

Adaptive Fuzzy Clustering with Granular Refinement and Triangulation

Anis Farhani Abdul Ghafar*, Syarifah Sakinah Syed Ahmad

Faculty of Information and Communications Technology, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia

Abstract—Traditional fuzzy C-means (FCM) typically uses a fixed Euclidean distance, which may be restrictive for heterogeneous or overlapping data. This working manuscript examines GFASM-FCM, which combines iteration-wise metric selection, ambiguity-focused membership sharpening, and lower-approximation prototype refinement. The originally reported 600-point synthetic results were Silhouette 0.82, DBI 0.23, PC 0.66, CE 0.60, ARI 0.77, NMI 0.88, and Purity 0.93; these are prior point estimates rather than independently revalidated 30-run summaries. The original point-centroid triangle-angle test is equivalent to distance ordering under a common metric and does not provide independent geometric evidence. Lower sets with membership threshold 0.60 are disjoint, so overlap control cannot be attributed to their Jaccard similarity. Broader claims about real-data generalization, statistically significant improvements, and timing superiority remain conditional on matched experimental results. The main contribution examined here is the sequential design; its incremental empirical validity requires further verification.

Keywords—Adaptive distance metric; fuzzy C-means; fuzzy clustering; granular computing; overlap verification; triangulation; unsupervised learning

I. INTRODUCTION

Clustering is one of the most popular and useful tools to explore structure in unlabeled data. The aim is straightforward - group observations similar to each other and separate those different from each other - but the idea of similarity is not fixed by the data alone. It depends on the representation, the proximity measure, and the rule to form the partition. For this reason, the geometry induced by a distance measure can strongly influence the final clustering result [3], [9]. Real datasets rarely follow one ideal shape. Compact regions can be found between elongated groups, correlated features might appear in a small part of the space, local densities can vary, and some observations may be between credible clusters. In this scenario, one global distance measure may be correct in one part of the space and wrong in another. Fuzzy C-Means (FCM) is derived from the fuzzy partitioning work of Dunn and Bezdek [1], [2], [33] and is well suited for clusters where the boundaries between clusters are not sharp or very close. Rather than assigning each observation to a single cluster, FCM estimates a set of memberships that sum to one. This soft representation has made the method useful in pattern recognition, image analysis, biomedical exploration, and other problems where the two clusters may overlap. The difficulty is that fuzzy membership still depends on the distances in the update equations. Most traditional FCM is based on Euclidean distance, which is simple and efficient but naturally favours roughly spherical groups with similar sizes. When

covariance, anisotropy, or unequal feature behaviour is important, the membership distribution can be systematically biased [5], [8], [24]. The same problem appears at the edges of clusters. For example, an observation could get two similar memberships, and both of them are mathematically valid. If that point contributes almost equally to two centroids, however, then the boundary can be blurred, and the prototypes may be pulled to an uncertain region. A wide range of FCM extensions have been proposed to capture noise, nonlinear structure, spatial structure, atypical observations, or uncertainty [4]-[6], [18]-[23], [31]. These methods have been valuable, but most of them are based on a single source of difficulty. Kernel methods change the similarity space, spatial variants provide neighbourhood information, and rough or granular approaches focus on uncertainty regions.

This study examines an iteration-wise fuzzy metric selection stage followed by selective membership sharpening and lower-approximation prototype refinement. The switching decision is global within each iteration, not region-specific; therefore, it should not be described as assigning different metrics simultaneously to different clusters. Further, the original triangle-angle comparison restates the distance ordering under a common metric and should be treated as an interpretive visualization rather than independent verification. The granular lower-set update emphasizes stable members, while claims involving boundary overlap depend on whether an actual overlap-dependent operation is present in the MATLAB code. These distinctions motivate the controlled ablations and reproduction requirements described later.

II. RELATED WORK AND RESEARCH GAP

A. Fuzzy C-Means and Distance Dependence

Classical FCM minimizes the dispersion between observations and cluster prototypes [1], [2]. Its strength is the soft partition itself: a point near a boundary can legitimately share membership across several clusters. But the quality of that partition is still closely tied to the distance model. If the representation is not in good agreement with the geometry, the distortion is passed to the membership matrix. Several extensions have attempted to reduce this dependence. Kernel FCM maps the data into a nonlinear feature space and can improve separation when the original geometry is difficult to describe with a simple centroid model [5]. The trade-off is that kernel choice and parameter tuning add more sensitivity and computation. Distance-aware formulations are more direct by changing the membership or distance to be the objective [7]. Hadi [8] further showed that separation measures for multivariate clusters should account for differences in location,

*Corresponding author.

elliptical shape, and direction rather than assuming a common spherical geometry.

Mahalanobis distance is attractive in this context because covariance and feature correlation are built into the measure, and it has been incorporated into fuzzy clustering for non-spherical image structures [24]. Minkowski distance provides a different type of flexibility because its order parameter controls the norm. Together, these observations suggest that the more useful question is not which single metric is universally best, but which metric is most defensible for the local structure observed at a given stage of clustering. Model-based clustering provides a complementary probabilistic approach to grouping observations and is useful context for comparisons involving Gaussian mixture models [10].

Covariance-sensitive fuzzy clustering has long been investigated as an extension of conventional FCM. Gustafson and Kessel introduced a fuzzy clustering approach incorporating adaptive covariance-based distance measures, allowing clusters with different orientations and elliptical structures to be represented more effectively. Gath and Geva subsequently developed a fuzzy clustering formulation that incorporates cluster-specific distributional characteristics, including covariance and density information. These approaches demonstrate that adapting cluster geometry is an established direction in fuzzy clustering research. Unlike methods that incorporate covariance information directly into their clustering formulation, the proposed FASM mechanism evaluates several candidate distance measures and selects an active metric during the iterative clustering process. Accordingly, its intended contribution lies in the metric-switching policy rather than the introduction of covariance-adaptive distance modelling itself [34]-[35].

B. Robust, Spatial, and Kernelized Fuzzy Clustering

Another major line of work has focused on robustness and contextual information. Askari [6] proposed a noise-resistant fuzzy clustering method, while Wang et al. [31] combined spatial and intensity constraints with membership linking for noisy image segmentation. Spatial-context models similarly use neighbouring observations to stabilize pixel assignments [23], and adaptive weighted spatial FCM has been applied to retinal-vessel segmentation [22]. More recent studies have compared noise-estimation strategies for preserving image structure under contamination [19]. Kernel possibilistic FCM combines nonlinear representation with a possibilistic treatment of atypical observations [20]. Beyond fuzzy clustering, image segmentation research has explored probabilistic bottom-up aggregation with cue integration [28] and contour-driven hierarchical segmentation [29]; these approaches provide broader context for boundary-aware image analysis rather than direct experimental baselines in this study.

These studies show that useful clustering decisions often require information beyond a raw Euclidean point-to-centroid distance. At the same time, many of the proposed mechanisms are specialized for a particular setting such as noisy images, spatial adjacency, or a selected similarity construction. Sparsity and multiple-kernel approaches offer another direction. Chen et al. [21] incorporated principal-component embedding into sparse FCM, while Tang et al. [18] proposed knowledge-

induced multiple-kernel fuzzy clustering. These approaches can be highly effective when an appropriate representation or kernel family is available.

GFASM-FCM takes a deliberately different approach. It keeps the centroid-based FCM structure for transparency and works with a small set of familiar distance measures. The adaptation is driven by observable local descriptors and a compact fuzzy rule base, allowing the reason for each metric switch to remain visible rather than being hidden inside a large learned representation. A recent review of rapid clustering methods also highlights the need to consider computational trade-offs and evaluation conditions when comparing clustering techniques [30].

C. Granular and Rough-Fuzzy Refinement

Granular computing is an interesting way to represent data at different levels of certainty and abstraction. In fuzzy clustering, information granules can distinguish reliable regions from uncertain boundaries instead of treating every membership value as equally informative. Hu et al. [11] linked fuzzy clustering to knowledge extraction and granulation, and Li et al. [12] used improved justifiable granularity to support feature selection. Pedrycz [13] investigated clustering in an order-2 information granule environment. Yang et al. [14] developed a linear fuzzy information-granule FCM method for time-series clustering. However, in [15], the fuzzy and rough notions of local boundary fuzzy rough were combined with the idea of justifiable granularity. Recently, fuzzy rough and granular prototype methods have been developed. FRCM combines fuzzy and rough ideas in a C-means framework [16], and Wang et al. [17] enrich a soft partition with granular prototypes. Boundary-aware granular models are particularly important for FCM because two similar memberships are not necessarily a mistake, and could represent a genuine transition region. For example, Zeng et al. [32] include local border points in fuzzy entropy clustering. Hence, in the present study, granular refinement plays a limited role. It separates stable members from ambiguous ones, uses this distinction when refining the prototypes, and preserves uncertainty in the final fuzzy representation.

Earlier research by Maji and Pal introduced Rough Fuzzy C-Means (RFCM), combining rough-set approximations with fuzzy clustering to distinguish more certain cluster members from observations located in overlapping or uncertain regions. This established a foundation for incorporating lower and boundary approximations into fuzzy-clustering procedures. The granular refinement component of GFASM-FCM builds upon these principles by applying lower-approximation-based prototype refinement after adaptive metric switching and ambiguity handling. Therefore, the contribution of the present framework is not the introduction of rough-fuzzy approximations, but their integration into the proposed sequential clustering procedure [36].

D. Geometric Verification and Research Gap

Geometric relationships provide a useful means of examining uncertainty in clustering, particularly when observations lie near the boundaries of neighbouring clusters. In traditional FCM, cluster associations are determined through distance-dependent membership values. Although fuzzy

memberships allow observations to belong partially to several clusters, competing membership values may introduce uncertainty into prototype estimation. This motivates a selective mechanism for identifying and managing ambiguous observations without removing the fuzzy characteristics of the partition.

In the proposed GFASM-FCM framework, triangulation is incorporated after the adaptive metric-switching stage to examine observations associated with two competing cluster centroids. Following the triangulation concept developed in this research, an ambiguous observation and its two nearest candidate centroids form a point-centroid-centroid triangle. The construction represents the spatial relationship between the observation and neighbouring cluster centres, allowing potentially redundant memberships to be identified and considered for adjustment before granular refinement.

The original triangulation formulation compares the sum of the angle at the observation and the angle at each candidate centroid. However, this angular comparison has an important mathematical limitation. When all three sides are measured within the same metric space, the ordering of the centroid angles is determined by the corresponding point-to-centroid distances. Therefore, the angle comparison does not provide independent evidence beyond the distance information already used in the fuzzy membership calculation. Instead, the triangulation stage is more appropriately interpreted as a geometrically expressed ambiguity-handling mechanism that reinforces the preferred cluster association for observations with competing memberships.

This distinction is particularly relevant to the adaptive distance framework. Euclidean and fixed-order Minkowski distances define common metric spaces, whereas cluster-specific Mahalanobis distances may involve different covariance matrices for different centroids. Combining such distances as the sides of a single triangle does not necessarily produce a geometrically valid construction. Therefore, angular calculations must be based on a consistently defined geometric space, and any membership adjustment must be distinguished from the metric-based association already established by FASM-FCM.

The research gap addressed by GFASM-FCM lies in integrating adaptive metric selection, selective ambiguity handling, and granular refinement within a sequential fuzzy-clustering framework. The FASM stage adapts the distance measure to the evolving characteristics of the data, while the triangulation stage identifies observations requiring further consideration because of competing cluster associations. Granular refinement subsequently distinguishes reliable cluster members from uncertain boundary observations and regulates their influence during prototype updating. The contribution of the triangulation stage should therefore be assessed through its effect on ambiguity handling and clustering stability, rather than through a claim of independent geometric verification. This integrated formulation provides a basis for investigating whether the additional processing stages improve clustering performance beyond metric adaptation alone.

III. PROPOSED GFASM-FCM METHOD

A. Baseline FCM Formulation

Let $X = \{x_i\}_{i=1}^n$ denote a dataset with $x_i \in \mathbb{R}^d$, let c be the number of clusters, $U = [u_{ij}]$ the fuzzy partition matrix, $V = \{v_j\}_{j=1}^c$ the cluster centroids, and $m > 1$ the fuzzifier. Classical FCM minimizes the membership-weighted objective

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m d^2(x_i, v_j) \quad (1)$$

subject to $0 \leq u_{ij} \leq 1$ and $\sum_{j=1}^c u_{ij} = 1$ for every observation. For a given distance $d_{ij} = d(x_i, v_j)$, the standard membership and centroid updates are

$$u_{ij} = [\sum_{k=1}^c (d_{ij}/d_{ik})^{2/(m-1)}]^{-1} \quad (2)$$

$$v_j = (\sum_{i=1}^n u_{ij}^m x_i) / (\sum_{i=1}^n u_{ij}^m) \quad (3)$$

Traditional FCM normally evaluates these updates with Euclidean distance throughout the run. GFASM-FCM leaves the fuzzy objective itself intact but replaces the fixed-metric assumption with an adaptive selection stage, followed by geometric verification and granular refinement.

B. Local Descriptors and Fuzzy Adaptive Switching Metric

At iteration t , the current partition is summarized using three descriptors: local variance, density, and dimensional variation. Weighted within-cluster variance reflects the spread of observations around a centroid. Density describes how concentrated the observations are in the region represented by a cluster, while dimensional variation captures unequal behaviour across features or changes in their relative contribution. Before entering the fuzzy inference system, the descriptors are normalized to comparable ranges.

Each descriptor is then mapped to interpretable linguistic categories. Variance and density use Low, Medium, and High labels; dimensional variation uses Insignificant, Moderate, and Significant. The rule base is intentionally compact. High variance combined with low density favours Mahalanobis distance because covariance-aware scaling is useful when a region is dispersed or elongated. High density with a roughly uniform shape favours Euclidean distance because its simpler geometry is sufficient. Significant dimensional variation favours Minkowski distance because the order parameter provides a more flexible norm. These rules should be read as context-dependent preferences rather than universal claims about the superiority of one metric. Fig. 1 shows the Fuzzy inference architecture used for the adaptive metric-switching stage.

The three candidate distances are;

$$d_E(x_i, v_j) = [\sum_{r=1}^d (x_{ir} - v_{jr})^2]^{1/2} \quad (4)$$

$$d_M(x_i, v_j) = [(x_i - v_j)^T S_j^{-1} (x_i - v_j)]^{1/2} \quad (5)$$

$$d_{Min}(x_i, v_j) = [\sum_{r=1}^d |x_{ir} - v_{jr}|^p]^{1/p} \quad (6)$$

where, S_j is the local covariance estimate and p is the Minkowski order. When S_j is poorly conditioned, a small diagonal regularizer can be added before inversion. This is particularly important for data with many features relative to the number of available observations.

The fuzzy inference output is represented by a metric-suitability vector $s = [s_E, s_M, s_{Min}]$, with non-negative values normalized to sum to one. In the final implementation, these scores guide selection rather than permanent distance fusion. The active metric is;

$$q^* = \arg \max_q s_q, \quad q \in \{E, M, Min\};$$
$$d_{ij} = d_{q^*}(x_i, v_j) \quad (7)$$

Eq. (7) makes the switching mechanism explicit. A single distance measure is active in each clustering iteration, even though the fuzzy scores retain a graded explanation for why it was selected. When the partition changes, the local descriptors and fuzzy scores are recomputed, so another metric can become active later in the run. In other words, the fuzzy layer provides soft evidence while the clustering layer executes a hard switch.

This design offers two practical advantages. It avoids committing the entire run to the geometry chosen at initialization, and it remains interpretable because each switch can be traced to the observed descriptors and the corresponding rules. It also avoids the additional complexity of learning a large parameterized metric model when the available data may not justify one.

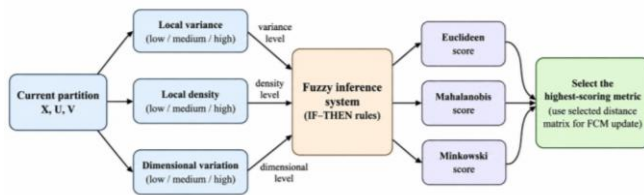


Fig. 1. Fuzzy inference architecture used for the adaptive metric-switching stage.

FASM stage extends traditional FCM by allowing the distance measure to respond to changes in the structure of the evolving partition. Rather than selecting one distance metric for the entire clustering process, FASM evaluates three characteristics of the current cluster regions: local variance, density, and dimensional variation. These descriptors provide the input to a FIS, which assesses the suitability of Euclidean, Mahalanobis, and Minkowski distances.

For a dataset containing n observations, d features, and c clusters, the descriptors are evaluated using the current membership matrix and cluster prototypes. Variance characterizes the dispersion of observations around their representative centroids, density reflects the concentration of observations within cluster regions, and dimensional variation describes differences in the distribution of feature values. Each descriptor is normalized before fuzzy inference so that the decision is not dominated by differences in numerical scale.

The FIS represents variance and density using three linguistic categories: Low, Medium, and High. Dimensional variation is described using Insignificant, Moderate, and Significant. These linguistic values are evaluated through predefined membership functions, enabling the algorithm to express gradual transitions between different data characteristics rather than relying on rigid classification thresholds.

Three principal decision rules are established. First, high variance combined with low density favours Mahalanobis distance because covariance-aware measurement can accommodate differences in dispersion and correlation. Second, high density associated with a relatively uniform cluster structure favours Euclidean distance, which provides a computationally straightforward measure for compact regions. Third, significant dimensional variation favours consideration of Minkowski distance, whose order parameter controls the contribution of coordinate-wise differences.

The outputs of the fuzzy inference process express the relative suitability of the three candidate distance measures. In the hard-switching formulation, the metric with the highest suitability score is selected for the current iteration and used in the subsequent membership update. The partition descriptors are then recalculated, allowing the active metric to change as the cluster structure evolves. Accordingly, the adaptation occurs across clustering iterations rather than through simultaneous application of multiple distance measures within a single update.

Although this approach retains the general membership representation of FCM, changing the active distance measure also changes the objective being evaluated. Therefore, the conventional monotonic-descent property of fixed-metric FCM cannot be assumed for the complete FASM procedure. Computational convergence must instead be evaluated using the observed changes in cluster prototypes, objective-function values, and metric-selection behaviour.

The iterative process is terminated when the maximum change in cluster centroids falls below the prescribed tolerance or the maximum number of iterations is reached. The experiments use a tolerance of 10^{-5} and a maximum of 100 iterations. These stopping conditions establish an operational convergence criterion rather than a theoretical guarantee of monotonic objective reduction.

The switching framework therefore provides an interpretable mechanism for adapting the distance model during fuzzy clustering. Its effectiveness is assessed through clustering-quality measures, empirical stability, convergence behaviour, and comparisons with traditional FCM under controlled initialization and parameter settings.

C. Triangulation-Based Verification of Competing Memberships

After the FASM membership update, GFASM-FCM identifies observations whose two strongest memberships are close enough to be considered genuinely competitive. Let a and b denote the two leading candidate clusters for observation x_i . The normalized redundancy indicator is

$$R_i = |u_{ia} - u_{ib}| / \max(u_{ia}, u_{ib}) \quad (8)$$

A small value of R_i indicates that neither membership clearly dominates. In the experiments, observations with $R_i < 0.20$ are passed to the geometric verification stage. The point x_i and the two centroids v_a and v_b define a triangle. Because the required side lengths are already available from the distance calculations, the interior angles can be obtained using the cosine rule.

Let φ_i be the angle at x_i and let φ_a and φ_b be the angles at the two candidate centroids. The comparison used in this study is

$$S_a = \varphi_i + \varphi_a, \quad S_b = \varphi_i + \varphi_b \quad (9)$$

If $S_a > S_b$, the geometry provides stronger support for cluster a; if $S_b > S_a$, cluster b is preferred. Because φ_i appears in both sums, the decision is equivalent to comparing the centroid angles. Within the triangle, that comparison provides a geometric confirmation of the shorter point-to-candidate-centroid relationship rather than relying only on membership magnitude.

Following the adaptive metric-switching stage, GFASM-FCM examines observations whose memberships indicate competing associations with neighbouring clusters. Traditional FCM permits partial membership in multiple clusters, which is useful for representing overlapping distributions. However, observations with similar membership values may also influence multiple prototypes and contribute to uncertainty in cluster-boundary estimation. The triangulation stage was introduced to identify such observations and provide a selective mechanism for controlling their influence before granular refinement.

For each observation x_i , let (a) and (b) denote the clusters associated with its two largest membership values, where $\mu_{ia} \geq \mu_{ib}$. The degree of competition is evaluated using the normalized redundancy indicator:

$$R_i = \frac{|\mu_{ia} \geq \mu_{ib}|}{\max(\mu_{ia} \geq \mu_{ib})} \quad (10)$$

Observations satisfying ($R_i < 20$) are regarded as ambiguous candidates and passed to the triangulation stage. This threshold is maintained consistently with the experimental settings of the proposed framework. Observations outside this condition retain their existing memberships.

The geometric construction consists of the observation and its two leading candidate centroids, forming a triangle with vertices x_i , v_a , and v_b . In the original formulation, the sum of the angle at the observation and the angle at each candidate centroid was compared to determine the preferred association. Nevertheless, for a geometrically valid triangle, the ordering of the centroid angles is determined by the corresponding opposite side lengths. Hence, when the point-to-centroid distances are calculated within a common metric space, the angular comparison reproduces the nearest-centroid ordering already represented by the fuzzy membership values. Fig. 2 illustrates this point-centroid-centroid construction for an observation with competing cluster memberships; the illustration is used to explain the ambiguity-handling step, not to claim independent evidence beyond the distance ordering.

Accordingly, the triangulation stage is interpreted as a selective ambiguity-handling procedure rather than an independent geometric verification mechanism. Its purpose is to identify observations with competing memberships and apply a controlled adjustment that can reduce the influence of weaker associations. The interpretation is consistent with the original objective of managing overlapping and potentially redundant

memberships while preserving the fuzzy structure of the resulting partition.

An attenuation-based implementation can be expressed by reducing the competing membership through a coefficient α , subject to ($0 < \alpha \leq 1$), followed by normalization so that the sum of memberships for each observation remains equal to one. The attenuation coefficient determines the strength of the adjustment: values closer to one preserve most of the original ambiguity, whereas smaller values produce a more concentrated membership distribution. The precise coefficient and membership-selection rule must correspond to the implementation used in the experiments.

Care is also required when applying the geometric construction to different distance measures. Euclidean distances provide a consistent basis for triangle-angle calculations. However, cluster-specific Mahalanobis distances may use different covariance matrices and cannot automatically be combined into a geometrically valid triangle. Consequently, the interpretation of triangle angles requires a common geometric space and should not be inferred directly from incompatible covariance-weighted distances.

The effect of this stage must be evaluated independently from the metric-switching and granular-refinement components. In particular, improvements in Partition Coefficient and reductions in Classification Entropy may result directly from attenuating and renormalizing fuzzy memberships. Therefore, these measures alone are insufficient to establish improved cluster validity. The contribution of the ambiguity-handling stage should also be assessed through internal clustering indices and agreement with available reference labels, using matched initializations and comparisons with an equivalent direct-distance membership-adjustment procedure.

Within GFASM-FCM, the resulting partition is subsequently passed to granular refinement, where reliable cluster members and uncertain boundary observations are treated differently during prototype updating. This sequence retains the three-stage structure of the proposed framework while providing a more precise interpretation of the role and limitations of triangulation.

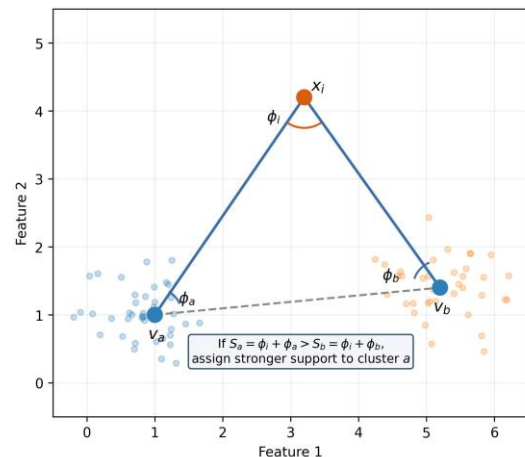


Fig. 2. Triangulation-based verification applied to observations with competing memberships.

D. Granular Refinement

The geometrically verified partition is then organized into information granules. The purpose is to distinguish observations that provide stable evidence for a cluster from those that still belong to an uncertain boundary region. For cluster j , the lower approximation is defined using the strong-membership threshold θ_L :

Observations that do not satisfy the strong-membership condition but still have a plausible association with a cluster remain in the upper or boundary approximation. The lower approximation therefore represents reliable cluster evidence, whereas the boundary explicitly preserves uncertainty. A threshold of $\theta_L = 0.60$ is used throughout the main experiments rather than being tuned separately for each dataset.

Neighbouring granules are compared with a Jaccard-like overlap measure

$$\text{Sim}(G_j, G_k) = |G_j \cap G_k| / |G_j \cup G_k| \quad (11)$$

A similarity value at or above τ indicates that two granules share a meaningful portion of their observations and should be considered during refinement, as shown in Fig. 3. The final centroid update emphasizes the stable members:

$$v_j^* = [\sum_{x_i \in G_j^L} u_{ij}^m x_i] / [\sum_{x_i \in G_j^L} u_{ij}^m] \quad (12)$$

Using only the lower approximation in the refined centroid calculation reduces the direct influence of weak or noisy memberships on the representative centre. The boundary observations are not removed. Their fuzzy memberships remain available for interpretation, so uncertainty is preserved even though those observations carry less weight in the final prototype update.

This separation between representation and prototype influence is central to the granular stage. It explains why the method can improve membership-oriented measures such as Partition Coefficient and Classification Entropy without eliminating FCM's ability to represent overlap.

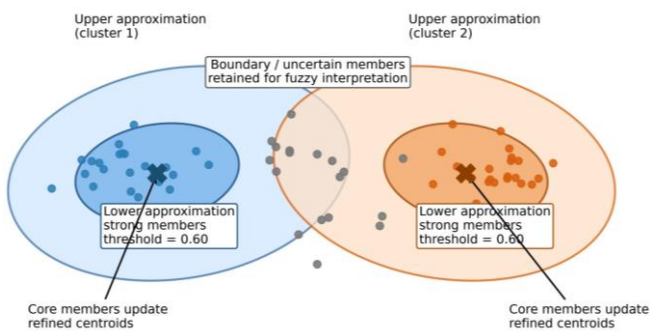


Fig. 3. Granular refinement through lower approximations, boundary regions, and overlap-aware updates.

Convergence is evaluated after the refined centroid update. The procedure terminates when the largest centroid change between consecutive iterations is below ε , or when the maximum number of iterations is reached:

$$\max_j \|v_j^{(t+1)} - v_j^{(t)}\|_2 < \varepsilon \quad (13)$$

Following the adaptive metric-switching and ambiguity-handling stages, the resulting fuzzy partition is refined using granular computing principles. The objective of this stage is to distinguish observations with strong cluster associations from those located in uncertain boundary regions. By emphasizing reliable observations during prototype estimation, granular refinement reduces the direct influence of weak memberships while retaining the fuzzy representation of uncertain observations.

For cluster j , the lower approximation is defined as

$$L_j = \{i: \mu_{ij} \geq \theta_L\}, \theta_L = 0.60 \quad (14)$$

The threshold is fixed throughout the principal experiments. Observations satisfying this condition are treated as reliable members of their corresponding clusters. Since fuzzy memberships are normalized such that their sum equals one for each observation, an observation cannot simultaneously satisfy the lower-approximation condition for two different clusters when the threshold exceeds 0.50. Therefore, the resulting lower approximations are mutually disjoint.

The lower approximation provides the basis for prototype refinement. Rather than allowing every observation to contribute to centroid estimation, only observations satisfying the strong-membership condition are included in the refined centroid calculation:

$$v_j^* = \frac{\sum_{i \in L_j} \mu_{ij}^m x_i}{\sum_{i \in L_j} \mu_{ij}^m} \quad (15)$$

Here, m represents the fuzzification parameter, and μ_{ij} is the membership of observation i in cluster j . This formulation assigns greater influence to observations exhibiting strong membership while excluding uncertain observations from the refined prototype calculation. Their memberships are nevertheless retained in the fuzzy partition for subsequent interpretation and evaluation.

The distinction between reliable and uncertain observations is consistent with the principles of rough-fuzzy clustering, in which lower approximations represent stronger cluster evidence and boundary regions describe uncertainty. Accordingly, the proposed refinement should be interpreted as a lower-approximation-weighted prototype update rather than an independently established overlap-aware centroid adjustment.

In the original formulation, granule similarity was described through a Jaccard measure with an overlap threshold of 0.30. However, the lower approximations defined using the 0.60 membership threshold are mutually disjoint. Therefore, their pairwise Jaccard similarity cannot provide a meaningful overlap criterion. Moreover, the specified prototype-update equation does not depend on the Jaccard similarity value. For this reason, overlap-based prototype modification is not claimed as an operational component of the present formulation.

The contribution of granular refinement is therefore evaluated through its effect on prototype estimation and the resulting clustering partition. Because membership concentration and entropy may also be influenced by preceding membership adjustments, improvements in these quantities are

interpreted alongside internal cluster-validity indices and external agreement measures. This distinction provides a more precise assessment of the granular stage within the overall GFASM-FCM framework.

E. Integrated Algorithm

Fig. 4 summarizes the complete GFASM-FCM workflow. The process begins with data input and preprocessing, continues through adaptive metric selection, the FCM update, selective triangulation, and granular refinement, and then checks convergence. If convergence is not achieved, the refined partition is returned to the next iteration. The diagram makes the sequential role of the three proposed mechanisms explicit.

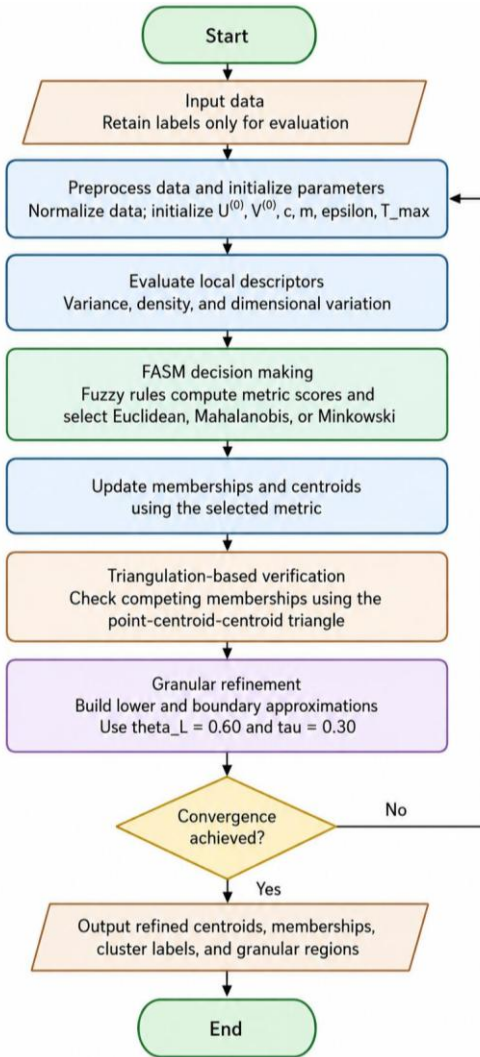


Fig. 4. Overall GFASM-FCM flowchart with data input, decision making, and convergence loop.

Adaptive distance modelling and uncertainty-based prototype refinement have been investigated extensively within the fuzzy-clustering literature. An important early contribution was introduced by Gustafson and Kessel, who extended fuzzy clustering through covariance-based distance modelling. Their approach accommodates clusters with different orientations and elliptical structures, thereby overcoming some of the geometric

restrictions associated with conventional Euclidean FCM (Algorithm 1).

Algorithm 1: GFASM-FCM

Input: X , c , m , ϵ , T_{max} , $\theta_L = 0.60$, and redundancy threshold = 0.20.

- 1) Min-max normalize X ; initialize a normalized fuzzy partition U and compute V .
- 2) Repeat until convergence or T_{max} :
 - a) Compute local variance, density, and dimensional variation.
 - b) Evaluate the fuzzy rules and obtain suitability scores s_E , s_M , and s_{Min} .
 - c) Select $q^* = \arg \max s_q$ and compute distances using the active metric.
 - d) Update U and V using the standard FCM equations.
 - e) Identify observations with competing memberships ($R_i < 0.20$).
 - f) For each candidate, form the point-centroid-centroid triangle, evaluate the angle criterion, attenuate the inconsistent membership, and renormalize.
 - g) Build lower and boundary granules, evaluate granule overlap using τ , and update centroids from the stable region.
- 3) Return refined centroids, fuzzy memberships, cluster labels, granular regions, and convergence information.

Gath and Geva subsequently developed fuzzy-clustering approaches incorporating distributional characteristics, cluster density, and covariance information. Their work further demonstrated the importance of accounting for differences in cluster shape and dispersion rather than assuming identical geometric properties across all clusters.

A related line of research integrates rough-set approximations with fuzzy clustering. Maji and Pal developed a rough-fuzzy possibilistic clustering framework that combines lower approximations, fuzzy boundary representations, and prototype estimation. These contributions established that reliable and uncertain observations can be treated differently when determining cluster representatives.

The present GFASM-FCM framework builds upon these established principles rather than introducing covariance-sensitive distances or granular approximations. Its distinguishing feature is the sequential integration of three procedures within an FCM-based framework. First, fuzzy adaptive metric switching selects an active distance measure according to the characteristics of the evolving partition. Second, an ambiguity-handling procedure examines observations with competing memberships and applies selective membership adjustment. Finally, lower-approximation-based refinement emphasizes observations exhibiting stronger cluster associations during prototype estimation.

Unlike covariance-adaptive methods that incorporate covariance information directly into their clustering formulation, the proposed FASM stage considers several candidate distance measures and selects an active metric during the iterative process. Similarly, the granular stage does not introduce the concept of lower approximations, but incorporates lower-approximation-based prototype refinement after metric adaptation and ambiguity handling.

Accordingly, the methodological contribution of GFASM-FCM lies in its switching policy, integration sequence, and the empirical assessment of these components. Any additional benefit over established covariance-adaptive and rough-fuzzy methods must be determined through controlled comparisons using consistent datasets, preprocessing, initialization, and evaluation criteria.

Covariance-sensitive fuzzy clustering predates the present work. Gustafson and Kessel proposed fuzzy covariance-based clustering; Gath and Geva developed mixture-density-oriented fuzzy clustering. Rough-fuzzy C-means methods, including work by Maji and Pal, already combine lower approximations and boundary influence. The proposed contribution should therefore be limited to the tested sequencing and switching policy, not the first use of covariance-adaptive distances or rough-fuzzy prototypes. Add these as properly numbered references and, if implementation permits, as matched experimental baselines.

F. Computational Considerations

The computational requirements of GFASM-FCM depend on the distance measure selected during each iteration and the additional processing introduced by ambiguity handling and granular refinement. For Euclidean and fixed-order Minkowski distances, the principal distance calculations require $O(ncd)$ operations per iteration, where n represents the number of observations, c denotes the number of clusters, and d is the number of features.

When Mahalanobis distance is selected, additional computations are required to estimate and process covariance matrices. If covariance matrices are estimated separately for each cluster, covariance estimation and matrix factorization can introduce computational costs on the order of $O(ncd^2 + cd^3)$ per iteration. The actual cost depends on the covariance-estimation procedure, numerical regularization, and implementation strategy.

The ambiguity-handling stage operates only on observations satisfying the predefined membership-competition condition. Consequently, it does not require geometric processing for every observation. Once the two leading candidate centroids have been identified, the additional operations associated with each selected observation are limited relative to the full distance-update procedure.

Granular refinement introduces membership thresholding and lower-approximation-based prototype estimation. Identifying observations satisfying the strong-membership condition requires examination of the fuzzy partition, while the refined centroid calculation uses only the selected lower-approximation members. This reduces the number of observations contributing to the final prototype calculation but does not necessarily reduce the total computational cost of the complete iteration.

The additional stages may therefore increase the work performed during an iteration. Any reduction in overall solution time must be interpreted in relation to the observed iteration count, stopping conditions, and implementation environment. Furthermore, because the active distance metric may change between iterations, the standard monotonic-convergence

property of fixed-metric FCM cannot be assumed without additional theoretical justification.

For these reasons, computational performance is evaluated empirically rather than through claims of universal asymptotic superiority. Meaningful runtime comparisons require consistent hardware, software, preprocessing, initialization, and timing procedures. Results obtained from different implementations or previously published experiments should be identified separately and should not be used to establish direct speedup ratios without equivalent experimental conditions.

If the active distance is Euclidean or Minkowski, the main FCM distance update cost is $O(ncd)$ per iteration. Mahalanobis switching introduces covariance estimation and matrix operations. If the covariance is updated independently for every cluster, then the cost would be $O(ncd^2 + cd^3)$. In practice, regularization and reuse of covariance factorizations can reduce this cost. We only use the triangulation stage for the ambiguous subset and not for all n observations. Since the other two candidate centroids are already known from the membership information or distance information, we need only a small number of side and angle calculations in each geometric check. Granular refinement adds membership thresholding and granule overlaps. Its practical cost depends on the number and shape of the granules, but focusing the centroid update on stable members can reduce the number of observations that have an impact on the later updates. GFASM-FCM therefore trades some local reasoning for a potentially better-conditioned clustering process. Runtime is of course implementation dependent. The timing results we show later describe the observed MATLAB implementation in a common environment and are not meant as a general statement that GFASM-FCM has a lower asymptotic cost than all the other methods.

G. Data and Controlled Synthetic Benchmark

The experimental evaluation considers a controlled synthetic dataset together with established biomedical, biological, email, and land-cover benchmarks. The datasets were selected to examine clustering behaviour under differences in feature dimensionality, data dispersion, overlapping observations, and sample size. The synthetic experiment provides a controlled setting for investigating clustering behaviour, whereas the real-world benchmarks provide opportunities to assess performance on naturally occurring heterogeneous data.

The principal synthetic dataset consists of 600 two-dimensional observations distributed equally among three Gaussian clusters. Each cluster contains 200 observations, with mean vectors $\mu_1 = (0, 0)$, $\mu_2 = (2.5, 2.5)$, and $\mu_3 = (5.0, 5.0)$. Unequal covariance structures are used to represent differences in local dispersion and orientation, while the experimental design considers lower-noise and higher-noise conditions.

The covariance matrices and noise-generation parameters must be reported explicitly to ensure that the experiment can be reproduced. The random-number initialization, data-generation sequence, and treatment of noise should also be documented because these factors can influence the degree of overlap and the resulting clustering indices. Both noise conditions should be evaluated using consistent algorithm settings so that changes in performance can be attributed more clearly to the data

characteristics rather than differences in experimental configuration.

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset provides a biomedical benchmark with 569 observations and 30 numerical features describing characteristics of cell nuclei. The dataset contains two reference classes, benign and malignant. Its correlated measurements and differences in feature scale provide a relevant setting for evaluating distance-dependent clustering.

The Abalone dataset contains 4,177 observations with eight original input attributes describing biological and physical characteristics. These include a categorical sex attribute and seven numerical measurements. The number of shell rings is provided as a target variable and is excluded from clustering inputs. Because the attributes have different measurement scales and include categorical information, their preprocessing must be specified before applying distance-based algorithms.

Spambase contains 4,601 observations and 57 features describing word frequencies, character frequencies, and capital-letter patterns in email messages. The reference labels distinguish spam from non-spam messages. The dataset provides a heterogeneous feature space in which individual variables can have substantially different distributions and numerical ranges.

The Urban Land Cover dataset contains 168 observations with 148 features derived from high-resolution aerial imagery. Its features describe spectral, geometric, and texture-related properties, while the reference labels represent nine land-cover categories. The relatively small number of observations compared with its feature dimensionality provides a challenging setting for clustering algorithms using covariance estimation.

The 1m2d dataset is included as a large-scale two-dimensional synthetic benchmark. Its purpose is to assess computational scalability with respect to the number of observations rather than clustering quality in a high-dimensional feature space.

Reference labels are excluded from the clustering inputs and retained exclusively for external evaluation. Numerical attributes are normalized using min-max scaling. Any categorical encoding, missing-value treatment, feature removal, and class-grouping decisions must be applied consistently across the compared algorithms and documented explicitly. The experimental protocol uses 30 runs for randomized methods, with shared initial fuzzy partitions for the FCM-based ablation.

The evaluation distinguishes internal clustering validity, external agreement with reference labels, and computational behaviour. Reporting these categories separately is necessary because favourable runtime does not establish clustering quality, while improved membership concentration does not necessarily indicate better recovery of the underlying data structure.

IV. EXPERIMENTAL DESIGN

A. Data and Controlled Synthetic Benchmark

The evaluation uses a controlled synthetic experiment and a wider group of heterogeneous benchmarks. The main synthetic dataset contains 600 two-dimensional observations and 200

points in each of three partially overlapping Gaussian clusters. The centres of the clusters are (0,0), (2.5,2.5), and (5.0,5.0).

We use different covariance structures to ensure that the groups are not identical spherical clouds and consider Gaussian perturbation at clean (low-noise) and higher-noise (high-noise) conditions. This construction deliberately adds several difficulties to the same problem: unequal local dispersion, correlated directions, and boundary observations with different groups. Spambase contains 4,601 observations and 57 features, while Urban Land Cover contains 168 observations and 148 features. The 1m2d dataset is used as a scalability test since it has a large number of observations, and we do not utilize it as a high-dimensional benchmark. Whenever reference labels are available, they are removed before clustering and only retained for external validation.

B. Experimental Settings

All methods receive the same preprocessed input. Numerical variables are min-max normalized to the interval [0,1]. For methods with random initialization, 30 independent runs will be conducted, and the mean and standard deviation are used for statistical reporting. Within a corresponding run, traditional FCM, FASM-FCM, and GFASM-FCM begin from the same normalized membership matrix. This common starting point reduces the chance that apparent differences are simply caused by favorable initialization, as indicated in Table I.

TABLE I. EXPERIMENTAL SETTINGS USED IN THE MAIN COMPARISONS

Component	Setting
Independent runs	30 (randomized methods)
Normalization	Min-max to [0,1]
Fuzzifier m	2.0
Tolerance ϵ	10^{-5}
Maximum iterations	100
FASM candidates	Euclidean / Mahalanobis / Minkowski
Lower threshold θ_L	0.60
Triangulation candidate	$R_i < 0.20$
Implementation	MATLAB, common environment

The common FCM fuzzifier is $m = 2$, the convergence tolerance is $\epsilon = 10^{-5}$, and the maximum number of iterations is 100. The lower-approximation threshold θ_L is fixed at 0.60, the granule-similarity threshold τ at 0.30, and the triangulation candidate condition at $R_i < 0.20$. MATLAB is used as the common implementation environment.

Execution-time comparisons require the same machine, MATLAB version, preprocessing, numerical precision, stopping conditions, and timing instrumentation. Previously reported cumulative runtimes for FCM-KMeans, FCM++, NFCM, and HOFM cannot establish directly comparable speed-up ratios without verified provenance and matched implementations. Accordingly, the unverified cross-study timing table is omitted, and no numerical cross-study speed-up ranking is claimed. Future matched-environment benchmarking should report per-dataset elapsed times, run-to-run variation, iterations, and the overhead of covariance estimation, metric

switching, ambiguity handling, and granular refinement when available.

C. Baselines and Validation Measures

To assess the effectiveness and stability of the proposed GFASM-FCM framework, the experimental evaluation considers clustering quality, agreement with reference labels, convergence behaviour, and computational performance. The evaluation protocol is designed to distinguish improvements associated with the proposed clustering stages from variation caused by random initialization.

For randomized clustering methods, the evaluation uses 30 independent runs for each dataset. Within each corresponding run, conventional FCM, FASM-FCM, and GFASM-FCM are initialized using the same normalized fuzzy partition matrix. This paired initialization strategy provides a controlled basis for assessing the incremental effects of adaptive metric switching and the subsequent ambiguity-handling and granular-refinement stages.

For each method and dataset, clustering performance is evaluated using complementary internal and external validation measures. Internal measures include the Silhouette coefficient, Davies–Bouldin Index (DBI), Dunn Index, Partition Coefficient (PC), and Classification Entropy (CE). External measures comprise the Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Purity, where reference labels are available. Computational evaluation considers the number of iterations and elapsed execution time.

The results of randomized methods are summarized using the arithmetic mean and sample standard deviation across the 30 runs. Mean values indicate the average observed performance, whereas standard deviations describe sensitivity to random initialization and variation between experimental runs. Deterministic baselines are reported separately when repeated execution under identical conditions produces the same clustering solution. Execution time is summarized across repeated measurements because computational timing may vary even for deterministic algorithms.

Statistical comparisons between FASM-FCM and GFASM-FCM will be conducted using a two-sided paired Wilcoxon signed-rank test. For each evaluation measure, paired differences are calculated from corresponding runs initialized under the same conditions. This nonparametric procedure evaluates whether the distribution of paired differences provides evidence of a systematic change in performance between the two methods. Zero differences and tied absolute differences are handled using a stated statistical convention.

Because multiple clustering indices and datasets are considered, the resulting p-values are adjusted using the Holm multiple-comparison procedure. Statistical significance is assessed at a family-wise significance level of 0.05. Alongside adjusted p-values, the analysis reports an appropriate paired effect size to characterize the magnitude of the observed differences.

The experimental protocol also requires consistent preprocessing and transparent parameter settings for all comparison methods. Gaussian Mixture Models (GMM) are

evaluated using explicitly documented covariance assumptions, covariance regularization, initialization settings, and convergence criteria. DBSCAN is evaluated using a defined neighbourhood radius, minimum neighbourhood size, and parameter-selection procedure that does not use reference labels to optimize the reported external validation results.

Where computationally feasible, established covariance-adaptive fuzzy clustering approaches, including Gustafson–Kessel and Gath–Geva, are included as additional baselines under matched experimental conditions. Any methods or datasets excluded because of computational limitations are identified explicitly, together with the reason for exclusion.

The statistical evaluation provides a basis for determining whether the performance differences between FASM-FCM and GFASM-FCM are consistent across repeated runs. In particular, small differences in external clustering agreement are interpreted in conjunction with their run-level variation, statistical significance, and effect sizes rather than being treated as conclusive evidence of improvement.

The main synthetic comparison includes GFASM-FCM, the intermediate FASM-FCM model, traditional FCM, K-Means, Gaussian Mixture Models (GMM), and DBSCAN. The FCM/FASM/GFASM sequence is especially important because it acts as an ablation study: it separates the benefit of adaptive metric switching from the additional effect of the complete geometric and granular refinement sequence. The large-scale timing comparison also includes FCM-KMeans, FCM++, NFCM, and HOFCM as reported in the source experiments.

Internal validation uses the Silhouette coefficient [25], Dunn Index, Davies–Bouldin Index (DBI) [26], Partition Coefficient (PC), and Classification Entropy (CE). Higher Silhouette, Dunn, and PC values are desirable, while lower DBI and CE values indicate more favourable separation or less ambiguous membership. External validation uses Adjusted Rand Index (ARI) [27], Normalized Mutual Information (NMI), and Purity. Runtime and iteration count are reported to characterize computational behaviour.

D. Evaluation Procedure, Reproducibility, and Fairness

The experimental procedure was designed to minimise differences caused by preprocessing, initialization, or access to class information. All comparison methods receive the same normalized representation. When reference labels are available, they are withheld during clustering and are used only after the clustering process has finished. This is essential because GFASM-FCM is an unsupervised method; using labels to guide initialization or parameter selection would change the nature of the comparison. For the FCM-based ablation, FCM, FASM-FCM, and GFASM-FCM use the same normalized fuzzy membership matrix within each run so that differences can be interpreted component by component. Thirty independent runs will be conducted for randomized methods. Aggregate performance is treated as the principal indicator, while run-to-run variation is used to assess stability.

The main GFASM-FCM parameters are fixed before the principal comparison: $m = 2$, $\varepsilon = 10^{-5}$, $T_{max} = 100$, $\theta_L = 0.60$, $\tau = 0.30$, and $R_i < 0.20$ for triangulation candidates. Keeping these values unchanged across the main experiments avoids tuning the

proposed method separately for every dataset after the outcomes are known. This matters particularly for a multi-stage algorithm, because aggressive dataset-specific tuning can make it difficult to distinguish a real design improvement from a parameter-selection advantage.

The same caution is applied to the timing results. Execution times are obtained in a common MATLAB environment and interpreted together with iteration counts. GFASM-FCM performs operations that plain FCM does not, including fuzzy metric selection, selective triangulation, and granular refinement. A favourable measured time is therefore meaningful only within the reported implementation and hardware setting. The timing comparison is intended to show that the additional stages remain manageable and can converge efficiently, not that GFASM-FCM has a universally lower theoretical complexity than all baselines.

E. Validation Metric Formulation and Interpretation

Several complementary measures are used because no single index captures every aspect of clustering quality. Internal indices examine compactness and separation without reference labels, whereas external indices compare the obtained partition with known classes when such labels exist. Membership-based measures are also retained because GFASM-FCM produces a fuzzy partition, and a meaningful improvement in membership certainty may occur even when global geometry changes only modestly.

For observation i , the Silhouette coefficient compares its average dissimilarity $a(i)$ to observations in the same cluster with the smallest average dissimilarity $b(i)$ to another cluster [25]:

$$s(i) = [b(i) - a(i)] / \max \{a(i), b(i)\} \quad (16)$$

The overall Silhouette score is the mean over all observations. Values closer to one indicate that observations are, on average, better matched to their assigned cluster than to neighbouring clusters. In this study, the measure helps assess whether adaptive metric switching improves the large-scale geometric arrangement of the partition.

The Davies-Bouldin Index provides a complementary view by comparing within-cluster scatter with the separation between cluster representatives [26]. For cluster j , let S_j represent within-cluster dispersion and M_{jk} the distance between centroids j and k :

$$DBI = (1/c) \sum_{j=1}^c \max_{k \neq j} [(S_j + S_k)/M_{jk}] \quad (17)$$

A smaller DBI is preferred because it indicates that clusters are compact relative to the distance between their centres. Reading DBI together with Silhouette reduces the risk of drawing a conclusion from only one geometric perspective.

Two measures are used to inspect the fuzzy partition directly. The Partition Coefficient summarizes the concentration of membership values:

$$PC = (1/n) \sum_{i=1}^n \sum_{j=1}^c u_{ij}^2 \quad (18)$$

Higher PC values indicate a more concentrated membership structure. Classification Entropy (CE) reflects membership dispersion and is defined as

$$CE = -(1/n) \sum_{i=1}^n \sum_{j=1}^c u_{ij} \ln(u_{ij}) \quad (19)$$

Lower CE is interpreted as less ambiguity in the reported experiments. PC and CE are especially relevant to the granular stage because refinement can reorganize uncertain memberships without necessarily producing another large change in Silhouette or DBI. They are therefore interpreted alongside the geometry-based indices.

External validation is reported with ARI, NMI, and Purity. ARI corrects partition agreement for chance [27], NMI measures the information shared by the produced clusters and the reference classes, and Purity measures the dominant reference class within each obtained cluster. Higher values are preferred for all three. Reference labels are used only at this evaluation stage and do not enter metric switching, triangulation, granule construction, centroid updates, or convergence.

V. RESULTS AND DISCUSSION

A. Internal Validation and Ablation

Table II reports the internal validation results for the 600-observation synthetic benchmark. Two effects are evident. First, adaptive metric switching improves the Silhouette coefficient from 0.68 to 0.82 and reduces the Davies-Bouldin Index from 0.30 to 0.23. This suggests that a substantial part of the geometric mismatch in traditional FCM is reduced when the active distance measure can respond to the evolving local structure. Second, the complete GFASM-FCM sequence further improves the membership-oriented measures. Although Silhouette and DBI remain unchanged relative to FASM-FCM, the Partition Coefficient increases from 0.62 to 0.66, Classification Entropy decreases from 0.68 to 0.60, and the Dunn Index increases from 1.45 to 1.48. Relative to traditional FCM, GFASM-FCM improves Silhouette by 20.6%, lowers DBI by 23.3%, improves PC by 22.2%, and reduces CE by 24.1%. These results are consistent with the intended division of roles: metric switching primarily improves global geometry, whereas the later stages refine ambiguous memberships and stable prototype evidence.

TABLE II. INTERNAL VALIDATION ON THE PRINCIPAL SYNTHETIC DATASET

Metric	FCM	FASM-FCM	GFASM-FCM
Silhouette $\uparrow$	0.68	0.82	0.82
Dunn $\uparrow$	1.25	1.45	1.48
DBI $\downarrow$	0.30	0.23	0.23
PC $\uparrow$	0.54	0.62	0.66
CE $\downarrow$	0.79	0.68	0.60

B. External Validation and Convergence

The reported ablation indicates that Silhouette and DBI are unchanged between FASM-FCM and GFASM-FCM, whereas PC changes from 0.62 to 0.66 and CE from 0.68 to 0.60. Such membership shifts are consistent with intentional sharpening and should not, by themselves, be interpreted as evidence of more accurate cluster structure. The observed ARI difference of 0.01 also cannot be assessed without paired variation and statistical testing. The stage-specific contribution remains unverified pending real-dataset, 30-run analyses.

Traditional FCM records $ARI = 0.65$, $NMI = 0.75$, and $Purity = 0.81$. Among the non-proposed baselines, GMM performs best in this synthetic experiment, but its ARI of 0.72 and NMI of 0.80 remain below the two adaptive FCM variants. K-Means is computationally fast but gives weaker agreement with the reference labels, illustrating why runtime alone is not sufficient when the data contain genuine overlap. DBSCAN performs least strongly on this controlled diagonal-overlap arrangement, although density-based methods can be advantageous for highly non-globular structures.

TABLE III. EXTERNAL VALIDATION RESULTS ON THE PRINCIPAL 600-OBSERVATION SYNTHETIC DATASET

Method	ARI	NMI	Purity
GFASM-FCM	0.77	0.88	0.93
FASM-FCM	0.76	0.85	0.91
Traditional FCM	0.65	0.75	0.81
K-Means	0.62	0.73	0.78
GMM	0.72	0.80	0.85
DBSCAN	0.55	0.68	0.70

Table III presents the previously reported external validation results for the principal synthetic dataset. Among the compared methods, GFASM-FCM recorded the highest reported values for Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Purity, with scores of 0.77, 0.88, and 0.93, respectively.

FASM-FCM achieved corresponding values of 0.76, 0.85, and 0.91, indicating relatively small differences between the metric-switching stage and the complete framework. Conventional FCM recorded lower values of 0.65, 0.75, and 0.81.

These observations suggest that adaptive metric switching may account for a substantial portion of the improvement relative to conventional FCM. However, the additional contribution of triangulation and granular refinement cannot be established from the reported point estimates alone.

In particular, the ARI difference between FASM-FCM and GFASM-FCM is 0.01. Without the run-level variability and paired statistical analysis, this difference cannot be interpreted as statistically significant. Therefore, the reported values are treated as descriptive results pending verification of their aggregation across repeated experiments.

The comparison also remains specific to the controlled synthetic dataset. Additional external validation on real-world benchmarks is required before broader conclusions can be drawn about general clustering performance.

Table IV summarizes the previously reported computational measurements on the principal 600-observation synthetic dataset. GFASM-FCM recorded an elapsed time of 3.4 seconds and 19 iterations, compared with 3.5 seconds and 25 iterations for FASM-FCM and 4.2 seconds and 30 iterations for conventional FCM.

K-Means recorded the lowest reported execution time of 2.9 seconds. The observed execution times suggest that the

additional stages in GFASM-FCM did not produce a larger total runtime in the reported experiment. However, this observation should not be interpreted as evidence of lower theoretical computational complexity.

TABLE IV. COMPUTATIONAL RESULTS ON THE PRINCIPAL SYNTHETIC DATASET

Method	Time	Iterations
GFASM-FCM	3.4 s	19
FASM-FCM	3.5 s	25
Traditional FCM	4.2 s	30
K-Means	2.9 s	15
GMM	4.0 s	28
DBSCAN	3.8 s	20

The proposed framework includes additional operations associated with adaptive metric selection, ambiguity handling, and granular prototype refinement. Consequently, any observed reduction in elapsed time may depend on the number of iterations, implementation efficiency, and computational environment.

The reported measurements require confirmation under a consistent execution environment, including identical preprocessing, documented stopping conditions, hardware specifications, and timing procedures. Iteration counts should also be interpreted according to the convergence mechanisms of the individual algorithms. Accordingly, the computational findings are treated as preliminary implementation-specific observations rather than evidence of general computational superiority.

GFASM-FCM converges in 19 iterations and 3.4 s in the reported environment, compared with 25 iterations for FASM-FCM and 30 iterations for traditional FCM. K-Means remains faster at 2.9 s, as expected from its simpler hard-assignment update. The key observation is therefore not that GFASM-FCM is universally the fastest method, but that its adaptive, geometric, and granular stages remain computationally manageable while preserving the strongest fuzzy and external validation results in this experiment.

C. Contribution of Triangulation and Granular Refinement

The ablation pattern also clarifies the specific roles of triangulation and granular refinement. FASM-FCM already reaches the final Silhouette and DBI values, indicating that the later stages are not primarily responsible for global cluster separation. Triangulation acts locally: it revisits observations whose two strongest memberships remain competitive and checks whether those memberships are geometrically consistent with the observation and its two candidate centroids. Granular refinement is more visible in PC and CE. From FASM-FCM to GFASM-FCM, PC increases from 0.62 to 0.66 (6.5%), while CE decreases from 0.68 to 0.60 (11.8%). The resulting partition therefore gives stronger influence to stable members without forcing every uncertain observation into an artificially crisp assignment.

D. Large Scale Benchmark Timing

The computational efficiency of fuzzy-clustering algorithms depends on dataset characteristics, algorithmic operations, convergence behaviour, and implementation conditions. Although GFASM-FCM incorporates additional processing stages beyond conventional FCM, these stages are intended to operate selectively within the iterative clustering procedure.

In the proposed framework, metric switching determines the active distance measure, ambiguity handling is restricted to observations with competing memberships, and granular refinement updates prototypes using reliable lower-approximation members. These design choices may affect the computational work performed during individual iterations and the total number of iterations required for termination.

However, comparisons involving previously published execution times cannot establish a fair performance ranking without equivalent hardware, software, datasets, preprocessing, and stopping conditions. The earlier accumulated timing comparison is therefore excluded from the revised evaluation because the computational environments and timing procedures of all comparator methods could not be established as equivalent.

The computational interpretation is consequently limited to measurements whose experimental conditions can be verified. Further matched-environment benchmarking is necessary to determine whether GFASM-FCM provides a reproducible runtime advantage over established fuzzy-clustering variants on heterogeneous and large-scale datasets.

E. Relation to Existing Fuzzy-Clustering Research

The empirical findings agree with two broader directions in the fuzzy-clustering literature. First, they reinforce the importance of distance geometry when clusters differ in scale, covariance, or orientation [7], [8], [24]. GFASM-FCM does not replace Euclidean distance with one globally fixed alternative. Instead, it allows the active metric to change as the partition evolves, keeping the simplicity of familiar distances while adding a local adaptation mechanism.

Second, the improvement in the membership-oriented measures is consistent with the role of granulation and fuzzy-rough boundary modelling reported in recent work [11]-[17]. The main difference lies in the order of operations. In GFASM-FCM, granular refinement is applied only after the membership values have been updated with an adaptive metric and ambiguous cases have been checked geometrically. The granules therefore operate on a partition that has already passed through two earlier correction stages.

The method also remains distinct from kernel and spatial FCM variants. Kernel models can represent nonlinear structure more flexibly [5], [18], [20], while spatial FCM is particularly effective when neighbourhood information has a direct domain meaning [22], [23], [31]. GFASM-FCM is most appropriate when the main difficulties are heterogeneous distance geometry, overlapping memberships, and uncertain boundaries in vector data. It should therefore be viewed as a complementary framework rather than a replacement for every specialized clustering model.

F. Robustness Across Different Data Characteristics

The benchmark collection was chosen to expose the algorithm to more than one form of difficulty. The controlled synthetic data make covariance differences and overlap visible, while the additional biomedical, image, heterogeneous, and large-scale benchmarks test whether the same processing sequence remains practical when structure is less predictable. Spambase offers 4,601 observations with 57 features, whereas Urban Land Cover represents the opposite balance with only 168 observations but 148 features. The 1m2d benchmark addresses a different issue altogether: it is a very large two-dimensional dataset used to examine scalability with respect to the number of observations.

These contrasts matter because the three GFASM-FCM components address different weaknesses. FASM is most valuable when one distance model does not describe every region equally well. Triangulation becomes relevant when two candidate memberships remain close enough that the distance-based update does not provide a convincing distinction. Granular refinement answers the later question of which observations should contribute most strongly to the final prototype after those earlier corrections have been made.

This manuscript presents a three-stage fuzzy-clustering design combining metric switching, an ambiguity-focused membership update, and lower-approximation prototype refinement. The existing synthetic experiment provides preliminary ablation results, but the incremental gain of the latter two stages is not established without controlled sharpening baselines and paired statistical testing. The angle ordering used in the original triangulation rule is equivalent to a distance ordering under a common metric, and lower approximations defined by membership ≥ 0.60 are disjoint. Conclusions about independent geometric verification, effective granule-overlap control, broad real-data superiority, and universal timing improvements are therefore withheld until implementation details and reproducible experiments are supplied.

The timing experiment provides a second robustness perspective. GFASM-FCM carries out more operations than plain FCM, yet its accumulated solution time remains favourable in the reported implementation. Selective processing appears to be important here: triangulation is restricted to ambiguous observations, and granular refinement emphasizes stable regions instead of treating all points as equally informative at every step. The resulting framework adds interpretable reasoning without making the iterative process impractical for the tested benchmarks.

G. Why the Sequential Order of GFASM-FCM Matters

The sequence FASM $\rightarrow$ triangulation $\rightarrow$ granular refinement is intentional. Metric adaptation is applied first because the interpretation of fuzzy membership depends on the distances used to construct the partition. If an unsuitable metric has already distorted those memberships, applying a boundary treatment immediately would risk refining uncertainty that was partly created by the wrong geometry. Triangulation follows the membership update because it requires two plausible candidate clusters for an observation and is therefore applied only where ambiguity remains.

Clear memberships are left unchanged. Granular refinement is placed last so that lower and boundary approximations are constructed from the most reliable membership information available in the iteration. In this order, GFASM-FCM addresses three related questions: how similarity should be measured, whether a competing membership is geometrically plausible, and which memberships should have the greatest influence on final prototype refinement.

VI. PRACTICAL IMPLICATIONS AND LIMITATIONS

GFASM-FCM is designed to remain interpretable. A metric switch can be related to observable descriptors such as variance, density, and dimensional variation; triangulation provides a geometric reason for revisiting only a small subset of ambiguous observations; and granular refinement explicitly distinguishes stable members from uncertain boundary observations. Several limitations should nevertheless be recognised. First, the method remains centroid-based, so highly irregular or strongly non-globular structures may still favour density-based alternatives. Second, GFASM-FCM remains sensitive to initialization; the use of 30 independent runs reduces dependence on a single starting point, but deterministic or metaheuristic initialization deserves further study. Third, the fuzzy descriptors and thresholds are intentionally interpretable rather than learned end-to-end, which may limit their flexibility in some datasets. Fourth, Mahalanobis switching depends on stable covariance estimation, making regularization important in high-dimensional, low-sample-size settings. Fifth, the large-scale runtime evidence is implementation-specific and should be replicated across software platforms and hardware. Finally, reproducibility would be strengthened by releasing source code, fixed random seeds, raw results for all runs, and sensitivity analyses for m , θ_L , τ , the Minkowski order, and the triangulation threshold.

VII. CONCLUSION

This study presented GFASM-FCM, an adaptive fuzzy-clustering framework that combines metric switching, triangulation-based verification, and granular refinement in a sequential process. The framework allows the distance model to respond to local structure, uses geometry as a second source of evidence when memberships remain competitive, and then separates stable evidence from uncertain boundary memberships during prototype refinement. On the principal synthetic benchmark, GFASM-FCM achieved a Silhouette coefficient of 0.82, DBI of 0.23, PC of 0.66, CE of 0.60, ARI of 0.77, NMI of 0.88, and Purity of 0.93. The ablation results indicate that adaptive metric switching contributes most strongly to the improvement in global cluster geometry, whereas the later stages primarily improve membership reliability. The reported heterogeneous benchmark timing results further indicate that the additional reasoning stages remain computationally manageable. Future work should examine more robust covariance estimation, automatic threshold selection, streaming and distributed implementations, and integration with learned feature representations.

REFERENCES

- [1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981.
- [2] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, nos. 2–3, pp. 191–203, 1984, doi: 10.1016/0098-3004(84)90020-7.
- [3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [4] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.
- [5] D. Graves and W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study," *Fuzzy Sets and Systems*, vol. 161, no. 4, pp. 522–543, 2010, doi: 10.1016/j.fss.2009.10.021.
- [6] S. Askari, "Noise-resistant fuzzy clustering algorithm," *Granular Computing*, vol. 6, no. 4, pp. 815–828, 2021.
- [7] B. A. Pimentel, R. de Amorim Silva, and J. C. S. Costa, "Fuzzy C-means clustering algorithms with weighted membership and distance," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 30, no. 4, pp. 567–594, 2022, doi: 10.1142/S0218488522500143.
- [8] A. S. Hadi, "A new distance between multivariate clusters of varying locations, elliptical shapes, and directions," *Pattern Recognition*, vol. 129, Art. no. 108780, 2022, doi: 10.1016/j.patcog.2022.108780.
- [9] V. Mehta, S. Bawa, and J. Singh, "Analytical review of clustering techniques and proximity measures," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5995–6023, 2020.
- [10] I. C. Gormley, T. B. Murphy, and A. E. Raftery, "Model-based clustering," *Annual Review of Statistics and Its Application*, vol. 10, pp. 573–595, 2023, doi: 10.1146/annurev-statistics-033121-115326.
- [11] X. Hu, Y. Tang, W. Pedrycz, K. Di, J. Jiang, and Y. Jiang, "Fuzzy clustering with knowledge extraction and granulation," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 4, pp. 1098–1112, 2023.
- [12] W. Li, S. Zhai, W. Xu, W. Pedrycz, Y. Qian, W. Ding, and T. Zhan, "Feature selection approach based on improved fuzzy C-means with principle of refined justifiable granularity," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 7, pp. 2112–2126, 2023.
- [13] W. Pedrycz, "Computing and clustering in the environment of order-2 information granules," *IEEE Transactions on Cybernetics*, vol. 53, no. 9, pp. 5414–5423, 2023.
- [14] Z. Yang, S. Jiang, F. Yu, W. Pedrycz, H. Yang, and Y. Hao, "Linear fuzzy information-granule-based fuzzy C-means algorithm for clustering time series," *IEEE Transactions on Cybernetics*, vol. 53, no. 12, pp. 7622–7634, 2023, doi: 10.1109/TCYB.2022.3184999.
- [15] T. Zhang, Y. Zhang, F. Ma, C. Peng, D. Yue, and W. Pedrycz, "Local boundary fuzzified rough K-means-based information granulation algorithm under the principle of justifiable granularity," *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 519–532, 2024.
- [16] B. Yu, Z. Zheng, M. Cai, W. Pedrycz, and W. Ding, "FRCM: A fuzzy rough C-means clustering method," *Fuzzy Sets and Systems*, vol. 480, Art. no. 108860, 2024, doi: 10.1016/j.fss.2024.108860.
- [17] R. Wang, K. Xu, and Y. Wang, "Augmentation of soft partition with a granular prototype-based fuzzy C-means," *Mathematics*, vol. 12, no. 11, Art. no. 1639, 2024, doi: 10.3390/math12111639.
- [18] Y. Tang, Z. Pan, X. Hu, W. Pedrycz, and R. Chen, "Knowledge-induced multiple kernel fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 14838–14855, 2023.
- [19] C. Wang, M. Zhou, W. Pedrycz, and Z. Li, "Comparative study on noise-estimation-based fuzzy C-means clustering for image segmentation," *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 241–253, 2024.
- [20] A. Farooq and K. H. Memon, "Kernel possibilistic fuzzy C-means clustering algorithm based on morphological reconstruction and membership filtering," *Fuzzy Sets and Systems*, vol. 477, Art. no. 108792, 2024.
- [21] J. Chen, J. Zhu, H. Jiang, H. Yang, and F. Nie, "Sparsity fuzzy C-means clustering with principal component analysis embedding," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 7, pp. 2099–2111, 2023.
- [22] S. Mahapatra, S. Agrawal, P. K. Mishro, and R. B. Pachori, "A novel framework for retinal vessel segmentation using optimal improved Frangi filter and adaptive weighted spatial FCM," *Computers in Biology and Medicine*, vol. 147, Art. no. 105770, 2022.

- [23] J. Xu, T. Zhao, G. Feng, M. Ni, and S. Ou, "A fuzzy C-means clustering algorithm based on spatial context model for image segmentation," *International Journal of Fuzzy Systems*, vol. 23, no. 3, pp. 816–832, 2021.
- [24] X. Zhao, Y. Li, and Q. Zhao, "Mahalanobis distance based on fuzzy clustering algorithm for image segmentation," *Digital Signal Processing*, vol. 43, pp. 8–16, 2015.
- [25] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [26] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [27] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, pp. 193–218, 1985.
- [28] S. Alpert, M. Galun, A. Brandt, and R. Basri, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 315–327, 2012.
- [29] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.
- [30] I. Shafi, M. Chaudhry, E. C. Montero, E. S. Alvarado, I. de la Torre Diez, M. A. Samad, and I. Ashraf, "A review of approaches for rapid data clustering: Challenges, opportunities, and future directions," *IEEE Access*, vol. 12, pp. 138086–138120, 2024.
- [31] Q. Wang, X. Wang, C. Fang, et al., "Robust fuzzy C-means clustering algorithm with adaptive spatial and intensity constraint and membership linking for noise image segmentation," *Applied Soft Computing*, vol. 92, Art. no. 106318, 2020.
- [32] Y. Zeng, Z. Xu, Y. He, and Y. Rao, "Fuzzy entropy clustering by searching local border points for the analysis of gene expression data," *Knowledge-Based Systems*, vol. 190, Art. no. 105309, 2020.
- [33] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.
- [34] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *Proc. IEEE Conf. Decision and Control*, 1978/1979, doi:10.1109/CDC.1978.268028.
- [35] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(7), 773–780, 1989, doi:10.1109/34.192473.
- [36] P. Maji and S. K. Pal, "RFCM: A Hybrid Clustering Algorithm Using Rough and Fuzzy Sets," *Fundamenta Informaticae*, 80(4), 475–496, 2007, doi:10.3233/FUN-2007-80408.