Journal of Neutrosophic and Fuzzy Systems

♦ ISSN: 2771-6449 Vol: 10, No. 02, 2025 ♦ PP. 10 – 15

♦ Article DOI: https://doi.org/10.54216/JNFS.100202

FULL LENGTH ARTICLE

Entropy–Remoteness Neutrosophic Fuzzy c-Means for Separating Boundary Ambiguity from Outlierness

Das Suman Das[1,*] **Ajoy Kanti**[2]

[1] Assistant Professor Grade II (Mathematics), Department of Education, National Institute of Technology Calicut, Kozhikode–673601, Kerala, India

[2] Associate Professor, Department of Mathematics, Tripura University, Agartala–799022, Tripura, India

Emails: dr.sumandas1995@gmail.com · ajoykantidas@gmail.com

Received: November 12, 2024 Revised: December 16, 2024 Accepted: January 17, 2025 ⋆ Corresponding author

ABSTRACT

A weak fuzzy assignment can mean two geometrically different things. A record may lie between otherwise legitimate clusters, so its membership vector $u_i = (u_{i1}, \ldots, u_{iK})$ is genuinely ambiguous; or it may be remote from every prototype, in which case diffuse membership is a symptom of outlierness rather than a boundary. Treating both cases through one fuzziness scalar makes the prototype update unable to explain *why* a record should have reduced influence. This paper constructs Entropy–Remoteness Neutrosophic Fuzzy c-Means (ERN-FCM) from two separate quantities. Boundary indeterminacy is $I_i = \alpha[-\sum_k u_{ik} \log u_{ik}]/\log K$, whereas robust remoteness is $F_i = 1 - \exp(-\tau\rho_i)$ with $\rho_i = [(\delta_i - q_{.90})/(q_{.90} - q_{.50} + \varepsilon)]_+$ and $\delta_i = \min_k \|x_i - v_k\|^2$. Their conjunction gives truth $T_i = (1 - I_i)(1 - F_i)$ and therefore $T_i + I_i + F_i = 1 + I_i F_i$. Prototypes are updated by $v_k^+ = \sum_i T_i u_{ik}^m x_i / \sum_i T_i u_{ik}^m$, so remote or highly ambiguous records contribute less without being hard-deleted. The numerical study deliberately separates the two geometries. On Wine with 10% gross contamination, mean adjusted Rand index rises from 0.8665 for ordinary FCM to 0.8975 for ERN-FCM; on Breast Cancer the corresponding values are 0.7014 and 0.7074, whereas Iris remains a counterexample where FCM is slightly better (0.6328 versus 0.6179). The remoteness-aware uncertainty $1 - T_i$ identifies gross outliers with AUC 0.9898–0.9956 at 10% contamination, and on a separate boundary benchmark I_i identifies bridge observations with AUC 0.9963. Paired bootstrap contrasts, component ablation, parameter sensitivity, and fixed-point diagnostics support a focused conclusion: ERN-FCM is most useful when cluster recovery and the *type* of uncertain observation must be distinguished, not as a universal replacement for FCM.

Keywords: Fuzzy c-means ▪ Single-valued neutrosophic information ▪ Clustering ▪ Entropy ▪ Outlier detection ▪ Boundary ambiguity ▪ Robust prototypes

1. INTRODUCTION

Consider a fuzzy partition $U = (u_{ik}) \in [0,1]^{n \times K}$ with $\sum_{k=1}^{K} u_{ik} = 1$. If $u_i \approx (1/2, 1/2, 0, \ldots)$, the observation may be a legitimate boundary record. If instead all prototype distances are large and similar, the same diffuse pattern can arise because the record is remote from every cluster. In both cases the ordinary fuzzy c-means update assigns a nonzero contribution u_{ik}^m to prototype k. The numerical appearance of uncertainty is similar, but the geometry is not. The central question of this paper is therefore not simply how to make clustering "more robust," but how to distinguish *ambiguity* from *outlierness* before they enter the prototype update.

Fuzzy sets formalize graded membership $\mu \in [0,1]$ rather

than binary inclusion [1]. Fuzzy c-means (FCM) operationalizes that idea by minimizing a weighted within-cluster sum of squares and retaining a full membership vector for each observation [2]. Noise clustering and possibilistic clustering subsequently addressed the fact that soft partitioning alone does not make a prototype estimator immune to atypical observations [3, 4]. These developments motivate robust weighting, but they do not by themselves provide two semantically distinct coordinates for boundary ambiguity and geometric remoteness.

Single-valued neutrosophic information permits truth, indeterminacy, and falsity degrees to coexist without requiring a unit-sum constraint [5]. Neutrosophic c-means has already shown that this representation can be integrated into clustering objectives [6]; similarity and entropy measures further clarify how uncertainty can be quantified inside the neutrosophic state space [7]. The present construction takes a narrower route. It does not introduce an additional noise cluster or kernel. Instead, it maps two observable features of the current fuzzy partition into a record-level state (T_i, I_i, F_i) and uses only the truth component to control prototype influence.

The resulting algorithm, ERN-FCM, is built from three equations. First, normalized partition entropy supplies I_i. Second, a quantile-normalized tail coordinate supplies F_i. Third, the conjunction

$$T_i = (1 - I_i)(1 - F_i)$$

attenuates the usual weight u_{ik}^m. The decomposition is deliberately asymmetric: I_i is high when membership is distributed across clusters, whereas F_i is high only when the nearest-cluster squared distance δ_i lies beyond a robust upper reference quantile. Hence a bridge point may have $I_i \gg F_i$, while a remote point may have $F_i \gg I_i$; the combined unresolved quantity $1 - T_i = I_i + F_i - I_i F_i$ reacts to either mechanism.

The paper is organized around this two-geometry distinction rather than around a conventional method-first narrative. Section 2 compresses the relevant clustering literature into the specific design requirements used here. Section 3 derives the record-level neutrosophic state from a fuzzy partition. Section 4 gives the fixed-point algorithm, and Section 5 establishes boundedness, scale behavior, conditional optimality, and influence attenuation. Section 6 then designs two different contamination experiments: gross radial outliers on three real benchmark datasets, and bridge ambiguity on a controlled three-cluster geometry. Section 7 presents seven numerical tables, after which Section 8 discusses where the method helps, where it does not, and why those failures are informative. All nine references predate 2021, so the manuscript is suitable for a 2025 volume without relying on later material.

2. RELATED WORK

2.1 Soft Partitioning and Atypical Observations

FCM assigns every $x_i \in \mathbb{R}^d$ to every cluster with membership u_{ik} and fuzzifier $m > 1$ [2]. This is appropriate for overlapping clusters, but membership softness is not identical to robust rejection. Davé's noise-cluster formulation explicitly allocates atypical points to a separate noise class [3], while possibilistic clustering relaxes the probabilistic partition constraint and interprets memberships as typicalities [4]. Those approaches establish an important principle used here: observations should not influence prototypes solely because the partition requires their memberships to sum to one.

2.2 Neutrosophic Clustering and Information

The single-valued neutrosophic representation $(T, I, F) \in [0,1]^3$ allows support, indeterminacy, and opposition to be recorded separately [5]. Guo and Sengur embedded neutrosophic memberships directly into c-means clustering and used indeterminacy to improve uncertain-data clustering [6]. Qin and Wang later developed similarity and entropy constructions for single-valued neutrosophic information [7]. ERN-FCM differs structurally from NCM: its I_i and F_i are *record-level diagnostic coordinates* derived from ordinary fuzzy memberships and robust distances, and the cluster-specific memberships remain the conventional u_{ik}. This makes the role of each uncertainty coordinate explicit in the prototype weight.

2.3 Evaluation Criterion

Cluster recovery is assessed externally by the adjusted Rand index, which corrects partition agreement for chance [8]. On uncontaminated data, the silhouette width provides an internal compactness/separation check independent of class labels [9]. These two criteria are used for evaluation only; neither enters the ERN-FCM update.

3. TWO FAILURE GEOMETRIES IN A FUZZY PARTITION

Let $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and let $V = \{v_k\}_{k=1}^K$ denote prototypes. Write

$$d_{ik}^2 = \|x_i - v_k\|_2^2, \qquad u_{ik} = \left[\sum_{\ell=1}^{K} \left(\frac{d_{ik}^2}{d_{i\ell}^2} \right)^{1/(m-1)} \right]^{-1}, \quad (1)$$

with the usual limiting convention when $d_{ik} = 0$. For $m = 2$, used in the experiments, membership is inversely proportional to squared distance after normalization.

3.1 Boundary Indeterminacy

The normalized membership entropy is

$$H_i = -\frac{1}{\log K} \sum_{k=1}^{K} u_{ik} \log u_{ik} \in [0,1]. \quad (2)$$

ERN-FCM uses a controlled indeterminacy amplitude

$$I_i = \alpha H_i, \qquad 0 < \alpha \leq 1. \quad (3)$$

Thus $I_i = 0$ for a crisp fuzzy assignment and $I_i = \alpha$ when $u_{ik} = 1/K$ for every k. The parameter α controls how strongly boundary ambiguity can attenuate a prototype update; the experiments use $\alpha = 0.5$.

3.2 Robust Remoteness

Entropy does not determine whether the observation is near any prototype. Define

$$\delta_i = \min_k d_{ik}^2, \quad (4)$$

and let $q_{.50}$ and q_q be empirical quantiles of $\{\delta_i\}$ with $q = 0.90$. A scale-free upper-tail coordinate is

$$\rho_i = [\frac{\delta_i - q_q}{q_q - q_{.50} + \varepsilon}]_+, \qquad F_i = 1 - e^{-\tau\rho_i}, \quad \tau > 0. \tag{5}$$

Hence the closest 90% of records have $F_i = 0$ at a given iteration, while progressively more remote records receive $F_i \uparrow 1$. The quantile difference replaces an absolute distance threshold and is recomputed from the current geometry.

3.3 Neutrosophic Truth and Unresolved Evidence

The two failure modes are fused multiplicatively:

$$T_i = (1 - I_i)(1 - F_i). \tag{6}$$

The triple $z_i = (T_i, I_i, F_i)$ is not forced to sum to one. Instead,

$$T_i + I_i + F_i = 1 + I_i F_i, \tag{7}$$

so excess mass occurs exactly when boundary ambiguity and remoteness coexist. The complementary unresolved score is

$$U_i = 1 - T_i = I_i + F_i - I_i F_i. \tag{8}$$

This identity is useful diagnostically: I_i can screen bridge observations, F_i can screen gross outliers, and U_i responds to either.

4. FIXED-POINT PROTOTYPE UPDATING

For fixed (U, T), define the adaptive weighted distortion

$$Q(V \mid U, T) = \sum_{i=1}^{n} T_i \sum_{k=1}^{K} u_{ik}^m \|x_i - v_k\|^2. \tag{9}$$

The prototype block has the closed-form minimizer

$$v_k^+ = \frac{\sum_{i=1}^n T_i u_{ik}^m x_i}{\sum_{i=1}^n T_i u_{ik}^m}. \tag{10}$$

For fixed T and V, the factor T_i multiplies every cluster term belonging to observation i and therefore cancels from the Lagrange equations for $\sum_k u_{ik} = 1$; the membership minimizer remains (1). ERN-FCM alternates these two exact conditional minimizers and then recomputes (I, F, T) from the updated fuzzy geometry. Because T itself changes, the procedure is a fixed-point reweighting scheme rather than minimization of one static objective.

5. MATHEMATICAL PROPERTIES

Proposition 1 (Bounds and diagnostic extrema)**.** *For every observation, $0 \le I_i \le \alpha$, $0 \le F_i < 1$, and $0 < T_i \le 1$. Moreover, $I_i = 0$ for a crisp partition, $I_i = \alpha$ for the uniform partition, $F_i = 0$ whenever $\delta_i \le q_q$, and $F_i \to 1$ as $\rho_i \to \infty$.*

Proof. The Shannon entropy of a K-category probability vector lies in $[0, \log K]$, which gives the first bound after normalization and multiplication by α. Equation (5) maps $\rho_i \in [0, \infty)$ to $[0, 1)$ monotonically. Equation (6) then gives $0 < T_i \le 1$ for finite ρ_i. □

Proposition 2 (Isotropic scale equivariance)**.** *If every observation is transformed as $x_i' = cx_i + b$ with $c > 0$, then ERN-FCM memberships and (T_i, I_i, F_i) are unchanged, while prototypes transform as $v_k' = cv_k + b$.*

Proof. Translation cancels in every difference. Under multiplication by c, all squared distances and both quantiles in (5) are multiplied by c^2; distance ratios in (1) and the normalized tail coordinate ρ_i are therefore invariant. The claim follows from (2)–(10). □

Proposition 3 (Conditional block optimality)**.** *For fixed T, the membership update (1) minimizes (9) with respect to each u_i under $u_{ik} \ge 0$ and $\sum_k u_{ik} = 1$, and (10) minimizes it with respect to each v_k.*

Proof. For observation i, the positive factor T_i multiplies all K cluster terms and therefore disappears from the membership first-order conditions. The resulting equations are the standard FCM equations. For V, differentiating (9) with respect to v_k gives $2\sum_i T_i u_{ik}^m (v_k - x_i) = 0$, yielding (10). □

Proposition 4 (Remote-point attenuation)**.** *For a record whose robust remoteness $\rho_i \to \infty$, the multiplicative prototype factor satisfies*

$$T_i u_{ik}^m \le e^{-\tau\rho_i} u_{ik}^m \to 0.$$

Thus its direct contribution to every numerator and denominator in (10) vanishes exponentially in ρ_i.

Proof. Because $1 - I_i \le 1$ and $1 - F_i = e^{-\tau\rho_i}$, Equation (6) gives $T_i \le e^{-\tau\rho_i}$. Since $0 \le u_{ik}^m \le 1$, the result follows. □

Remark 1. *The propositions do not imply monotone descent of a single global objective after T is recomputed. ERN-FCM is therefore reported as a fixed-point reweighting method. Numerical center-shift diagnostics are included in Section 7 rather than replacing this limitation with an unsupported convergence theorem.*

Algorithm 1 Entropy–Remoteness Neutrosophic Fuzzy c-Means

Require: Data X, clusters K, fuzzifier m, α, τ, quantile q, tolerance ε_V

Initialize $V^{(0)}$ by k-means++ seeds; set $t \leftarrow 0$
repeat
  Compute d_{ik}^2 and fuzzy memberships $u_{ik}^{(t)}$ by (1)
  Compute $I_i^{(t)} = \alpha H_i^{(t)}$ from (2)
  Set $\delta_i^{(t)} = \min_k d_{ik}^2$ and obtain $q_{.50}^{(t)}, q_q^{(t)}$
  Compute $F_i^{(t)}$ by (5) and $T_i^{(t)} = (1 - I_i^{(t)})(1 - F_i^{(t)})$
  for $k = 1, \ldots, K$ **do**
    Update $v_k^{(t+1)}$ by (10)
  end for
  $t \leftarrow t + 1$
until $\|V^{(t)} - V^{(t-1)}\|_F < \varepsilon_V$ or $t = t_{\max}$
return V, U and diagnostic triples (T_i, I_i, F_i)

6. EXPERIMENTAL DESIGN

6.1 Two Different Uncertainty Geometries

The experiments intentionally avoid mixing all difficult observations into one contamination label. The first study uses

Table 1. Clean-data clustering over 20 initializations.

Data	Method	ARI	SD	Silh.
Iris	K-means	.6114	.0151	.4609
	FCM	.6303	.0000	.4584
	ERN-FCM	.5888	.0853	.4611
Wine	K-means	.9001	.0085	.2851
	FCM	.8975	.0000	.2849
	ERN-FCM	.9149	.0000	.2859
Breast cancer	K-means	.6599	.0089	.3439
	FCM	.6829	.0000	.3400
	ERN-FCM	.7189	.0000	.3285

the standardized Iris, Wine, and Breast Cancer benchmark matrices. For contamination fraction $\eta \in \{0.05, 0.10, 0.15\}$, round$(\eta n)$ gross points are appended in random directions at radii between 1.20 and 1.65 times the 95th percentile of the original Euclidean norms. Original labels are retained only for evaluating cluster recovery; the algorithms never see them.

The second study is a controlled boundary benchmark with three two-dimensional Gaussian clusters centered at $(-2.2, 0)$, $(2.2, 0)$, and $(0, 3.5)$, with 120 records per cluster. Bridge observations are generated near convex combinations $\lambda c_a + (1-\lambda)c_b$, $\lambda \in [0.43, 0.57]$, of two different centers. These bridge records are not remote by construction; the intended diagnostic is therefore I_i, not F_i. Contamination fractions are 0.05–0.20.

6.2 Comparators, Parameters, and Metrics

K-means, ordinary FCM, and ERN-FCM share standardized inputs and k-means++ initialization. The proposed method uses $m = 2$, $\alpha = 0.50$, $\tau = 0.80$, and $q = 0.90$ before comparison. Clean-data performance is summarized by adjusted Rand index (ARI) [8] and silhouette width [9]. Under contamination, ARI is computed only on the original records, so an algorithm is not rewarded for assigning synthetic contaminants to an arbitrary class. Gross-outlier screening is measured by AUC(F) and AUC(U); bridge screening is measured by AUC(I) and AUC(U).

Real-data results use 20 paired random initializations per condition. The boundary experiment uses 50 replicates. For each real dataset and contamination rate, a paired nonparametric bootstrap with 5000 resamples estimates the 95% interval of $\Delta\text{ARI} = \text{ARI}_{\text{ERN-FCM}} - \text{ARI}_{\text{FCM}}$. A component ablation replaces T_i by 1, $1 - I_i$, or $1 - F_i$. Sensitivity is evaluated on $\alpha \in \{0.25, 0.50, 0.75\}$ and $\tau \in \{0.4, 0.8, 1.2\}$.

7. NUMERICAL RESULTS

7.1 Clean Clustering Is Not Uniformly Improved

Table 1 is important because it prevents robustness gains from being interpreted without their clean-data cost. ERN-FCM improves Wine ARI from 0.8975 to 0.9149 and Breast Cancer from 0.6829 to 0.7189, but on Iris its mean ARI is 0.5888, below FCM at 0.6303. Silhouette values are close on Iris and Wine; for Breast Cancer, the higher external agreement of ERN-FCM comes with a lower silhouette (0.3285 versus 0.3400), indicating that label recovery and Euclidean compactness are not interchangeable objectives.

Table 2. Gross-contamination results. ARI is evaluated on original records; AUCs are ERN-FCM diagnostics.

Data	η	K-means	FCM	ERN-FCM	AUC(U)
Iris	.05	.6119	.6337	.6096	.9965
	.10	.6148	.6328	.6179	.9922
	.15	.6097	.6296	.6290	.9877
Wine	.05	.8949	.8808	.9131	.9986
	.10	.8818	.8665	.8975	.9956
	.15	.8786	.8547	.8860	.9891
Breast cancer	.05	.6637	.6939	.7137	.9923
	.10	.6681	.7014	.7074	.9898
	.15	.6663	.7039	.7060	.9816

Table 3. Paired bootstrap contrast $\Delta\text{ARI} = \text{ARI}_{\text{ERN-FCM}} - \text{ARI}_{\text{FCM}}$.

Data	η	Mean	95% interval	Win rate
Iris	.05	-.0240	[-.0541,-.0028]	.20
	.10	-.0149	[-.0394,-.0004]	.35
	.15	-.0006	[-.0038,.0030]	.30
Wine	.05	.0323	[.0282,.0363]	1.00
	.10	.0311	[.0259,.0361]	.95
	.15	.0313	[.0245,.0385]	.95
Breast cancer	.05	.0197	[.0185,.0209]	1.00
	.10	.0060	[.0030,.0090]	.75
	.15	.0021	[.0000,.0039]	.60

7.2 Gross Outliers: Recovery and Screening

Table 2 reports the main gross-contamination experiment. Wine shows the clearest recovery gain: at $\eta = 0.10$, ERN-FCM gives ARI $= 0.8975$ compared with 0.8665 for FCM and 0.8818 for k-means. Breast Cancer gains are smaller but consistent through all three rates. Iris behaves differently: ERN-FCM is worse than FCM at 5% and 10% and becomes essentially tied by 15%. This dataset-specific reversal is retained because it reveals the cost of attenuating high-entropy points when those points are structurally useful to a compact, overlapping partition.

The diagnostic signal is more stable than the clustering gain. At $\eta = 0.10$, AUC(U) equals 0.9922, 0.9956, and 0.9898 for Iris, Wine, and Breast Cancer, respectively. Even when the remoteness-only score F weakens at $\eta = 0.15$ because appended points alter the empirical upper-tail reference, the combined $U = I + F - IF$ remains above 0.98 on all three datasets.

7.3 Paired Uncertainty in the ARI Differences

Table 3 gives paired bootstrap contrasts. Wine has positive intervals at every contamination rate, including $\Delta\text{ARI} = 0.0311$ with 95% interval $[0.0259, 0.0361]$ at 10%. Breast Cancer is also positive at 5% and 10%, while the 15% interval begins effectively at zero. Iris reverses sign: at 10%, $\Delta\text{ARI} = -0.0149$ with interval $[-0.0394, -0.0004]$. The table therefore supports conditional rather than universal superiority.

7.4 Boundary Ambiguity Is a Different Diagnostic Problem

The bridge experiment in Table 4 isolates ambiguity without creating radially remote points. Cluster recovery is nearly perfect for both FCM and ERN-FCM, so there is little scope for an ARI gain. The meaningful result is instead diagnostic: AUC(I) remains approximately 0.996 from 5% through 20% bridge contamination. The combined AUC(U) remains around 0.993 for ERN-FCM. Thus the same method produces a different dominant coordinate depending on contamination geometry: F is informative for gross radial points, whereas I

Table 4. Controlled boundary benchmark (50 replicates).

η	Method	ARI	AUC(I)	AUC(U)
.05	FCM	.9982	.9962	.9921
	ERN-FCM	.9980	.9961	.9931
.10	FCM	.9982	.9964	.9925
	ERN-FCM	.9982	.9963	.9936
.15	FCM	.9982	.9964	.9922
	ERN-FCM	.9982	.9963	.9938
.20	FCM	.9982	.9963	.9913
	ERN-FCM	.9982	.9963	.9935

Table 5. Component ablation at 10% contamination.

Setting	FCM	$1-I$	$1-F$	ERN-FCM
Iris / gross	.6328	.6326	.6247	.6288
Wine / gross	.8700	.8866	.8787	.9010
Breast cancer / gross	.6969	.7029	.7026	.7065
Boundary benchmark	.9994	.9994	.9994	.9994

is informative for bridges.

7.5 Ablation and Parameter Sensitivity

Table 5 separates the two gates at 10% gross contamination. On Wine, entropy gating alone raises ARI from 0.8700 to 0.8866, remoteness gating gives 0.8787, and their conjunction reaches 0.9010. On Breast Cancer both gates help modestly and the conjunction is again best. Iris is the counterexample: remoteness attenuation lowers recovery, and the full model does not exceed FCM. On the easy boundary benchmark, all variants already recover the three generating clusters, so the value of entropy lies in diagnosis rather than in changing ARI.

Table 6 shows that the pooled mean ARI across the three real 10% gross-contamination settings and the 10% boundary setting varies only from 0.8021 to 0.8089 over the tested grid. The default $(\alpha,\tau)=(0.50,0.80)$ lies inside this plateau rather than at an isolated optimum. Increasing τ has little effect once upper-tail points are already strongly suppressed; similarly, $\alpha=0.50$ and 0.75 are almost indistinguishable in pooled recovery.

7.6 Fixed-Point Behavior

The empirical stopping behavior is summarized in Table 7. Median iteration counts are 24 for Iris and 18 for both Wine and Breast Cancer. Twenty-nine of the 30 displayed runs reach the 10^{-5} center-shift tolerance before the 60-iteration cap. One Iris run reaches the cap with final shift 2.8×10^{-3}, confirming why the paper does not claim global convergence. Mean I is largest for Wine (0.4145), while mean F remains small because the remoteness gate is deliberately activated only above the 90th percentile.

8. DISCUSSION

The numerical evidence points to a distinction between *robust clustering* and *uncertainty diagnosis*. ERN-FCM improves prototype recovery when a subset of observations is both atypical and influential, as in Wine and Breast Cancer. The mechanism is visible in (10): multiplying u_{ik}^m by T_i reduces the leverage of points whose membership entropy, remoteness, or both are large. In Wine at 10% contamination, the paired improvement 0.0311 is not a consequence of a hard trimming quota; it arises from continuous weights $T_i\in(0,1]$.

Table 6. Sensitivity of pooled mean ARI.

α	$\tau=.4$	$\tau=.8$	$\tau=1.2$
.25	.8021	.8033	.8048
.50	.8085	.8084	.8089
.75	.8075	.8075	.8078

Table 7. Fixed-point diagnostics at 10% gross contamination.

Data	Med. iter.	Max iter.	Max shift	$\bar{I}$	$\bar{F}$
Iris	24	60	2.8×10^{-3}	.2606	.0672
Wine	18	24	$<10^{-5}$	.4145	.0400
Breast cancer	18	23	$<10^{-5}$	.3933	.0322

Iris exposes the opposite regime. Its natural class overlap means that high fuzzy entropy is not necessarily nuisance variation. When informative boundary records receive $1-I_i<1$, the prototypes can move away from a partition that better matches the external class labels. The negative bootstrap intervals at 5% and 10% contamination quantify this cost. This is not merely a parameter failure: it demonstrates that "uncertain" and "uninformative" are not synonyms. A future adaptive version could let α depend on a cluster-validity diagnostic rather than fixing a common entropy penalty.

The diagnostic coordinates are more stable than the recovery effect. Gross radial contamination produces F_i and U_i AUCs near one because the points occupy the upper tail of δ_i. Bridge contamination behaves differently: it need not be remote, but it is constructed near competing cluster regions, so I_i is the informative coordinate. The identity $U_i=I_i+F_i-I_iF_i$ then serves as a soft logical union of these two failure modes. This separation is the main reason to retain the full neutrosophic triple rather than collapsing uncertainty into one scalar before interpretation.

Three limitations matter. First, the 90th-percentile reference in (5) assumes that a large majority of records are not gross outliers; if contamination approaches the reference quantile, F_i can lose contrast, as seen in the declining AUC(F) at 15%. Second, the method is Euclidean and isotropically scale-equivariant, not invariant to arbitrary feature-wise rescaling; standardization is therefore part of the experimental protocol. Third, recomputation of (I,F,T) breaks the monotonicity guarantee available for each fixed-T surrogate. The observed fixed-point stability is encouraging but is not a substitute for a global convergence theorem.

9. CONCLUSION

This paper separated two geometries that ordinary fuzzy partition softness can conflate. Starting from memberships u_{ik}, entropy produces boundary indeterminacy I_i, robust upper-tail distance produces remoteness falsity F_i, and their conjunction gives truth $T_i=(1-I_i)(1-F_i)$. The resulting prototype update is a truth-weighted FCM step, while the same triple supplies diagnostic quantities for uncertain records. Mathematically, the construction is bounded, isotropically scale-equivariant, conditionally optimal for fixed truth weights, and exponentially suppresses sufficiently remote records.

The experiments support a qualified conclusion. ERN-FCM improves cluster recovery on Wine and Breast Cancer under gross contamination but not consistently on Iris. Its stronger and more portable result is diagnostic: U_i detects gross outliers with AUC near one in the main settings, while I_i detects

controlled bridge ambiguity with AUC about 0.996. The method is therefore best viewed as a clustering-and-diagnosis layer for datasets in which one must distinguish *between-cluster uncertainty* from *outside-cluster atypicality*. Future work can replace the fixed quantile and entropy amplitude by data-adaptive rules and investigate a global objective whose reweighting map inherits a formal descent property.

REFERENCES

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[3] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, no. 11, pp. 657–664, 1991.

[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] H. Wang, F. Smarandache, Y. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," *Multispace and Multistructure*, vol. 4, pp. 410–413, 2010.

[6] Y. Guo and A. Sengur, "NCM: Neutrosophic c-means clustering algorithm," *Pattern Recognition*, vol. 48, no. 8, pp. 2710–2724, 2015.

[7] K. Qin and L. Wang, "New similarity and entropy measures of single-valued neutrosophic sets with applications in multi-attribute decision making," *Soft Computing*, vol. 24, pp. 16 165–16 176, 2020.

[8] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, pp. 193–218, 1985.

[9] 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.