical formulation of the fuzzy clustering model and outlines its optimization mechanism.

2.2.2 Quantile-Based Partitioning

This method divides the dataset $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}$ into three performance levels—Low, Moderate, and High—based on empirical tertiles, which are the 33.3rd percentile and 66.7th percentile of the sorted dataset. This approach is commonly used in educational data mining to obtain relative performance clusterings that are evenly distributed in terms of the number of individuals in each cluster. Let $x_1 \leq x_2 \leq \dots \leq x_N$ be the ordered version of the dataset. Then, q_1 is the first tertile (33.3rd percentile), defined as the score below which 33.3% of the data fall and q_2 is the second tertile (66.7th percentile), defined as the score below which 66.7% of the data fall.

These thresholds are computed by the empirical quantile function:

$$q_1 = Q_X(0,333), \quad q_2 = Q_X(0,667), \quad (4)$$

where $Q_X(p)$ denotes the p -th quantile of the dataset X . Once q_1 and q_2 are determined, the partitions are defined as:

$$C_L = \{x_i \in X \mid x_i \leq q_1\}, \quad (5)$$

$$C_M = \{x_i \in X \mid q_1 < x_i \leq q_2\}, \quad (6)$$

$$C_H = \{x_i \in X \mid x_i > q_2\}. \quad (7)$$

Each cluster—Low, Moderate, and High—thus contains approximately one-third of the dataset. This ensures a balanced cluster size and avoids the distortion that may arise from skewed distributions when using fixed thresholds. It is particularly useful for identifying relative performance bands when absolute performance cutoffs are not pre-defined.

This method supports equitable comparisons and is beneficial for statistical modeling, though it lacks the pedagogical alignment of threshold-based classification grounded in institutional grading policies. Nevertheless, it provides a robust empirical baseline for comparing the granularity and consistency of fuzzy clustering results.

2.2.3 Mean–Standard Deviation-Based Partitioning

The third classification scheme applies a norm-referenced approach, in which performance levels are defined relative to the empirical mean and standard deviation of the dataset. This method is grounded in classical measurement theory and is widely used in standardized testing and norm-based assessment models.

Let μ and σ denote the sample mean and sample standard deviation of the dataset $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}$, computed as

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i, \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}. \quad (8)$$

Using μ and σ , we define two dynamic cut-off values:

Lower cut-off: $\mu - 0.5\sigma$

Upper cut-off: $\mu + 0.5\sigma$

The partitioning rules are then established as:

$$C_L = \{x_i \in X \mid x_i \leq \mu - 0.5\sigma\}, \quad (9)$$

$$C_M = \{x_i \in X \mid \mu - 0.5\sigma < x_i \leq \mu + 0.5\sigma\}, \quad (10)$$

$$C_H = \{x_i \in X \mid x_i > \mu + 0.5\sigma\}. \quad (11)$$

This approach assumes that the data are approximately symmetrically distributed and uses the standard deviation to model the spread of student performance around the central tendency (mean). The choice of ± 0.5 standard deviation as the cutoff bandwidth is motivated by its ability to provide a central zone of moderate performers, while identifying the lower and higher tails of the distribution as Low and High performers, respectively.

One advantage of this method is that it adapts to the characteristics of the dataset, unlike fixed-thresholding which uses static, possibly arbitrary cutoffs. However, its effectiveness depends on the shape of the data distribution—if the data are heavily skewed or multimodal, this method may misrepresent actual conceptual clusterings.

Despite this limitation, mean-standard deviation partitioning provides a valuable statistical benchmark for evaluating clustering models and offers an alternative, distribution-aware classification method that reflects performance variation in a quantitative manner.

2.3 Fuzzy Clustering Formulation

While manual classification schemes rely on fixed or data-dependent hard boundaries, fuzzy clustering offers a mathematically grounded alternative that captures the inherent uncertainty and gradual nature of conceptual learning. In contrast to hard clustering methods, fuzzy clustering allows each data point to belong to multiple clusters with varying degrees of membership. This property is particularly beneficial in educational contexts where student understanding does not always fit neatly into discrete categories.

This study employs the Fuzzy C-Means (FCM) algorithm to discover latent conceptual clusterings within the total score data of students. FCM is an unsupervised learning method that partitions a set of numerical vectors into c fuzzy clusters by minimizing an objective function based on the distance between data points and cluster centers, weighted by membership degrees.

Let $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}$ be the set of total score of N students. The goal of FCM is to find [25], [26], [27], [28]

A set of cluster centers $V = \{v_1, v_2, \dots, v_c\} \subset \mathbb{R}$, and

A membership matrix $U = [u_{ij}] \in [0, 1]^{c \times N}$, where u_{ij} represents the degree of membership of data point x_j to cluster i .

subject to the constraint

$$\sum_{i=1}^c u_{ij} = 1 \text{ for all } j = 1, 2, \dots, N. \quad (12)$$

The bijective function minimized by FCM is

$$J_m(U, V) = \sum_{j=1}^N \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|^2, \quad (13)$$

where $m > 1$ is the fuzzifier exponent, controlling the degree of cluster fuzziness (typically set to $m = 2$), and $\|\cdot\|$ denotes the Euclidean norm [25], [26], [27].

The minimization of $J_m(U, V)$ is performed using an iterative optimization procedure with the following update rules:

Cluster centers update

$$v_i = \frac{\sum_{j=1}^N u_{ij}^m x_j}{\sum_{j=1}^N u_{ij}^m}, \quad (14)$$

Membership degrees update

$$u_{ij} = \left(\sum_{k=1}^c \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad (15)$$

The iterative process continues until the maximum change in membership values between two successive iterations is less than a predefined threshold ε . In this study, FCM is implemented for various candidate values of c . For each value of c , the resulting partition is evaluated using two clustering validity indices to determine the optimal number of clusters that best fits the data structure.

Unlike previous studies that assume a fixed number of clusters (e.g., $c = 3$ for Low–Moderate–High), this study introduces a data-driven approach using clustering validity indices to determine the optimal number of conceptual clusterings. This adaptive mechanism provides a more faithful representation of student learning heterogeneity and supports the development of an explainable clustering framework that is tailored to the structure of actual student data.

2.4 Validity Indices for Optimal Clustering

Determining the optimal number of clusters (cc) is a fundamental challenge in clustering analysis, particularly when the objective is to uncover interpretable and pedagogically meaningful clusterings of students. While educational studies often assume $c = 3$ to reflect intuitive categories such as *Low*, *Moderate*, and *High* conceptual mastery, such assumptions may oversimplify the complexity and heterogeneity of real-world student performance data. Therefore, this study adopts a data-driven approach based on clustering validity indices to objectively determine the optimal number of clusters.

Clustering validity indices are quantitative measures used to evaluate the quality of clustering results based on internal criteria such as compactness (intra-cluster cohesion), separation (inter-cluster distinction), and in fuzzy clustering, the degree of overlap among memberships. In this study, we employ the Tang–Sun–Sun (TSS) Index as the primary criterion for selecting the optimal number of clusters due to its ability to balance compactness and separation while being robust to fuzziness. To enhance the comprehensiveness of the evaluation, the Xie–Beni (XB) Index is also used as a complementary benchmark.

2.4.1 Tang–Sun–Sun Index (TSS)

The Tang–Sun–Sun (TSS) index is a fuzzy validity measure that simultaneously considers intra-cluster variance and inter-cluster distance, while also addressing the problem of small denominators in cluster separation terms. It is formally defined as [26]:

$$TSS(U, V, X) = \frac{\sum_{i=1}^c \sum_{k=1}^N u_{ik}^2 \|x_k - v_i\|^2}{\min_{1 \leq i, j \leq c, i \neq j} \|v_i - v_j\|^2 + \frac{1}{c}} + \frac{\frac{1}{c(c-1)} + \sum_{i=1}^c \sum_{j=1, i \neq j}^c u_{ik}^2 \|v_i - v_j\|^2}{\min_{1 \leq i, j \leq c, i \neq j} \|v_i - v_j\|^2 + \frac{1}{c}}. \quad (16)$$

The first term captures the intra-cluster dispersion weighted by squared membership values, while the second term introduces a normalized penalty for small inter-cluster distances. The small constant $\frac{1}{c}$ is added to avoid division by zero and to stabilize the index under tight clusters. A lower TSS value indicates better clustering performance in terms of compactness and separability.

The TSS index is particularly well-suited for educational datasets where overlapping conceptual boundaries and varying cluster scales are common, and where explainability and interpretive fidelity are essential.

2.4.2 Xie-Beni Index (XB)

To complement the TSS index, we also apply the Xie–Beni (XB) index, a widely used fuzzy clustering validity measure. The XB index is defined as [27]:

$$XB(U, V) = \frac{\sum_{i=1}^c \sum_{j=1}^N u_{ij}^2 \|x_j - v_i\|^2}{N \cdot \min_{i \neq k} \|v_i - v_k\|^2}. \quad (17)$$

where the numerator represents the total within-cluster variance, weighted by fuzzy membership degrees, and the denominator penalizes clusterings with poor separation. Lower XB values are desirable, indicating more compact clusters with greater inter-cluster distinction.

While the XB index is computationally efficient and interpretable, it may be sensitive to the presence of closely located cluster centers, which can cause the denominator to approach zero. Thus, it is used in this study as a supplementary comparison index to validate and triangulate the results obtained from the TSS index.

Cluster Selection Procedure

To determine the optimal number of clusters c , the FCM algorithm is executed for a range of candidate values $c \in \{2, 3, 4, 5, 6\}$. For each clustering configuration, two validity indices were computed: the Tang–Sun–Sun (TSS) Index as the primary criterion, and the Xie–Beni (XB) Index as a complementary measure. The TSS index was used to identify the number of clusters that offered the best balance between intra-cluster compactness and inter-cluster separation, with lower TSS values indicating higher-quality clustering. Meanwhile, the XB index served as a secondary indicator to validate whether the configuration selected by TSS was also supported by traditional fuzzy clustering metrics.

The final selection of the optimal cluster count was made by identifying the value of c that minimized the TSS index while ensuring that the XB index remained acceptably low. In this study, both indices converged to suggest $c = 3$ as the most appropriate number of clusters. This dual-index validation strategy reinforces confidence in the clustering solution, ensuring that the configuration is not only mathematically optimal, but also educationally meaningful and interpretable.

Importantly, the use of validity indices in this study is not limited to performance evaluation. Rather, it constitutes a core mechanism within the proposed explainable clustering framework. By relying on internal validity criteria rather than external assumptions or heuristics, the framework enables the discovery of latent conceptual clusterings that emerge directly from student performance data. This approach aligns with the principles of data-driven decision-making in learning analytics, promoting objectivity, fairness, and adaptability.

The resulting fuzzy clustering model—optimized using these validity indices—was subsequently analyzed and compared with conventional manual classification schemes to assess its alignment, granularity, and pedagogical relevance. The next section describes the procedures and metrics used in this comparative evaluation. To provide a clearer summary of the cluster selection mechanism described above, the procedural steps are organized into Algorithm 1.

Algorithm 1. Cluster Selection Procedure Using TSS and XB Validity Indices**Input:**

- Dataset X
- Candidate cluster counts $C = \{2, 3, 4, 5, 6\}$

Output:

- Optimal number of clusters (c^*)

Step 1 — Initialization

Prepare the set of candidate cluster counts C .

Step 2 — FCM Execution and Validity Calculation

For each value of c in C , perform the following steps:

- 2.1. Apply the Fuzzy C-Means (FCM) algorithm to dataset X using c clusters.
- 2.2. Compute the Tang–Sun–Sun validity index $TSS(c)$.
- 2.3. Compute the Xie–Beni validity index $XB(c)$.

Step 3 — Identification of Candidate Solutions

- 3.1. Identify c_TSS as the value of c that yields the minimum $TSS(c)$.
- 3.2. Identify c_XB as the value of c that yields the minimum $XB(c)$.
- 3.3. Form a candidate set $C_candidate$ that contains c_TSS , c_XB , and any other c values that jointly exhibit relatively low $TSS(c)$ and $XB(c)$.

Step 4 — Educational Interpretability Check

From $C_candidate$, select the value of c that:

- 4.1. Maintains low $TSS(c)$ and $XB(c)$, and
- 4.2. Produces clusters that are educationally meaningful and interpretable (for example, mapping naturally to Low, Moderate, and High performance levels in an OBE setting).

Denote this selected value as c^* .

Step 5 — Return Result

Return c^* as the optimal number of clusters.

2.5 Evaluation and Comparison

To assess the performance and pedagogical interpretability of the fuzzy clustering results, this study conducts a structured evaluation by comparing the output of the optimized fuzzy clustering model with the baseline manual classification schemes described in Section 2.2. The objective is to evaluate how well the data-driven fuzzy clusters align with existing deterministic clusterings, and to explore whether the fuzzy model reveals additional latent structures that are educationally meaningful. The comparison framework is composed of three complementary components: cross-tabulation analysis, visual inspection, and quantitative agreement metrics.

2.5.1 Cross-Tabulation Analysis

Cross-tabulation is used as a descriptive tool to examine the overlap between cluster assignments obtained from fuzzy clustering and those from manual classification. The fuzzy clustering results are first converted into crisp labels by applying the maximum membership rule, that is:

$$\mathcal{F}_j = \arg \max_i u_{ij}, \quad (18)$$

where \mathcal{F}_j is the crisp cluster label for student x_j , and u_{ij} is the fuzzy membership degree of x_j to cluster i . For each manual classification method (threshold-based, quantile-based, and mean-standard deviation-based), a **contingency matrix** is constructed to compare student assignments:

$$CT(a, b) = \{x_j \in X \mid \text{Manual}(x_j) = a, \mathcal{F}_j = b\}, \quad (19)$$

where $a \in \{\text{Low}, \text{Moderate}, \text{High}\}$ and $b \in \{1, 2, \dots, c\}$. The resulting tables allow qualitative evaluation of alignment and deviation between the fuzzy and manual clusterings.

2.5.2 Visualization for Explainability

To support explainability and interpretability, graphical visualizations are used to examine the structure and distribution of clusters:

Membership Degree Plot: for each data point x_j , the corresponding fuzzy membership values u_{ij} across clusters are plotted to illustrate partial associations.

Bar Plots or Histograms: used to show the frequency distribution of students within each fuzzy cluster and to compare with manual cluster sizes.

Heatmaps of Crosstabs: enhance the interpretability of contingency matrices, making it easier to observe areas of strong or weak alignment.

These visual tools are instrumental in identifying ambiguous cases, transitional learners, and potential subclusters within broader conceptual categories.

2.5.3 Quantitative Agreement Metrics

While cross-tabulation and visualization provide qualitative insights, quantitative agreement metrics offer formal evaluation of similarity between clustering outcomes. Two standard metrics are employed:

Adjusted Rand Index (ARI):

Measures the similarity between two data clusterings while correcting for chance. It ranges from -1 to 1 , with 1 indicating perfect agreement and 0 corresponding to random labeling.

$$ARI(A, B) = \frac{\text{Index} - \text{Expected Index}}{\text{Max Index} - \text{Expected Index}}, \quad (20)$$

Normalized Mutual Information (NMI):

Derived from information theory, NMI measures the amount of shared information between two partitions. It ranges from 0 to 1 , where higher values indicate greater consistency.

$$NMI(A, B) = \frac{2 \cdot I(A; B)}{H(A) + H(B)}, \quad (21)$$

where $I(A; B)$ is the mutual information between clusterings A and B , and $H(\cdot)$ denotes entropy.

These metrics are computed between the fuzzy clustering (after argmax conversion) and each manual classification. Their values provide an objective assessment of clustering agreement and serve as validation tools for interpreting the explainability and structural relevance of the fuzzy clusters. The combination of cross-tabulation, visualization, and clustering agreement metrics allows for a multi-angle evaluation of the clustering results. This comprehensive framework ensures that both statistical robustness and educational meaning are considered when comparing clustering models.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics of Student Performance

To establish a foundational understanding of students' conceptual mastery prior to clustering analysis, we first examined the descriptive statistics of scores associated with each Course Learning Outcome (CLO) as well as the final aggregated scores. Each CLO score is derived from two components: a structured assignment score and an examination score. [Table 1](#) presents the mean, standard deviation, minimum, quartiles, and maximum values for each of the CLO components, along with the overall final score.

Table 1. Descriptive Statistics of Student Scores

Component	Mean	Std	Min	Max	Q1	Q2 (Median)	Q3
CLO 1 Assignment	92.60	11.08	37.5	100.0	90.63	95.32	100.0
CLO 1 Exam	52.43	23.33	0.0	98.0	37.00	51.00	70.00
CLO 2 Assignment	92.60	11.08	37.5	100.0	90.63	95.32	100.0
CLO 2 Exam	37.02	19.10	0.0	82.0	23.38	34.50	50.00
CLO 3 Assignment	92.60	11.08	37.5	100.0	90.63	95.32	100.0
CLO 3 Exam	67.94	23.63	0.0	98.0	57.00	76.00	82.00
CLO 4 Assignment	92.60	11.08	37.5	100.0	90.63	95.32	100.0
CLO 4 Exam	58.95	26.38	0.0	100.0	35.00	64.50	80.00

Overall, the assignment scores across all CLOs exhibited high consistency, with a mean of 92.60 and a relatively low standard deviation of 11.08 . This suggests that the majority of students performed well in assignment components, likely due to the availability of resources and extended time for completion. In

contrast, the examination scores demonstrated more variability. For instance, the average score for CLO 1 Exam was 52.43 with a standard deviation of 23.33, and the lowest observed score was 0.0, indicating a wider spread in conceptual mastery when students were assessed under time constraints and without external support.

Among all examination components, CLO 2 Exam had the lowest average score (mean = 37.02) and the most compressed upper quartile ($Q3 = 50.0$), which may point to higher conceptual difficulty or gaps in instructional effectiveness for the topic of "Infinite Sequences and Series." This contrasts with CLO 1 Exam and CLO 3 Exam, which had higher median values (51.0 and 72.0 respectively), reflecting stronger average understanding in those areas.

3.2 Optimal Cluster Determination Using Validity Indices

A critical component in clustering analysis is determining the number of clusters, c , that best represents the latent structure of the data. In many educational studies, the number of conceptual clusters is often arbitrarily set to three (e.g., Low, Moderate, High). However, such a priori assumptions may not accurately capture the heterogeneity and progression of student understanding in real-world data. Therefore, this study adopts a data-driven approach to identify the optimal number of clusters using internal validity indices tailored for fuzzy clustering.

In particular, we evaluate clustering performance using two widely used metrics: the Tang–Sun–Sun (TSS) Index and the Xie–Beni (XB) Index. While XB focuses on the trade-off between intra-cluster compactness and inter-cluster separation, the TSS index extends this by incorporating a regularization term to stabilize the influence of minimum inter-cluster distances and to penalize degenerate clusterings more effectively.

The Fuzzy C-Means (FCM) algorithm was executed for a range of cluster numbers from $c = 2$ to $c = 6$, and both indices were computed for each configuration. Table 2 summarizes the resulting values, and Figure 1 presents the corresponding line plots.

Table 2. Validity Indices for Various Numbers of Clusters

Number of Clusters (c)	Tang–Sun–Sun Index (TSS)	Xie–Beni Index (XB)
2	1.049144	0.147920
3	0.951104	0.175077
4	0.963723	0.183232
5	0.978256	0.069927
6	1.038883	0.194231

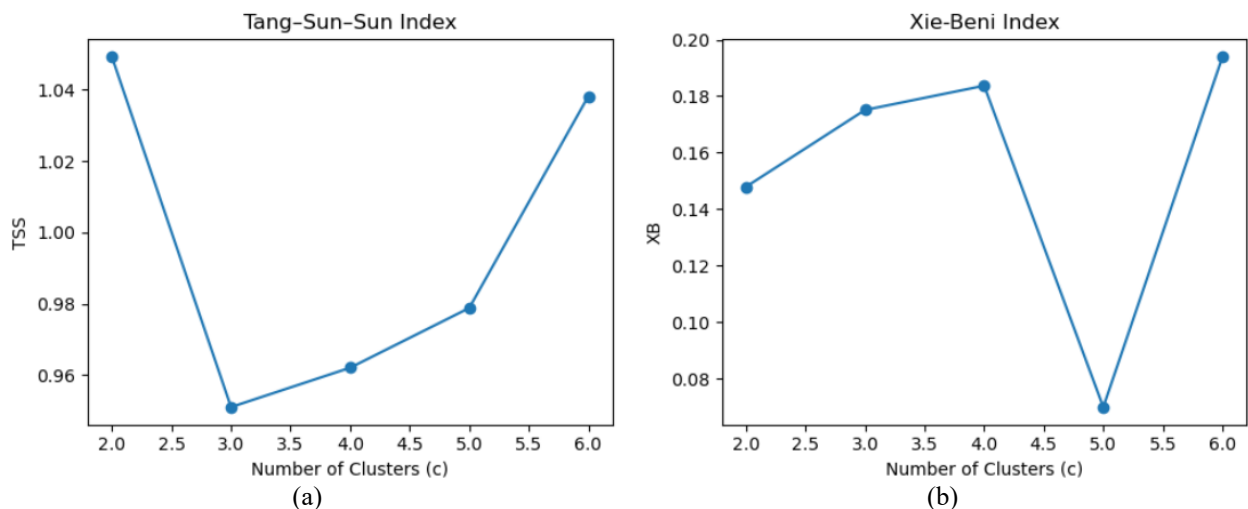


Figure 1. Validity Index Trends Across Cluster Configurations

(a) The Tang-Sun-Sun Index, (b) The Xie-Beni Index

Fig. 1 shows that the TSS index achieved its lowest value at $c = 3$, indicating that a three-cluster configuration offers the best balance between within-cluster compactness and between-cluster separation. Although the XB index reached its minimum at $c = 5$, this configuration may result in overly fragmented clusters that are less interpretable in educational contexts, especially when the aim is to identify broad conceptual stages in student learning.

The Tang–Sun–Sun index, by design, is more robust to irregular inter-cluster distances and better captures the stability of fuzzy partitions. This makes it particularly suited for evaluating conceptual progression where clusters are expected to exhibit gradual transitions rather than sharp boundaries. Therefore, the selection of $c = 3$ as the optimal number of clusters is well justified, as it aligns both mathematically and pedagogically with the goal of modeling explainable learning trajectories.

The three-cluster configuration also aligns with the typical categorization of students into Low, Moderate, and High performance levels. However, unlike rigid threshold-based divisions, the fuzzy clustering approach preserves partial membership and soft transitions, making it more adaptable to individual learning patterns and more interpretable from an instructional standpoint. This optimal configuration ($c = 3$) will be used in the subsequent sections to interpret the fuzzy membership structure and to compare it against conventional deterministic classification methods.

3.3 Cluster Interpretation and Membership Distribution

Following the identification of the optimal number of clusters ($c = 3$) using the Tang–Sun–Sun (TSS) index, this section presents a comprehensive interpretation of the resulting fuzzy clusters and their pedagogical meaning. The discussion is organized into three parts: a semantic interpretation of each cluster, analysis of the distribution and balance of student assignments, and alignment of fuzzy clustering results with conventional manual classification schemes.

3.3.1 Cluster Semantic Interpretation

The fuzzy clustering algorithm assigns each student to a conceptual cluster based on their normalized final score, allowing for partial membership across multiple clusters. To enhance interpretability, the three resulting clusters were sorted according to their centroid values in ascending order, enabling consistent mapping onto educational performance categories. Cluster 0 corresponds to students with the lowest centroid and is interpreted as the *Low* performance cluster. Cluster 1, with intermediate centroid values, is interpreted as the *Moderate* cluster, while Cluster 2, which comprises students with the highest scores, is interpreted as the *High* performance cluster.

This semantic labeling allows the clusters to be understood not just as mathematical clusterings but as meaningful representations of conceptual mastery levels within an Outcome-Based Education (OBE) framework. The strength of fuzzy clustering lies in its ability to assign soft memberships, reflecting the fact that conceptual understanding is often a continuum rather than a set of rigid categories. For example, a student may exhibit 85% membership in the Moderate cluster and 15% in the Low cluster, suggesting a transitional level of understanding that may not be captured in crisp classification models.

3.3.2 Cluster Distribution and Balance

Once the cluster labels were finalized using the maximum membership rule, we examined the distribution of students across the three clusters. Of the 84 students analyzed, 15 students (17.9%) were assigned to the Low cluster (Cluster 0), 31 students (36.9%) to the Moderate cluster (Cluster 1), and 38 students (45.2%) to the High cluster (Cluster 2). This distribution suggests a right-skewed pattern of conceptual mastery, where nearly half of the students demonstrated strong performance, a substantial portion occupied an intermediate zone, and a smaller cluster showed signs of conceptual struggle.

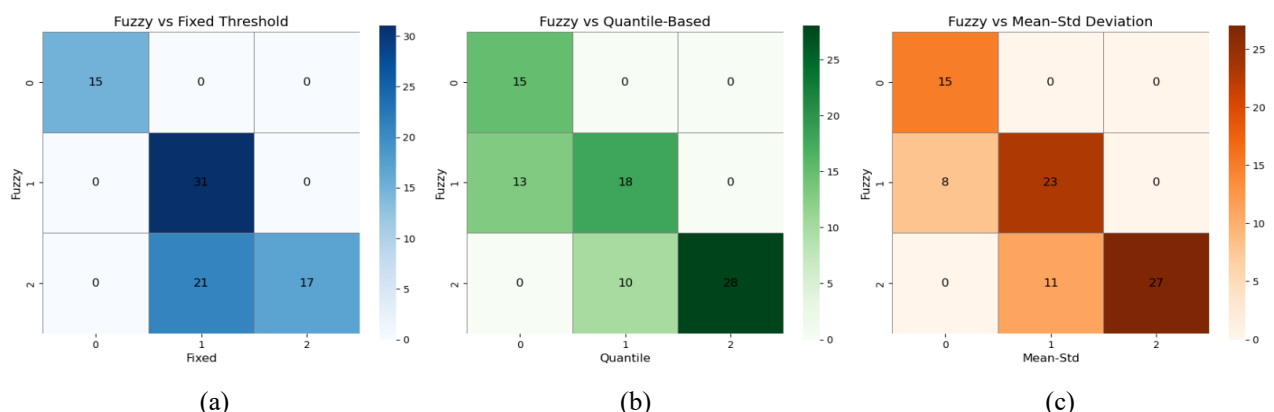


Figure 2. Heatmaps Comparing the Fuzzy Clustering Results with Three Manual Classification Schemes (a) Fuzzy vs Fixed Threshold, (b) Fuzzy vs Quantile-Based, (c) Fuzzy vs Mean-Std Deviation

Importantly, this distribution emerged organically from the data without relying on externally imposed thresholds. The resulting cluster sizes were neither artificially equalized nor disproportionately unbalanced, which strengthens the argument that the fuzzy clustering framework is well-suited to reflect actual variations in student understanding. It also affirms that the framework can produce educationally meaningful clusterings that go beyond traditional rigid banding schemes.

A comparison of cluster sizes across the different classification schemes further highlights the advantages of the fuzzy clustering model. The fuzzy method produced an organically balanced distribution of students—15 in the Low cluster, 31 in Moderate, and 38 in High—closely reflecting the actual performance dynamics of the cohort. In contrast, the fixed thresholding method assigned a disproportionately large number of students (52) to the Moderate category, suggesting a potential overgeneralization caused by rigid cut-off values. Meanwhile, the quantile-based approach produced an artificially uniform distribution (28 per cluster), disregarding meaningful differences in score density across the spectrum. The mean–standard deviation method, though more flexible, still resulted in a Moderate cluster (34) that exceeded both the Low (23) and High (27) categories, pointing to a bias toward central clustering. These differences are visualized in Figure 3, which compares cluster sizes across all classification approaches. Collectively, these discrepancies reaffirm the value of the fuzzy approach in preserving interpretive fidelity while remaining adaptive to the underlying data structure.

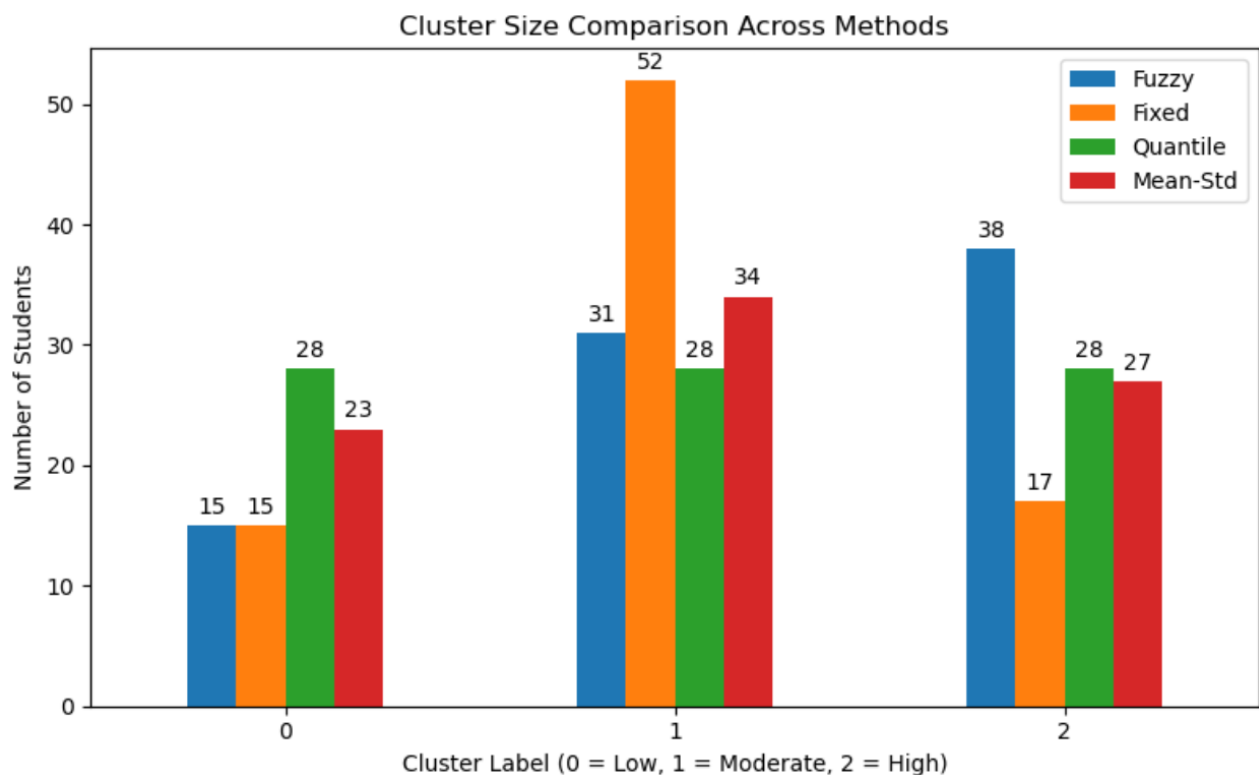


Figure 3. Cluster Size Comparison Across Fuzzy Clustering and Three Manual Classification Methods

3.3.3 Alignment with Manual Classification Schemes

To further validate the relevance of the fuzzy clusters, we compared the assigned fuzzy labels with three widely used manual classification schemes: fixed thresholding, quantile-based partitioning, and mean–standard deviation-based classification. The fixed threshold method followed institutional grading guidelines using absolute score cutoffs (e.g., ≤ 60 for Low, 61–80 for Moderate, and > 80 for High). In contrast, quantile-based and mean–standard deviation-based schemes are data-driven but make assumptions about score distribution shapes and boundaries.

Cross-tabulation analysis revealed a strong alignment between fuzzy clustering and the fixed threshold method. All 15 students assigned to the Low fuzzy cluster were also categorized as Low in the fixed threshold scheme. Similarly, all 31 students in the Moderate fuzzy cluster were mapped consistently to the Moderate band. The High fuzzy cluster showed the greatest variation, with 17 of the 38 students being labeled as High in the fixed threshold method, while 21 were categorized as Moderate—suggesting that fuzzy clustering captures transitional performance levels that rigid cutoffs fail to identify.

The comparison with the quantile-based scheme revealed greater divergence. Although all Low-cluster students were assigned to the lowest quantile, the Moderate and High clusters overlapped substantially with the middle quantile cluster. This outcome is expected given that quantile partitioning enforces equally sized clusters regardless of score distribution, thus obscuring meaningful conceptual gaps. In the case of the mean–standard deviation-based scheme, similar patterns were observed. Students in the Low fuzzy cluster were correctly mapped to the lower band, but those in the Moderate and High clusters were distributed across adjacent bands due to the symmetrical assumptions underlying this classification method.

These comparisons are illustrated in Fig. 2, which presents side-by-side heatmaps of the crosstab relationships between fuzzy clustering and each manual classification scheme. The heatmaps clearly show strong diagonal alignment with fixed thresholding, partial dispersion with the mean–standard deviation approach, and broader overlap in the quantile-based scheme. Overall, while manual classification methods provide a useful point of reference, they are limited by their reliance on fixed or statistical thresholds that may not align with the latent structure of student performance. Fuzzy clustering, in contrast, reveals overlapping boundaries and transitional learners, offering a richer and more flexible model for conceptual interpretation and formative assessment.

3.4 Comparative Evaluation with Manual Classification Schemes

To further evaluate the reliability and practical value of the fuzzy clustering approach, a comparative analysis was conducted against three conventional manual classification schemes. Each scheme is commonly used in educational assessment to categorize students into performance levels, albeit based on different underlying assumptions. The fixed thresholding scheme employs absolute cutoffs—typically scores of 60 and 80—to classify students into Low, Moderate, and High performance bands. This approach is aligned with institutional grading policies and is straightforward to implement. However, it imposes rigid boundaries that may not reflect the actual distribution or dynamics of student learning.

The quantile-based scheme partitions the dataset into equal-sized clusters using empirical tertiles (33.3rd and 66.7th percentiles). This method ensures that each cluster contains roughly the same number of students, which is statistically convenient but often pedagogically arbitrary. Because it does not consider conceptual mastery thresholds or instructional benchmarks, it may cluster together students with substantially different learning profiles. The third scheme, mean–standard deviation-based classification, categorizes students relative to the dataset’s central tendency. It defines Moderate performance as lying within ± 0.5 standard deviations from the mean, with Low and High categories assigned to scores below and above this range, respectively. This method adjusts to the shape and spread of the data but may struggle with skewed or multimodal distributions, which are common in educational datasets.

When cross-tabulated with the fuzzy clustering results, the fixed threshold scheme showed the highest alignment, particularly for the Low and Moderate categories. All 15 students in the fuzzy Low cluster were also classified as Low by the fixed threshold scheme. Similarly, all 31 students in the fuzzy Moderate cluster aligned with the Moderate category. The fuzzy High cluster showed partial overlap: 17 students matched with the fixed High category, while 21 were assigned to Moderate. This suggests that fuzzy clustering was able to identify nuanced conceptual elevation that extended beyond rigid cutoff points. In contrast, the quantile-based classification displayed more fragmentation. Although all students in fuzzy Cluster 0 remained in the lowest quantile, those in Cluster 1 were split between the first and second quantiles. Meanwhile, Cluster 2 spanned both the second and third quantiles. This pattern illustrates how the quantile method, by enforcing equal cluster sizes, may mask meaningful conceptual differences—especially among students near boundaries.

The mean–standard deviation approach demonstrated a similar partial alignment. While fuzzy Cluster 0 mapped neatly to the Low cluster, students in Cluster 1 and 2 were distributed across the Moderate and High categories. Notably, some students in fuzzy Cluster 2—indicative of high conceptual performance—were placed in the Moderate band by this method, likely due to the compressing effect of the standard deviation window. These results emphasize a key strength of the fuzzy clustering framework: its ability to model gradual transitions and uncertainty in student learning. Unlike the manual schemes, which force all students into discrete and mutually exclusive categories, fuzzy clustering acknowledges and represents the ambiguous or transitional nature of conceptual development. For example, a student with 85% membership in the High cluster and 15% in Moderate may be confidently considered a strong performer, while one with a 55–45% split between Moderate and High reflects a student whose performance is evolving—information that is critically important for formative assessment and targeted instructional support.

Furthermore, fuzzy clustering does not rely on externally imposed cutoffs or statistical heuristics. It derives cluster boundaries and membership degrees directly from the data, thus reflecting the intrinsic structure of student performance. This makes it both flexible and explainable—qualities that are essential in educational contexts where individual variation, instructional context, and longitudinal development play significant roles. In conclusion, while manual classification schemes offer familiarity and simplicity, they are inherently limited in their capacity to capture the full spectrum of student understanding. Fuzzy clustering, by contrast, provides a data-driven, interpretable, and pedagogically meaningful alternative that enhances our ability to assess, differentiate, and support student learning in a nuanced and dynamic way. To complement the qualitative comparisons, we computed two widely recognized clustering agreement metrics—Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI)—to quantitatively evaluate the alignment between the fuzzy clustering results and each manual classification scheme. These metrics provide an objective measure of similarity that corrects for chance agreement (ARI) and quantifies shared information content (NMI). The results are summarized in [Table 3](#).

Table 3. Quantitative Agreement Between Fuzzy Clustering and Manual Classification Schemes.

Comparison	ARI	NMI
Fuzzy vs Fixed	0.405	0.629
Fuzzy vs Quantile	0.437	0.550
Fuzzy vs Mean–Std	0.462	0.569

As shown in [Table 3](#), the highest ARI (0.462) was obtained in the comparison between fuzzy clustering and the mean–standard deviation scheme, indicating a moderate level of agreement in cluster assignments. The fixed thresholding scheme, while pedagogically interpretable, exhibited the lowest ARI (0.405), suggesting a larger divergence in exact label assignments. However, it yielded the highest NMI (0.629), implying that although label assignments may differ, the general information structure is relatively well preserved.

These results support the earlier visual findings: while fuzzy clustering captures underlying conceptual structure, its flexible and overlapping nature inevitably leads to partial mismatches with rigid schemes. Therefore, the modest ARI scores are not indicative of weakness but rather reflect the greater granularity and nuance provided by the fuzzy model. Taken together, the combination of ARI and NMI validates the structural relevance of fuzzy clustering while reinforcing its explainability advantage over conventional deterministic methods.

These findings parallel the results reported by [28] in *Expert Systems with Applications*, where fuzzy clustering—combined with a transition-matrix analysis—was shown to reveal latent learner transitions and performance gradients that traditional deterministic methods failed to capture. Similarly, in the present study, fuzzy clusters expose nuanced conceptual boundaries and identify transitional learners whose memberships straddle Moderate and High clusters. This behavior cannot be reproduced by fixed-threshold, quantile-based, or mean–standard deviation-based approaches, which inherently enforce rigid, non-overlapping categories.

From an Outcome-Based Education (OBE) perspective, the fuzzy clustering results provide clearer alignment with the progressive and cumulative nature of CLO achievement. OBE emphasizes developmental progression rather than categorical separation, and the partial memberships generated by fuzzy clustering map naturally onto this philosophy. Students with mixed memberships represent those who are gradually advancing their conceptual understanding—an insight critical for formative assessment and targeted intervention. Thus, fuzzy clustering not only produces mathematically coherent partitions but also yields pedagogically interpretable structures that support evidence-based instructional decision-making.

3.5 Visualization and Explainability

In the context of explainable artificial intelligence (XAI) and educational data mining, the ability to visualize and interpret the structure of data-driven models is essential—particularly when the objective is to support formative assessment and personalized feedback. Accordingly, this study incorporates several forms of visual analysis to reinforce the explainability of the fuzzy clustering results and to provide deeper insight into how students are clustered based on conceptual performance levels.

The first layer of explainability is presented through membership-based cluster visualization. By leveraging the soft membership matrix U produced by the Fuzzy C-Means algorithm, each student's degree of association with all three clusters (Low, Moderate, High) can be examined. This allows us to move beyond

binary classification and instead understand each student's conceptual profile on a continuum. Students located near cluster boundaries—those with non-trivial membership values in more than one cluster—can be flagged as transitional learners who may benefit from targeted instructional interventions.

To illustrate this, a stacked bar chart is provided in Figure 4, showing the fuzzy membership degrees for all students across the three clusters. Students with strong dominance in one cluster appear with a single dominant color, while transitional students exhibit bars with more balanced color segments, indicating overlapping conceptual mastery. **Fig. 4** Stacked bar chart showing fuzzy membership degrees of each student across three conceptual clusters: Low (green), Moderate (orange), and High (blue). This visualization illustrates the soft nature of fuzzy clustering and helps identify students with overlapping or transitional performance profiles.

To complement this view, **Fig. 5** presents a set of scatter plots depicting the relationship between final scores and degrees of membership to each fuzzy cluster. These plots reveal the smooth and continuous transitions across cluster boundaries, with students at mid-score ranges showing partial memberships in adjacent clusters. Such visualizations help in identifying "cusp learners"—students whose performance lies on the threshold between two conceptual bands—and provide a richer context for formative assessment.

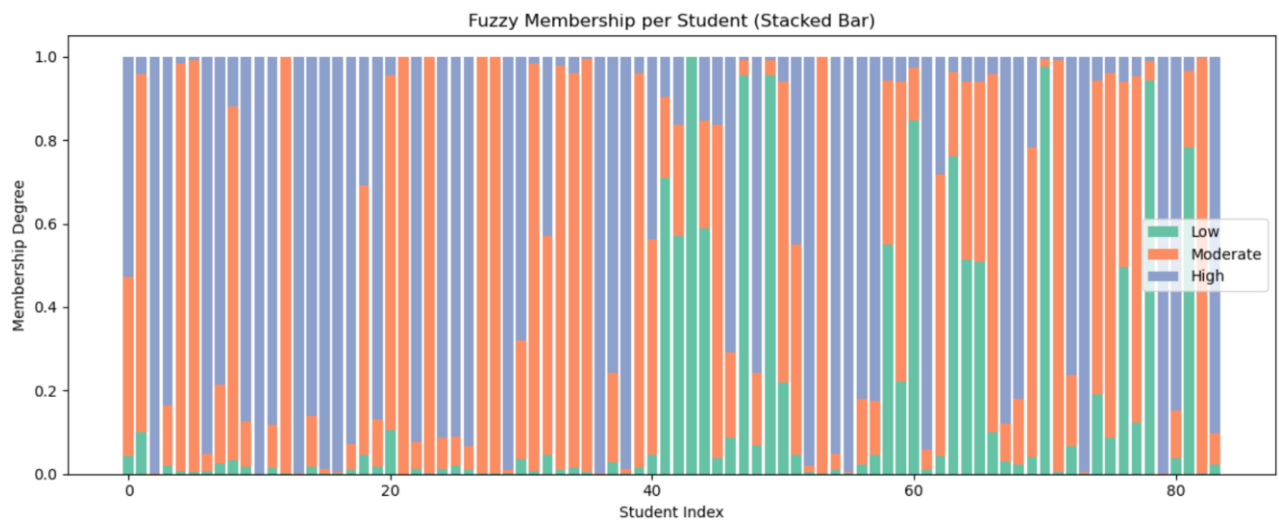


Figure 4. Stacked Bar Chart Showing Fuzzy Membership Degrees of Each Student Across Three Conceptual Clusters

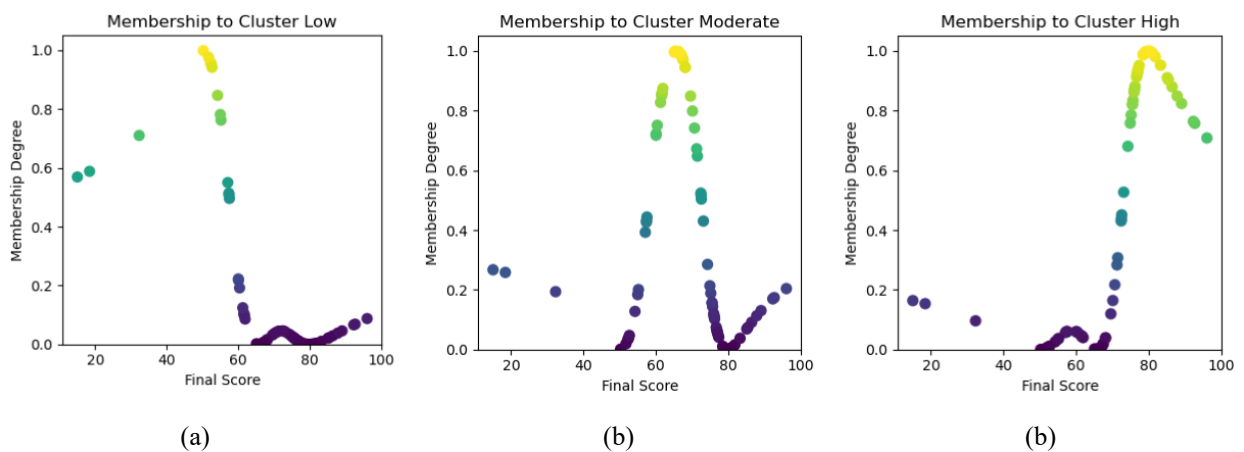


Figure 5. Scatter Plots Showing the Relationship Between Students' Final Scores and Their Membership Degrees to Each Fuzzy Cluster

(a) Membership to Cluster Low, (b) Membership to Cluster Moderate, (c) Membership to Cluster High

Beyond student-specific profiles, the study also presents cross-tabulation heatmaps comparing the fuzzy clustering labels with three manual classification schemes: fixed thresholding, quantile-based partitioning, and mean–standard deviation-based classification. These visualizations, shown earlier in **Fig. 2**, clearly illustrate the areas of alignment and misalignment between the fuzzy and manual classifications. For instance, the heatmap comparing fuzzy clustering and fixed thresholding reveals a strong diagonal structure, indicating high agreement across all performance levels. In contrast, the heatmaps for quantile and mean–

standard deviation methods show more dispersed patterns, highlighting the fuzziness and overlap in conceptual boundaries that those rigid methods fail to capture.

Additionally, Fig. 3 presents a bar chart comparison of cluster sizes across all classification methods. This plot not only confirms the numerical differences reported earlier but also makes them visually salient. The fuzzy clustering model yields a naturally balanced distribution of cluster sizes, whereas the fixed thresholding method produces an overrepresented Moderate cluster. The quantile-based approach enforces uniformity by design, and the mean–standard deviation model tends to concentrate students toward the middle cluster. This visual comparison reinforces the argument that the fuzzy clustering model reflects the intrinsic structure of the dataset more faithfully than its manual counterparts.

These visualization techniques serve a dual purpose. On one hand, they act as validation tools for evaluating the internal coherence and external alignment of the clustering results. On the other hand, they function as communication tools, enabling educators, researchers, and stakeholders to intuitively grasp the structure of student learning trajectories without needing to interpret raw numerical outputs or dense mathematical formulations. From a pedagogical standpoint, visualizing membership degrees provides actionable insights. Educators can identify students who are near the borders between clusters and may be at risk of regression or in need of enrichment. For example, a student with membership values of 0.48 in Moderate and 0.47 in High might not be flagged by traditional grading schemes but is clearly on the cusp of performance elevation. Such insights are particularly valuable in formative contexts where real-time intervention can significantly alter learning outcomes. In conclusion, the visualizations embedded in this study enhance the transparency and interpretability of the fuzzy clustering framework. They illustrate not only *what* the model has learned from the data, but also *why* certain students are clustered together, and *how* these clusterings relate to conventional assessment strategies. These qualities position fuzzy clustering not merely as a predictive technique, but as an interpretable, teacher-friendly diagnostic tool for advancing student learning in Outcome-Based Education environments.

4. CONCLUSION

This study has introduced an explainable fuzzy clustering framework for modeling student learning trajectories in an Outcome-Based Education (OBE) context. By applying the Fuzzy C-Means algorithm with optimal cluster determination via the Tang–Sun–Sun (TSS) index, the framework offers a data-driven alternative to manual classification. The resulting three clusters—Low, Moderate, and High performance—naturally reflect pedagogical categories while also accommodating transitional learners through soft membership, thus capturing the nuanced continuum of student understanding. Comparative analysis with threshold-based, quantile, and statistical banding methods revealed that conventional schemes often obscure meaningful variation in student performance. In contrast, the fuzzy approach preserves gradation and semantic interpretability. Visualizations further supported the model's transparency, offering educators intuitive insights into individual and group learning profiles. In sum, this study positions fuzzy clustering as a pedagogically grounded and interpretable technique for student assessment, supporting fairer and more responsive educational decision-making. Beyond its clustering accuracy, the framework also enhances explainability in educational assessment. The soft membership values provide a transparent rationale behind each student's placement, allowing instructors to identify borderline or transitional learners whose conceptual mastery is still evolving. This contrasts sharply with rigid deterministic schemes, where cut-off boundaries obscure meaningful learning gradients. By revealing how close or distant a student is from adjacent conceptual levels, the model supports more informed formative feedback, equity-driven decision-making, and targeted instructional intervention that aligns directly with OBE principles. This study is limited by its focus on a single-course dataset collected within one semester, which constrains the generalizability of the clustering structure across broader instructional contexts. The one-dimensional aggregated CLO score also limits the ability to capture longitudinal patterns of conceptual development. Future work should incorporate multi-course, multi-semester datasets to construct more authentic learning trajectories and examine how conceptual mastery evolves across curricular stages. Further extensions may include integrating fuzzy transition matrices or temporal fuzzy models to characterize dynamic learning progress more comprehensively.

Author Contributions

Rustam: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing-Original Draft, Writing-Review and Editing. Koredianto Usman: Validation, Writing-Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

The authors gratefully acknowledge the financial support provided by Telkom University with research grant number is RCMS/RES/Penelitian Internal/15187/4/2025.

Acknowledgment

The authors thank the anonymous reviewers for their valuable comments and suggestions.

Declarations

The authors declare that there is no conflict of interest regarding the publication of this article.

Declaration of Generative AI and AI-assisted Technologies

Generative AI tools were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

- [1] M. M. M. Syeed, A. S. M. Shihavuddin, M. F. Uddin, M. Hasan, and R. H. Khan, "OUTCOME BASED EDUCATION (OBE): DEFINING THE PROCESS AND PRACTICE FOR ENGINEERING EDUCATION," *IEEE Access*, vol. 10, pp. 119170–119192, 2022. doi: <https://doi.org/10.1109/ACCESS.2022.3219477>
- [2] R. Katawazai, "IMPLEMENTING OUTCOME-BASED EDUCATION AND STUDENT-CENTERED LEARNING IN AFGHAN PUBLIC UNIVERSITIES: THE CURRENT PRACTICES AND CHALLENGES," *Heliyon*, vol. 7, no. 5, 2021. doi: <https://doi.org/10.1016/j.heliyon.2021.e07076>
- [3] A. W. Clemmons, D. A. Donovan, E. J. Theobald, and A. J. Crowe, "USING THE INTENDED–ENACTED–EXPERIENCED CURRICULUM MODEL TO MAP THE VISION AND CHANGE CORE COMPETENCIES IN UNDERGRADUATE BIOLOGY PROGRAMS AND COURSES," *CBE—Life Sciences Education*, vol. 21, no. 1, p. ar6, 2022. doi: <https://doi.org/10.1187/cbe.21-02-0054>
- [4] Q. I. Ali, "TOWARDS MORE EFFECTIVE SUMMATIVE ASSESSMENT IN OBE: A NEW FRAMEWORK INTEGRATING DIRECT MEASUREMENTS AND TECHNOLOGY," *Discover Education*, vol. 3, no. 1, p. 107, 2024. doi: <https://doi.org/10.1007/s44217-024-00208-5>
- [5] J. Cain, M. Medina, F. Romanelli, and A. Persky, "DEFICIENCIES OF TRADITIONAL GRADING SYSTEMS AND RECOMMENDATIONS FOR THE FUTURE," *Am J Pharm Educ*, vol. 86, no. 7, p. 8850, 2022. doi: <https://doi.org/10.5688/ajpe8850>
- [6] Z. Li, Z. A. Pardos, and C. Ren, "ALIGNING OPEN EDUCATIONAL RESOURCES TO NEW TAXONOMIES: HOW AI TECHNOLOGIES CAN HELP AND IN WHICH SCENARIOS," *Comput Educ*, vol. 216, p. 105027, 2024. doi: <https://doi.org/10.1016/j.compedu.2024.105027>
- [7] Q. Zhang, S. Chen, and W. Liu, "BALANCED KNOWLEDGE TRANSFER IN MTTL-CLINICALBERT: A SYMMETRICAL MULTI-TASK LEARNING FRAMEWORK FOR CLINICAL TEXT CLASSIFICATION," *Symmetry (Basel)*, vol. 17, no. 6, p. 823, 2025. doi: <https://doi.org/10.3390/sym17060823>
- [8] D. C. Youvan, "THE FUZZY-HOTT CONSCIOUSNESS MODEL: A HYBRID FRAMEWORK FOR UNDERSTANDING AMBIGUITY AND STRUCTURE IN HUMAN THOUGHT," 2024.
- [9] J. Stanja, W. Gritz, J. Krugel, A. Hoppe, and S. Dannemann, "FORMATIVE ASSESSMENT STRATEGIES FOR STUDENTS' CONCEPTIONS—THE POTENTIAL OF LEARNING ANALYTICS," *British Journal of Educational Technology*, vol. 54, no. 1, pp. 58–75, 2023. doi: <https://doi.org/10.1111/bjet.13288>
- [10] A. E. Ezugwu *et al.*, "A COMPREHENSIVE SURVEY OF CLUSTERING ALGORITHMS: STATE-OF-THE-ART MACHINE LEARNING APPLICATIONS, TAXONOMY, CHALLENGES, AND FUTURE RESEARCH PROSPECTS," *Eng Appl Artif Intell*, vol. 110, p. 104743, 2022. doi: <https://doi.org/10.1016/j.engappai.2022.104743>
- [11] S. E. Hashemi, F. Gholian-Jouybari, and M. Hajiaghahi-Keshteli, "A FUZZY C-MEANS ALGORITHM FOR OPTIMIZING DATA CLUSTERING," *Expert Syst Appl*, vol. 227, p. 120377, 2023. doi: <https://doi.org/10.1016/j.eswa.2023.120377>
- [12] E. Stamkou, D. Keltner, R. Corona, E. Aksoy, and A. S. Cowen, "EMOTIONAL PALETTE: A COMPUTATIONAL MAPPING OF AESTHETIC EXPERIENCES EVOKED BY VISUAL ART," *Sci Rep*, vol. 14, no. 1, p. 19932, 2024. doi: <https://doi.org/10.1038/s41598-024-69686-9>

- [13] U. Ahmed *et al.*, "SIGNATURE-BASED INTRUSION DETECTION USING MACHINE LEARNING AND DEEP LEARNING APPROACHES EMPOWERED WITH FUZZY CLUSTERING," *Sci Rep*, vol. 15, no. 1, p. 1726, 2025. doi: <https://doi.org/10.1038/s41598-025-85866-7>
- [14] A. Y. Lee and J. T. Hancock, "DEVELOPING DIGITAL RESILIENCE: AN EDUCATIONAL INTERVENTION IMPROVES ELEMENTARY STUDENTS' RESPONSE TO DIGITAL CHALLENGES," *Computers and Education Open*, vol. 5, p. 100144, 2023. doi: <https://doi.org/10.1016/j.caeo.2023.100144>
- [15] X. Wang, Y. Chen, J. Jin, and B. Zhang, "FUZZY-CLUSTERING AND FUZZY NETWORK BASED INTERPRETABLE FUZZY MODEL FOR PREDICTION," *Sci Rep*, vol. 12, no. 1, p. 16279, 2022. doi: <https://doi.org/10.1038/s41598-022-20015-y>
- [16] M. A. Uddin, Y. Yoon, M. H. Head, and Q. O. Abiona, "GROUPING FACTORS IN BRIDGE COMPONENT DETERIORATION: A DATA-DRIVEN COMPARATIVE ANALYSIS USING MULTIFACETED SIMILARITY MEASURES," *Journal of Bridge Engineering*, vol. 30, no. 6, p. 4025031, 2025. doi: <https://doi.org/10.1061/JBENF2.BEENG-7016>
- [17] A. M. Ikotun, F. Habyarimana, and A. E. Ezugwu, "CLUSTER VALIDITY INDICES FOR AUTOMATIC CLUSTERING: A COMPREHENSIVE REVIEW," *Heliyon*, 2025. doi: <https://doi.org/10.1016/j.heliyon.2025.e41953>
- [18] S. Zhu, L. Xu, and E. D. Goodman, "HIERARCHICAL TOPOLOGY-BASED CLUSTER REPRESENTATION FOR SCALABLE EVOLUTIONARY MULTIOBJECTIVE CLUSTERING," *IEEE Trans Cybern*, vol. 52, no. 9, pp. 9846–9860, 2021. doi: <https://doi.org/10.1109/TCYB.2021.3081988>
- [19] F. Wang, S. Qing, C. Sa, Q. Lai, and A. Chang, "REMOTE SENSING EVALUATION OF TROPHIC STATUS IN THE DAIHAI LAKE BASED ON FUZZY CLASSIFICATION," *Water (Basel)*, vol. 16, no. 21, p. 3032, 2024. doi: <https://doi.org/10.3390/w16213032>
- [20] 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*, vol. 12, no. 19, p. 9467, 2022. doi: <https://doi.org/10.3390/app12199467>
- [21] Á. López-Oriona, J. A. Vilar, and P. D'Urso, "QUANTILE-BASED FUZZY CLUSTERING OF MULTIVARIATE TIME SERIES IN THE FREQUENCY DOMAIN," *Fuzzy Sets Syst*, vol. 443, pp. 115–154, 2022. doi: <https://doi.org/10.1016/j.fss.2022.02.015>
- [22] J. A. Supovitz, C. B. Ebby, and G. Collins, "PATHWAYS TO PERFORMANCE: THE EXPERIMENTAL IMPACTS OF LEARNING TRAJECTORY-ORIENTED FORMATIVE ASSESSMENT IN MATHEMATICS," *American Journal of Education*, vol. 130, no. 4, pp. 621–651, 2024. doi: <https://doi.org/10.1086/730994>
- [23] C. Cammies, J. A. Cunningham, and R. K. Pike, "NOT ALL BLOOM AND GLOOM: ASSESSING CONSTRUCTIVE ALIGNMENT, HIGHER ORDER COGNITIVE SKILLS, AND THEIR INFLUENCE ON STUDENTS' PERCEIVED LEARNING WITHIN THE PRACTICAL COMPONENTS OF AN UNDERGRADUATE BIOLOGY COURSE," *J Biol Educ*, vol. 58, no. 3, pp. 588–608, 2024. doi: <https://doi.org/10.1080/00219266.2022.2092191>
- [24] B. Albreiki, T. Habuza, and N. Zaki, "FRAMEWORK FOR AUTOMATICALLY SUGGESTING REMEDIAL ACTIONS TO HELP STUDENTS AT RISK BASED ON EXPLAINABLE ML AND RULE-BASED MODELS," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 49, 2022. doi: <https://doi.org/10.1186/s41239-022-00354-6>
- [25] Rustam *et al.*, "MODIFIED POSSIBILISTIC FUZZY C-MEANS ALGORITHM FOR CLUSTERING INCOMPLETE DATA SETS," *Acta Polytechnica*, vol. 61, no. 2, pp. 364–377, 2021. doi: <https://doi.org/10.14311/AP.2021.61.0364>
- [26] Rustam, A. Y. Gunawan, and M. T. A. P. Kresnowati, "DATA DIMENSIONALITY REDUCTION TECHNIQUE FOR CLUSTERING PROBLEM OF METABOLOMICS DATA," *Heliyon*, vol. 8, no. 6, 2022. doi: <https://doi.org/10.1016/j.heliyon.2022.e09715>
- [27] Rustam and K. Usman, "A NOVEL FUZZY CLUSTERING FRAMEWORK WITH MANHATTAN DISTANCE AND WEIGHTED MEDIAN CENTROIDS FOR OUTLIER-RESILIENT DATA ANALYSIS," in *IEEE Access*, vol. 13, pp. 190964–190979, 2025, doi: <https://doi.org/10.1109/ACCESS.2025.3629679>
- [28] Rustam, D. N. Anggraini, K. Usman, and L. Gaur, "A NOVEL FUZZY CLUSTERING APPROACH WITH TRANSITION MATRIX FOR EXPLAINABLE EVALUATION OF SOCIAL MEDIA-BASED DIGITAL LITERACY INTERVENTIONS," *Expert Systems with Applications*, vol. 298, p. 129769, 2026, doi: <https://doi.org/10.1016/j.eswa.2025.129769>
- [29] Rustam, A. Y. Gunawan, and M. T. A. P. Kresnowati, "ARTIFICIAL NEURAL NETWORK APPROACH FOR THE IDENTIFICATION OF CLOVE BUDS ORIGIN BASED ON METABOLITES COMPOSITION," *Acta Polytechnica*, vol. 60, no. 5, pp. 440–447, 2020, doi: <https://doi.org/10.14311/AP.2020.60.0440>
- [30] Rustam, A. Y. Gunawan, and M. T. A. P. Kresnowati, "THE HARD C-MEANS ALGORITHM FOR CLUSTERING INDONESIAN PLANTATION COMMODITY BASED ON METABOLITES COMPOSITION," in *Journal of Physics: Conference Series*, vol. 1315, p. 012085, 2019, IOP Publishing, doi: <https://doi.org/10.1088/1742-6596/1315/1/012085>
- [31] G. A. Salsabila, S. Sa'idah, and R. Rustam, "ANEMIA DETECTION THROUGH CONJUNCTIVA ON EYES USING HISTOGRAM OF ORIENTED GRADIENTS AND DECISION TREE CLASSIFICATION," in *Proc. 2022 8th Int. Conf. on Science and Technology (ICST)*, Yogyakarta, Indonesia, 2022, pp. 1–6, doi: <https://doi.org/10.1109/ICST56971.2022.10136290>
- [32] Rustam, "PRETREATMENT METHODS FOR ENHANCING MACHINE LEARNING PERFORMANCE ON METABOLOMICS DATA," *IEEE Access*, vol. 13, pp. 80133–80148, 2025, doi: <https://doi.org/10.1109/ACCESS.2025.3567153>