Neutrosophic Computing and Machine Learning, Vol. 45, 2026

University of New Mexico

Non-Normalization Is Not Independence: An Empirical Audit of the Indeterminacy Coordinate in Five Neutrosophication Methods on Clinical Data

La no normalización no es independencia: una auditoría empírica de la coordenada de indeterminación en cinco métodos de neutralización en datos clínicos.

Maikel Yelandi Leyva-Vázquez [1,2,3,*], Lorenzo Cevallos-Torres [2], Omar Mar Cornelio [4], Alexis Matheu Pérez [3] and Florentin Smarandache [5]

1 Universidad Bolivariana del Ecuador, Guayaquil, Ecuador. E-mail: myleyvav@ube.edu.ec. ORCID: https://orcid.org/0000-0001-7911-5879

2 Universidad de Guayaquil, Guayaquil, Ecuador.

3 Centro de Investigación Institucional, Universidad Bernardo O'Higgins, Santiago, Chile. E-mail: alexis.matheu@ubo.cl

4 Universidad de las Ciencias Informáticas, La Habana, Cuba.

5 University of New Mexico, Mathematics, Physics and Natural Science Division, Gallup, NM 87301, USA. E-mail: smarand@unm.edu

* Corresponding author: myleyvav@ube.edu.ec

Abstract. Neutrosophication converts a crisp value into a triple (T, I, F) of truth, indeterminacy and falsity degrees. A common way of arguing that a method is "truly neutrosophic" is to observe that $T + I + F \neq 1$ or that $F \neq 1 - T$. Neither observation shows that the third coordinate carries information not already contained in the other two. We audit five neutrosophication transformations (a proposed K-Means + sigmoid method, and the parabolic, threshold-distance, kernel-density and triangular-fuzzy methods) on six clinical attributes of the Cleveland heart-disease data ($n = 299$), separating normalization, complementarity, data adaptivity, empirical non-redundancy, stability and downstream utility as distinct, separately tested properties. An equation-level audit shows that all five transformations are functions of a single scalar and that four of them derive I analytically from T. Cross-validated regressions of I on (T, F) give out-of-sample R^2 of 1.00, 0.99, 1.00 and 0.98 for the four complementary methods and 1.00 for K-Means at full numerical precision, falling to 0.48 when T and F are recorded to two decimals: the K-Means coordinate is non-redundant only through the numerically negligible tails of saturated sigmoids (T, $F < 0.01$ for 63 % of observations). A nearest-neighbour test confirms the pattern. In a leakage-safe, same-classifier ablation, adding I to (T, F) changed the ROC-AUC of a logistic regression by -0.009 (95 % -0.035 to +0.017) for K-Means and by at most 0.008 for the other methods, and no neutrosophic representation exceeded the raw six attributes by more than 0.006 AUC with any of three classifiers. The K-Means I correlated weakly with external uncertainty signals (AUC 0.56 against misclassification), and two ambiguity-based alternatives were not better. K-Means results were stable across seeds with k-means++ but not with the original random initialisation, and sensitive to the sigmoid slope, to the choice of K and to scaling. The main lesson is methodological: not every three-component representation contains three independent dimensions of information, and neutrosophication methods should not be evaluated by $T + I + F$ alone. Because a univariate transformation cannot add information about its input, the constructive consequence is that future indeterminacy coordinates must be grounded in separately auditable evidence (reliability, additional sources, resampling, time, modalities, agents); the paper closes with a prior-art-audited taxonomy of such evidence-grounded designs, ten design principles and a validation checklist.

Keywords: neutrosophication; neutrosophic sets; indeterminacy; functional redundancy; normalization; complementarity; K-means; uncertainty quantification; medical data

Resumen. La neutrosofización convierte un valor nítido en una tripleta (T, I, F) de grados de verdad, indeterminación y falsedad. Una forma común de argumentar que un método es "verdaderamente neutrosófico" es observar que T + I + F ≠ 1 o que F ≠ 1 - T. Ninguna de estas observaciones muestra que la tercera coordenada contenga información que no esté ya contenida en las otras dos. Analizamos cinco transformaciones de neutrosofización (un método propuesto de K-Means + sigmoide,

y los métodos parabólico, de distancia umbral, de densidad kernel y triangular difuso) en seis atributos clínicos de los datos de cardiopatías de Cleveland (n = 299), separando la normalización, la complementariedad, la adaptabilidad de los datos, la no redundancia empírica, la estabilidad y la utilidad posterior como propiedades distintas, probadas por separado. Una auditoría a nivel de ecuación muestra que las cinco transformaciones son funciones de un único escalar y que cuatro de ellas derivan I analíticamente de T. Las regresiones con validación cruzada de I sobre (T, F) dan un R^2 fuera de muestra de 1,00, 0,99, 1,00 y 0,98 para los cuatro métodos complementarios y 1,00 para K-Means con precisión numérica completa, que cae a 0,48 cuando T y F se registran con dos decimales: la coordenada K-Means no es redundante solo a través de las colas numéricamente despreciables de sigmoides saturadas (T, F < 0,01 para el 63 % de las observaciones). Una prueba del vecino más cercano confirma el patrón. En una ablación segura de fugas y del mismo clasificador, agregar I a (T, F) cambió el ROC-AUC de una regresión logística en -0,009 (IC del 95 %: -0,035 a +0,017) para K-Means y en un máximo de 0,008 para los otros métodos, y ninguna representación neutrosófica superó los seis atributos brutos en más de 0,006 AUC con ninguno de los tres clasificadores. El K-Means I se correlacionó débilmente con las señales de incertidumbre externas (AUC 0,56 frente a la clasificación errónea), y dos alternativas basadas en la ambigüedad no fueron mejores. Los resultados de K-Means fueron estables entre semillas con k-means++ pero no con la inicialización aleatoria original, y sensibles a la pendiente sigmoide, a la elección de K y al escalado. La lección principal es metodológica: no toda representación de tres componentes contiene tres dimensiones de información independientes, y los métodos de neutrosofización no deben evaluarse solo con T + I + F. Dado que una transformación univariada no puede añadir información sobre su entrada, la consecuencia constructiva es que las coordenadas de indeterminación futuras deben fundamentarse en evidencia auditable por separado (fiabilidad, fuentes adicionales, remuestreo, tiempo, modalidades, agentes). El artículo concluye con una taxonomía, revisada por el estado de la técnica, de dichos diseños basados en evidencia, diez principios de diseño y una lista de verificación de validación.

Palabras clave: neutrosofización; conjuntos neutrosóficos; indeterminación; redundancia funcional; normalización; complementariedad; K-medias; cuantificación de la incertidumbre; datos médicos

1 Introduction

Neutrosophic sets extend fuzzy sets by attaching to each element three degrees, truth T, indeterminacy I and falsity F, whose sum is not constrained to one [1, 2, 3]. Applying neutrosophic models to numerical data requires a neutrosophication step that maps a crisp value to a triple. Image-processing work introduced the recipes still in use: a truth degree from a local mean, an indeterminacy from a local statistic and a falsity equal to 1 − T [4, 5, 6], clustering on the neutrosophic domain [7], and neutrosophic c-means with a unit-sum constraint [8]. Statistical and decision-making applications have added parabolic, threshold and membership-function recipes [3]. A comparative evaluation of these transformations was proposed by the present authors in a preprint [9]; to our knowledge it remains the only comparative study, and no peer-reviewed comparison exists.

That preprint, like much of the applied literature, argued that a method is "truly neutrosophic" when $T + I + F \neq 1$ and $F \neq 1 - T$. This paper starts from the observation that neither fact establishes what it is taken to establish. Whether a representation is normalized ($T + I + F = 1$) is one property; whether F is the complement of T is a second; whether I is a function of (T, F), so that the triple is the graph of a function over the polar square and adds no degree of freedom, is a third; whether that third coordinate carries information usable by a downstream task is a fourth; and whether the transformation is stable and data-adaptive are still others. The theoretical companion of this paper formalizes the first three distinctions: a derived coordinate I = f(T, F) is representationally redundant, a primitive coordinate is representationally independent when states with equal (T, F) and different I exist, and normalization is a separate operation that quotients the state space along rays and deletes the magnitude $T + I + F$ [10]. It also shows that representational independence does not imply inferential relevance: a coordinate may distinguish states without changing what follows from them. The present paper does not enter inference; it asks the empirical question that precedes it.

The scientific question is therefore reformulated. Instead of asking which method is the most neutrosophic, we ask how common neutrosophication procedures differ in normalization, complementarity, data adaptivity, functional redundancy and downstream information preservation, and, as a testable sub-question, whether the proposed K-Means representation generates an empirically non-redundant third coordinate beyond the polar pair (T, F), and whether that coordinate is useful for a task. The K-Means method may pass, pass partially or fail; all three outcomes are reported as found.

The contributions are: (i) a systematic comparison of five transformations on one public medical data set with identical preprocessing; (ii) an explicit separation of normalization, complementarity, data adaptivity and empirical non-redundancy, with an equation-level audit; (iii) empirical non-redundancy tests, a cross-validated predictability of I from (T, F) at full and at finite resolution, and a nearest-neighbour test with pre-declared tolerances; (iv) a leakage-safe, same-classifier downstream ablation of [T, F] against [T, I, F], raw attributes and normalized coordinates, with repeated cross-validation and paired effect sizes; (v) stability and sensitivity analyses (initialisation seed, number of clusters, sigmoid slope, scaling, a discrete attribute) and an external valida-

tion of I against misclassification, model disagreement and bootstrap instability, including two ambiguity-based alternatives to the original definition. Section 2 fixes the vocabulary, Section 3 the data and methods, Section 4 the results, Section 5 the discussion and Section 6 the conclusions.

2 Normalization, Complementarity, Redundancy and Independence

Let a transformation map a scaled value $x \in [0, 1]$ to $(T, I, F) \in [0, 1]^3$. The following properties are distinct and are tested separately in this paper.

Normalization. $T + I + F = 1$ is imposed. None of the five methods examined imposes it; the sum is a magnitude, and normalization to the simplex deletes it [10, Theorem 8.1]. A sum below or above one is a magnitude fact, not a quality.

Complementarity. $F = 1 - T$. It holds by construction for four of the five methods; its absence (non-complementarity) is a fact about F, not about I.

Functional redundancy. I = f(T, F) for some function f. Then the state set is the graph of f over the polar square and the third coordinate adds no degree of freedom [10, Definition 3.4 and Theorem 3.8]. Representational independence is the contrary: states with equal (T, F) and different I exist [10, Theorem 3.9]. In data these are tested empirically, relative to a class of function approximators and a measurement resolution; a low out-of-sample R^2 is empirical non-redundancy relative to the tested classes, not mathematical independence.

Statistical independence is a probabilistic notion (independence of random variables) and is neither tested nor claimed. Inferential independence concerns whether a coordinate changes logical consequence [10, Theorem 5.8]; a coordinate can be representationally distinct yet inferentially inert. Nothing in this paper bears on inference.

2.1 Three senses of independence and the polar-fibre criterion

Three properties are easily conflated under the word "independent", and the manuscript keeps them apart. Formal representational independence is a property of a state set: a coordinate I is representationally redundant when it is a function of the other coordinates on that set, and representationally independent when two states agree on the other coordinates and differ on I [10, Definition 3.4]. Empirical informational novelty is a property of a data set, a transformation and a numerical resolution: it holds when I cannot be recovered from (T, F) by a flexible learner at the resolution at which the data are recorded. Independent evidence provenance is a property of the measurement process: it holds when I is computed from an input that is not itself a function of the scalar x from which T and F are computed. The first can hold without the third (a primitive I in a formal state space), the second can hold without the third (Section 4.2 shows how), and only the third can add information about the proposition rather than about x. "Independent" in the third sense means separately grounded; it does not mean statistically independent.

For a univariate transformation $\Phi(x) = (T(x), I(x), F(x))$ write $P(x) = (T(x), F(x))$ for the polar map. The companion's redundancy result [10, Theorems 3.8–3.9] has, in this setting, the following operational form, which we state as an application rather than as new mathematics.

Polar-fibre criterion. I is a function of the polar pair on $\Phi(X)$, that is $I = f(T, F)$ for some f, if and only if I is constant on every fibre of P: whenever $P(x_1) = P(x_2)$, $I(x_1) = I(x_2)$. Equivalently, I fails to be a function of (T, F) exactly when some pair x_1, x_2 has $T(x_1) = T(x_2)$, $F(x_1) = F(x_2)$ and $I(x_1) \neq I(x_2)$. (If I is constant on fibres, $f(t, f) := I(x)$ for any x in the fibre is well defined; the converse is immediate. This is the elementary fact that a map factors through another map exactly when it is constant on the latter's fibres, and the collision pair of [10, Theorem 3.9] is a fibre carrying two values of I.)

Corollary (injective polar map). If P is injective on X, every fibre is a single point, so $I = I \circ P^{-1}$ is a function of (T, F) on P(X), and the indeterminacy coordinate of a deterministic univariate neutrosophication carries no information about x beyond the polar pair. The corollary is elementary; its use here is that it turns an empirical question into a checkable property of the transformation. The T and F of Equation (1) are strictly monotone in x, so the K-Means polar map is injective in exact arithmetic and its I is provably a function of (T, F); non-redundancy can only appear when injectivity is lost, which is what rounding does to saturated sigmoids. The four complementary methods have T = x or T strictly monotone in x, so their polar maps are injective too, and their I is a function of T alone.

The consequence for design is the central methodological point of this paper. If x is the only informational input, no transformation Φ can create information beyond x; a third coordinate can at most re-encode information that the polar projection discarded. Independent evidential meaning cannot be obtained merely by adding another deterministic transformation of the same scalar observation; it requires either information discarded by the polar projection or an additional evidential input. The first route is what Section 4 audits; the second is the subject of Section 5.4.

3 Materials and Methods

3.1 Data

The Cleveland heart-disease data of the UCI repository [11] contain 303 records; 299 are complete on the six continuous or ordinal attributes used here: age, resting systolic blood pressure, serum cholesterol, maximum heart rate, ST depression induced by exercise and number of major vessels coloured by fluoroscopy (an integer in 0–3). Descriptive statistics computed on the 299 complete cases are: age 54.53 ± 9.02 years; resting pressure 131.67 ± 17.71 mmHg; cholesterol 247.1 ± 51.91 mg/dL; maximum heart rate 149.51 ± 22.95 bpm; ST depression 1.05 ± 1.16 mm; vessels 0.67 ± 0.94. (The cholesterol and heart-rate figures in the preprint [9] were those of the 303-record file.) The binary label used in the downstream task is presence of disease (num > 0; 46.5 %). Each attribute is min-max scaled to [0, 1]; whenever a transformation is evaluated out of sample, scaling constants, centroids and densities are estimated on the training folds only [12, 13].

3.2 The five transformations

All formulas are those implemented in the neutrolab library, version 1.4.0 [14], which produced the results of the preprint; $\varepsilon = 10^{-6}$. K-Means + sigmoid (proposed in [9]): one-dimensional K-means with $K = 3$ gives sorted centroids $c_{low} < c_{mid} < c_{high}$ and within-cluster standard deviations σ; with slope $s = 10$,

$$T(x) = \frac{1}{1 + exp\left(-s\,(x - c_{high}) \,/\, (\sigma_{high} + \varepsilon)\right)},\quad I(x) = \frac{1}{1 + |x - c_{mid}| \,/\, (\sigma_{mid} + \varepsilon)},\quad F(x) = \frac{1}{1 + exp\left(s\,(x - c_{low}) \,/\, (\sigma_{low} + \varepsilon)\right)} \tag{1}$$

Parabolic [3], with α = 0.5:

$$T(x) = x,\ I(x) = 4\alpha x(1 - x),\ F(x) = 1 - x \tag{2}$$

Threshold distance [6], with θ = 0.5 and $\lambda = 5$:

$$T(x) = \frac{1}{1 + exp\left(-10(x - \theta)\right)},\ I(x) = exp(-\lambda(x - \theta)^2),\ F(x) = 1 - T(x) \tag{3}$$

Kernel density estimation, with a Gaussian kernel, Silverman's bandwidth and the density $\hat{f}$ min-max normalized:

$$T(x) = x,\ I(x) = 1 - f_{norm}(x),\ F(x) = 1 - x \tag{4}$$

Triangular fuzzy membership [15], with sets L = (−0.5, 0, 0.5), M = (0.25, 0.5, 0.75) and H = (0.5, 1, 1.5):

$$T(x) = x,\ I(x) = 1 - max\{\mu_L(x), \mu_M(x), \mu_H(x)\},\ F(x) = 1 - x \tag{5}$$

Equation-level audit. In (2)–(5) $F = 1 - T$ and T = x (or a fixed monotone function of x), so I is an analytic function of T: 4αT(1 − T) in (2); exp(−λ(logit(T)/10)²) in (3); a deterministic, data-dependent function of T in (4); a piecewise-linear function of T in (5). These four coordinates are redundant by construction. In (1) T and F are logistic functions with different centres and spreads, so $F \neq 1 - T$, and I is a function of x but has no closed form in (T, F). With s/σ of order 100 on the unit scale, T and F saturate: over the middle of the range both are numerically close to zero while I varies. Two parameter artefacts are also visible: $\lambda = 5$ in (3) makes I ≥ exp(−1.25) = 0.29 everywhere, and the supports in (5) make max $\mu \geq 0.5$ and hence $I \leq 0.5$.

3.3 Descriptive properties and the complementarity index

For every method and attribute we report the means of T, I and F; the mean, median, standard deviation, minimum, maximum and 5th and 95th percentiles of $T + I + F$; and, in place of the preprint's "T + F consistency", the complementarity index

$$CI = 1 - \frac{1}{n}\sum_{i=1}^{n} |\,T_i + F_i - 1\,| \tag{6}$$

which equals 1 under complementarity and measures only closeness to $F = 1 - T$. Range and entropy of I are reported as descriptors, not as quality metrics.

3.4 Empirical non-redundancy tests

Predictability test. For each method and attribute, and pooled over attributes, I is regressed on (T, F) with eight learners (linear, polynomial of degree 2 and 4, ridge, random forest, gradient boosting, k-nearest neighbours, a small multilayer perceptron) under 5-fold cross-validation; out-of-sample R^2, MAE and RMSE are reported. As a control, I is regressed on the scalar x. Because a mathematically exact dependence may be recoverable only through numerically negligible values, the test is repeated with T and F rounded to three and to two decimals (a resolution-aware variant). Nearest-neighbour test. Among 20,000 random pairs of observations (pooled over attributes), pairs with $|\Delta T| < \varepsilon$ and $|\Delta F| < \varepsilon$, for pre-declared $\varepsilon \in \{0.01, 0.02, 0.05\}$, are compared with all pairs on $|\Delta I|$. Dimensionality. Principal components of the standardized (T, I, F) and the rank of their covariance are reported as complementary evidence only.

On the data, all five polar maps are injective at full precision on the 378 distinct attribute values (no fibre carries two values of I), which is why the exact-precision R^2 of Section 4.2 is close to one for every method; at a resolution of 0.01 the K-Means polar map collapses the 378 values onto 100 pairs, 21 of which carry ranges of I up to 1.00, whereas the four complementary methods collapse onto 233–262 pairs with ranges of I at most 0.13 (results/polar_injectivity.csv).

3.5 Downstream ablation (same classifier, leakage-safe)

Three classifiers (logistic regression with standardization, random forest with 200 trees, gradient boosting; scikit-learn defaults otherwise) are trained on eight feature sets: the six raw scaled attributes; for each method, [T, F] (12 features) and [T, I, F] (18 features); and for K-Means, [T, I, F] normalized to unit sum. Evaluation uses 5-fold stratified cross-validation repeated 5 times (25 folds, identical folds for all feature sets). The neutrosophication is a pipeline step fitted on the training folds only. Metrics: accuracy, balanced accuracy, macro-F1, ROC-AUC and Brier score. The quantity of interest is the paired difference Δ = performance([T, I, F]) − performance([T, F]) on identical folds, with its mean, 2.5–97.5 percentile range over folds, a Wilcoxon signed-rank p-value and Cohen's d; percentile ranges over correlated folds are descriptive [16, 17].

3.6 External validation and alternative definitions of I

Three external signals of uncertainty were computed from the raw attributes: misclassification of the random forest in out-of-fold prediction; disagreement among the three classifiers' out-of-fold predictions; and instability of the random-forest prediction over 100 bootstrap refits. For each attribute the K-Means I was compared with these signals by ROC-AUC and Spearman correlation. Two ambiguity-based alternatives were computed from the same fitted centroids, where d_(1) ≤ d_(2) are the distances to the two nearest centroids and w_k are softmax memberships with scale σ_k:

$$I_{amb}(x) = 1 - \frac{d_{(2)} - d_{(1)}}{d_{(2)} + d_{(1)}}, \quad I_{ent}(x) = -\frac{1}{\ln K}\sum_{k=1}^{K} w_k(x) \ln w_k(x) \qquad (7)$$

3.7 Stability and sensitivity

K-Means was refitted with 50 seeds under two initialisations: the random-point initialisation without restarts implemented in neutrolab, and k-means++ with 10 restarts [18]; centroid and coordinate variability are reported, together with the downstream AUC over 10 seeds. Cluster validity for $K = 2$ to 5 was measured with the silhouette [19], Calinski–Harabasz and Davies–Bouldin indices. The sigmoid slope was varied over {2, 5, 10, 20}; min-max scaling was compared with 1–99 % winsorized min-max and with rank (quantile) scaling. The attribute with four integer values was examined separately.

4 Results

4.1 Descriptive properties

Table 1 gives the pooled statistics of $T + I + F$. For K-Means the mean is 0.638 but the median is 0.500, the sum ranges from 0.001 to 1.210 and exceeds one in 22 % of observations; the preprint's two statements, that the sum "consistently exceeds 1" and that it averages 0.639, were therefore both incomplete. For the complementary methods the sum equals 1 + I and is at least one by arithmetic. The complementarity index is 0.29 for K-Means and 1.00 for the others: non-complementarity, nothing more.

Table 1: Magnitude $T + I + F$ pooled over the six attributes ($n = 1{,}794$ values per method), complementarity index and descriptors of I.

Method	mean	median	SD	min	max	Q05	Q95	% > 1	Compl. index	mean I	range I	entropy I
K-Means	0.638	0.500	0.336	0.00	1.21	0.25	1.21	22	0.29	0.35	0.89	2.10
Parabolic	1.331	1.406	0.175	1.00	1.50	1.00	1.50	84	1.00	0.33	0.49	1.51

Threshold	1.711	1.790	0.247	1.29	2.00	1.29	2.00	100	1.00	0.71	0.69	1.97
KDE	1.306	1.227	0.298	1.00	2.00	1.00	1.88	100	1.00	0.31	1.00	2.30
Fuzzy	1.350	1.387	0.230	1.00	1.67	1.00	1.67	84	1.00	0.35	0.66	2.08

Note: range and entropy of I are descriptors, not quality metrics; entropy over 20 bins on [0, 1].

4.2 Empirical non-redundancy

Table 2 and Figure 1 give the predictability of I from (T, F). Per attribute, the best learner reaches out-of-sample R^2 of 1.000 (Parabolic), 0.988 (Threshold), 0.999 (KDE) and 0.981 (Fuzzy): their indeterminacy is fully recoverable, as the equation audit predicted. Pooled over attributes the KDE value drops to 0.78 only because the density is attribute-specific. For K-Means the per-attribute R^2 is 0.996 at full numerical precision (pooled 0.96, best learner k-nearest neighbours), that is, I is recoverable from (T, F). The recovery, however, runs through the tails: with T and F rounded to three decimals the mean R^2 falls to 0.62, and to 0.48 at two decimals, while the other methods keep $R^2 \geq 0.97$. For 63 % of the observations (attributes other than the vessel count) both T and F are below 0.01 while I varies with standard deviation about 0.18. The K-Means coordinate is thus empirically non-redundant at any practical resolution, and redundant in exact arithmetic; the difference is entirely due to saturation of the two logistic coordinates, which discard the position of x in the middle of the range. Regressing I on x directly gives $R^2 \geq 0.88$ for every method and attribute: all coordinates are functions of one scalar.

Table 2: Out-of-sample R^2 of I predicted from (T, F): pooled over attributes by learner (columns 2–8), mean over attributes of the best learner at exact, 0.001 and 0.01 resolution, and nearest-neighbour ratio mean |ΔI| (near pairs) / mean |ΔI| (random pairs).

Method	**linear**	**poly-2**	**poly-4**	**RF**	**GBM**	**kNN**	**MLP**	**best per attribute (exact)**	**rounded 0.001**	**rounded 0.01**	**NN ratio $\varepsilon = 0.02$**
K-Means	0.32	0.46	0.47	0.78	0.77	0.96	0.46	0.996	0.62	0.48	0.87
Parabolic	0.34	1.00	1.00	1.00	1.00	1.00	0.99	1.000	0.98	0.97	0.02
Threshold	0.18	0.84	0.97	1.00	1.00	1.00	0.91	0.988	0.99	0.98	0.17
KDE	0.12	0.12	0.24	0.78	0.60	0.48	0.29	0.999	1.00	1.00	0.53
Fuzzy	0.15	0.45	0.62	1.00	1.00	1.00	0.99	0.981	0.98	0.98	0.05

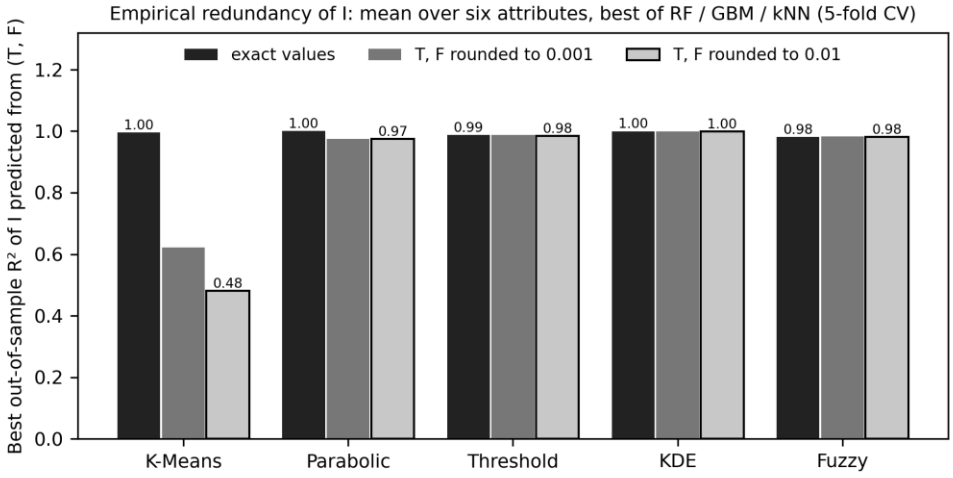

Figure 1: Out-of-sample R^2 of the indeterminacy coordinate predicted from (T, F) with eight learners, pooled over the six attributes.

The nearest-neighbour test (Table 2, last column; Figure 2) tells the same story at the level of pairs. For pairs whose T and F differ by less than 0.02, the mean |ΔI| is 0.004 for Parabolic, 0.014 for Fuzzy, 0.047 for Threshold and 0.174 for KDE, against 0.243 for K-Means, which is 0.87 of the value for random pairs: knowing (T, F) to two decimals tells almost nothing about I under K-Means, and 49 % of such pairs differ in I by more than 0.2. The pattern is unchanged for $\varepsilon = 0.01$ and 0.05. Principal components (Table 3) agree: the covariance of the standardized triple has rank 2 for the four complementary methods and rank 3 for K-Means, whose third component explains 13 % of variance. This is complementary evidence about geometry, not about logical independence.

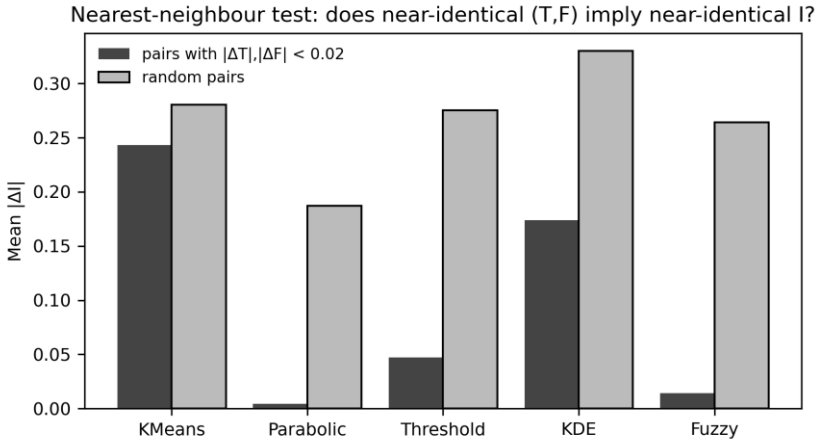

Figure 2: Mean |ΔI| for pairs with |ΔT| < 0.02 and |ΔF| < 0.02 versus random pairs.

Table 3: Principal-component structure of the standardized (T, I, F), pooled over attributes.

Method	**PC1**	**PC2**	**PC3**	**rank of covariance**	**participation ratio**
K-Means	0.49	0.38	0.13	3	2.50
Parabolic	0.82	0.18	0.00	2	1.41
Threshold	0.76	0.24	0.00	2	1.57
KDE	0.73	0.27	0.00	2	1.64
Fuzzy	0.75	0.25	0.00	2	1.60

4.3 Downstream ablation

Table 4: Same-classifier downstream ablation, ROC-AUC over 25 identical folds (5 × 5 stratified CV), transformations fitted on training folds only.

Method	**Classifier**	**AUC [T, F]**	**AUC [T, I, F]**	**Δ AUC (2.5–97.5 %)**	**Wilcoxon p**	**Cohen d**	**AUC [I only]**
K-Means	logreg	0.812	0.803	-0.009 (-0.035, +0.017)	0.009	-0.61	0.608
	rf	0.816	0.816	+0.000 (-0.017, +0.032)	0.711	+0.01	0.735
	gbm	0.793	0.790	-0.003 (-0.040, +0.048)	0.288	-0.14	0.693
Parabolic	logreg	0.840	0.839	-0.001 (-0.031, +0.029)	0.830	-0.05	0.795
	rf	0.811	0.806	-0.005 (-0.023, +0.028)	0.019	-0.40	0.789
	gbm	0.787	0.786	-0.000 (-0.034, +0.032)	0.968	-0.03	0.760
Threshold	logreg	0.826	0.833	+0.008 (-0.043, +0.046)	0.122	+0.30	0.798
	rf	0.810	0.805	-0.005 (-0.021, +0.025)	0.049	-0.37	0.788
	gbm	0.789	0.792	+0.003 (-0.033, +0.029)	0.353	+0.16	0.758
KDE	logreg	0.840	0.837	-0.003 (-0.044, +0.025)	0.677	-0.14	0.831
	rf	0.811	0.813	+0.002 (-0.027, +0.042)	0.677	+0.13	0.814
	gbm	0.787	0.791	+0.004 (-0.046, +0.045)	0.216	+0.17	0.776
Fuzzy	logreg	0.840	0.835	-0.005 (-0.030, +0.025)	0.264	-0.24	0.745
	rf	0.811	0.809	-0.002 (-0.019, +0.017)	0.476	-0.16	0.731
	gbm	0.787	0.781	-0.006 (-0.045, +0.027)	0.191	-0.29	0.713
raw attributes	logreg / rf / gbm	—	0.839 / 0.817 / 0.787	—	—	—	—
K-Means normalized	logreg / rf / gbm	—	0.815 / 0.816 / 0.793	+0.012 / -0.001 / +0.004 vs raw K-Means	—	—	—

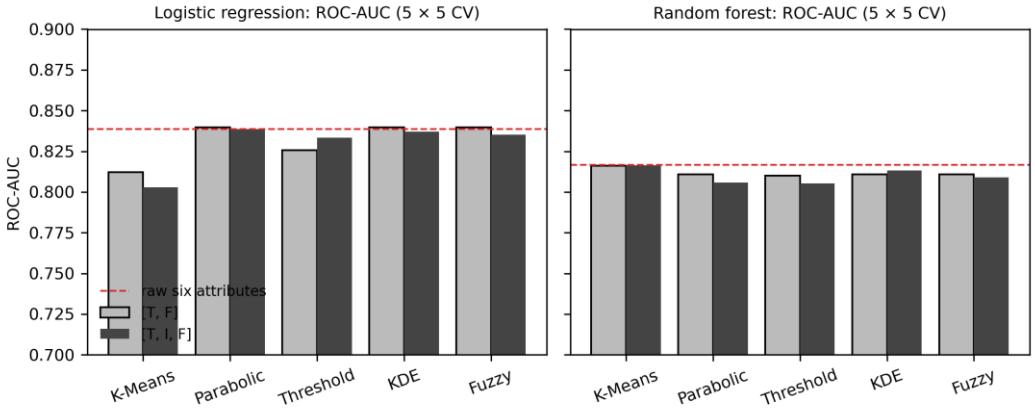

Figure 3: ROC-AUC of [T, F] and [T, I, F] representations for each method and two classifiers, with the raw-attribute reference (dashed).

Table 4 and Figure 3 give the ablation. With logistic regression, adding I to (T, F) changed the AUC by -0.009 (95 % -0.035 to +0.017) for K-Means, -0.001 (95 % -0.031 to +0.029) for Parabolic, +0.008 (95 % -0.043 to +0.046) for Threshold, -0.003 (95 % -0.044 to +0.025) for KDE and -0.005 (95 % -0.030 to +0.025) for Fuzzy. With the random forest the K-Means gain was +0.000 (95 % -0.017 to +0.032) and with gradient boosting -0.003 (95 % -0.040 to +0.048). For no classifier did the K-Means coordinate produce a paired gain whose 2.5–97.5 % range excluded zero. The raw six attributes reached AUC 0.839 (logistic), 0.817 (random forest) and 0.787 (gradient boosting); the best neutrosophic representation for each classifier was Parabolic [T, I, F] at 0.839, K-Means at 0.816 and Threshold at 0.792. Normalizing the K-Means triple to unit sum changed the AUC by +0.012 (logistic), -0.001 (random forest) and +0.004 (gradient boosting) relative to the raw triple: normalization deletes the magnitude [10] and, for a linear classifier, also the monotone information carried by the saturated coordinates. The I coordinate alone (last column) classifies far worse than (T, F) for every method.

4.4 External validation of I and alternative definitions

Averaged over the six attributes, the K-Means indeterminacy discriminated misclassified from correctly classified records with AUC 0.56, disagreeing from agreeing classifiers with AUC 0.57, and correlated with bootstrap instability at $\rho = 0.10$ and with proximity to the decision boundary at $\rho = 0.18$ (Table 5). The two ambiguity-based alternatives, I_{amb} (closeness to the boundary between two nearest clusters) and I_{ent} (entropy of soft memberships), gave AUC 0.55 and 0.53 against misclassification, and their predictability from (T, F) was R^2 0.91 and 0.93. None of the three definitions is a substantive external uncertainty signal at the attribute level: an attribute's position relative to its own clusters is not the model's uncertainty about the patient.

Table 5: External validation of three definitions of the K-Means indeterminacy coordinate against model-based uncertainty signals (random forest, out-of-fold).

I definition (mean over attributes)	AUC vs misclassification	AUC vs classifier disagreement	ρ with bootstrap instability	ρ with boundary proximity	mean	SD
I_{oria}	0.56	0.57	0.10	0.18	0.35	0.11
I_{amb}	0.55	0.56	0.06	-0.03	0.39	0.11
I_{ent}	0.53	0.50	0.06	0.07	0.41	0.09

4.5 Stability and sensitivity

Seeds. With the random-point initialisation of the original implementation, 50 seeds produced on average 5.5 distinct centroid solutions per attribute, a middle-centroid standard deviation of 0.050 and a mean absolute deviation of I of 0.098; with k-means++ and ten restarts the corresponding values were 2.3, 0.002 and 0.005. The algorithm is stochastic; reproducibility requires a fixed seed, and stability requires restarts. The downstream AUC of the K-Means [T, I, F] representation over ten seeds was 0.803 ± 0.001. Number of clusters. The silhouette did not prefer $K = 3$ for any attribute: it selected $K = 2$ for age, $K = 4$ for ca, $K = 2$ for chol, $K = 5$ for oldpeak, $K = 2$ for thalach, $K = 5$ for trestbps; the method is defined only for $K = 3$, so K is an assumption imposed on the data, not a data-driven choice. Slope. Table 6 shows that the slope governs everything that made K-Means look distinctive: with $s = 2$ the share of saturated observations is 14 %, the complementarity index 0.31 and the resolution-free R^2 of I from (T, F) 1.00; with $s = 20$ they are 63 %, 0.29 and 0.80. Scaling. For

cholesterol, the most skewed attribute, min-max, winsorized and rank scaling moved the middle centroid from 0.29 to 0.40 and 0.49 and the mean I from 0.40 to 0.39 and 0.35. Discrete attribute. The vessel count takes four values; three clusters map them to T = (0.00, 0.00, 0.00, 1.00), I = (0.00, 1.00, 0.00, 0.00), F = (0.50, 0.00, 0.00, 0.00): a look-up table with a degenerate middle state, not a graded membership. It is excluded from the sensitivity summaries and flagged in every table.

Table 6: Sensitivity of the K-Means transformation to the sigmoid slope (logistic regression, 5 × 3 CV).

slope s	mean T	mean I	mean F	mean sum	compl. index	% saturated	R^2 I from (T,F)	AUC [T,F]	AUC [T,I,F]
2	0.11	0.35	0.20	0.66	0.31	14	1.00	0.828	0.826
5	0.09	0.35	0.20	0.64	0.29	47	0.97	0.818	0.811
10	0.09	0.35	0.20	0.64	0.29	58	0.96	0.813	0.803
20	0.09	0.35	0.20	0.64	0.29	63	0.80	0.809	0.800

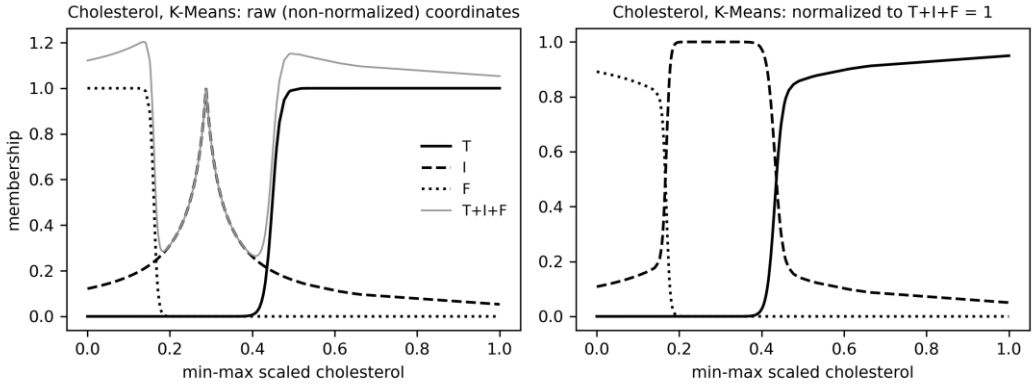

Figure 4: K-Means coordinates for cholesterol as functions of the scaled value: raw (left, with the magnitude $T + I + F$) and normalized to unit sum (right).

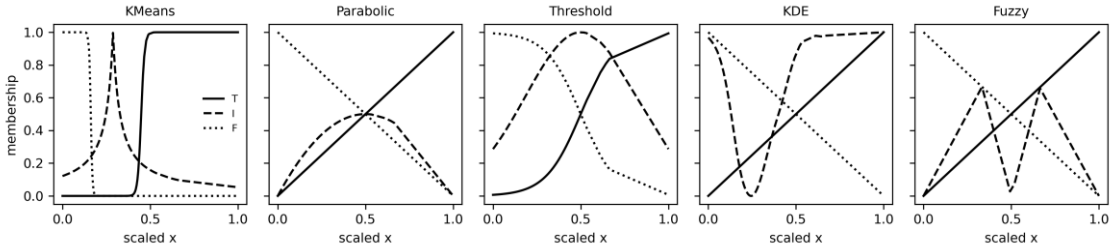

Figure 5: T, I and F as functions of the scaled cholesterol value for the five methods; every coordinate is a function of the single scalar x.

5 Discussion

5.1 Non-normalization is not independence

Non-normalization is not independence. The preprint's argument ran from $T + I + F \neq 1$ and $F \neq 1 - T$ to "true independence". The audit shows why the inference fails twice: four methods are non-normalized and yet have an I that is an exact function of T, and the fifth is non-complementary and yet has an I that is an exact function of (T, F) in full arithmetic. The magnitude $T + I + F$ is not a quality score in either direction, and a complementarity index of 0.29 says only that F is not 1 − T.

5.2 What the audit says about current neutrosophication

What K-Means actually adds. The K-Means coordinate is non-redundant at any practical resolution because the two logistic coordinates discard the middle of the range. The third coordinate therefore records where x lies among three clusters after T and F have stopped recording it. That is a reparameterization of one scalar into three pieces, one of which is active where the others are flat; it is data-adaptive, and it is not a second source of information. The theoretical companion makes the same point abstractly: a coordinate that varies while the others are

fixed is representationally independent, but that independence is a property of the state space, not of any inference or task [10]. Here the task settles the matter: with a linear classifier the extra coordinate did not help, with tree ensembles it made no reliable difference, and no neutrosophic representation exceeded the raw attributes by more than 0.006 AUC, because every transformation at best reparameterizes x and at worst discards part of it.

Descriptors are not quality. High entropy and range of I are consequences of the transformation's shape and of parameters (λ, the triangular supports, the slope). The slope sensitivity makes this explicit: the same data produce a "very neutrosophic" or a "nearly fuzzy" K-Means depending on s.

5.3 Semantic validity of indeterminacy

Semantics of I. Membership in the middle cluster is a statement about position, not about uncertainty. Without a clinical proposition ("cholesterol is high") and an external criterion, the coordinate should be described as a data-driven three-state membership, which is what the ablation and the external validation treat it as. The weak association with misclassification, disagreement and instability confirms that attribute-level position is not model-level uncertainty; the ambiguity-based alternatives are more defensible as definitions but are not, on these data, strong uncertainty signals either. Under Belnap–Dunn and annotated readings, a single I would in any case conflate ignorance and conflict [20, 21], and subjective logic and Dempster–Shafer theory derive their third mass rather than measure it [22, 23]; intuitionistic and picture fuzzy sets bound the third coordinate by the first two [24, 25]. What is distinctive about the non-normalized neutrosophic triple is representational freedom, and the present results show that freedom must be earned empirically.

5.4 From scalar reparameterization to evidence-enriched neutrosophication

The failure of a derived I to supply an independent source of information does not show that neutrosophic indeterminacy cannot be operationalised. It shows that an indeterminacy coordinate must be grounded in an evidential phenomenon that the polar pair does not already encode. This section states the design consequence and reviews, with their prior art, the families of neutrosophication in which such grounding is possible. Nothing in it is a validated method; it is a research agenda, and each item is classified as existing work, ongoing work of the authors, or a proposed direction. Neutrosophication itself has so far been defined only informally, as the conversion of a crisp or fuzzy quantity into a triple [71, 6], and applied through ad hoc procedures that treat the forward step as a solved sub-task [42, 72, 73, 7]; no statement of what such a procedure must satisfy was found, which is the gap that Section 5.5 addresses.

Two types. In scalar-derived neutrosophication (Type A) a single observation x is mapped to (T(x), I(x), F(x)); all five methods audited here are of this type, as are most fuzzy, threshold, density and clustering recipes. In evidence-enriched neutrosophication (Type B) the coordinates are computed from evidence channels that are separately grounded, $\Phi(e_T, e_I, e_F; z) = (T(e_T, z_T), I(e_I, z_I), F(e_F, z_F))$, where the channels may overlap or be statistically dependent but have distinguishable semantics and provenance. The central hypothesis of the agenda is that a neutrosophic coordinate should be justified by the evidential phenomenon it represents, not by the shape of the function that produces it. The word "independent" in "independent evidence channels" therefore means separately grounded, never $P(E_T, E_I, E_F) = P(E_T)P(E_I)P(E_F)$; statistical dependence between channels is expected and permitted, and the requirement is non-recoverability of I from (T, F) together with a distinct meaning.

Conflict is not indeterminacy. Strong evidence for a proposition and strong evidence against it give high T and high F; that state is conflict, and in the free triple it is represented in the polar plane, not in I [10, Section 7]. Low T and low F with little evidence is insufficiency. High I should be reserved for an explicitly justified source of indeterminacy: abstention, ambiguity of the evidence, absence of a modality, unreliability of an instrument. Whether a future design needs a separate conflict coordinate C, or whether C should remain a derived quantity C = c(T, F) as in annotated logic [21] and in the companion's contradiction measures [10], is an open design question; this paper does not propose a four-component representation, and it warns against collapsing conflict into I, which several aggregation recipes do implicitly.

Reliability-aware neutrosophication. The simplest Type B design keeps T = T(x) and F = F(x) and sets I = g(q), where q is the reliability of the observation: instrument precision, test–retest variability, assay reliability, the width of a confidence interval, missingness or source credibility. Two observations with the same x have the same (T, F) and different I when their q differs, which is a fibre of the polar map carrying two values of I: the polar-fibre criterion is satisfied by construction and by provenance, not by saturation. The nearest existing construct, the neutrosophic Z-number, attaches a reliability degree to each of the three coordinates rather than deriving I from reliability [29], neutrosophic statistics represents measurement imprecision as intervals [30], and the (value, reliability) pair itself originates outside neutrosophy [31].

Multi-source neutrosophication. With m sources each supplying support s_i, opposition r_i, unresolved evidence u_i and reliability w_i, one may aggregate $T = 1 - \Pi_j(1 - w_i\ s_i)$, $F = 1 - \Pi_j(1 - w_i\ r_i)$, $I = 1 -$

Π_j(1 − w_i u_i). These operators are illustrations, not recommendations: they are the probabilistic-sum family and inherit its assumptions. Multi-source and multi-sensor neutrosophic fusion, including combinations with Dempster–Shafer evidence theory and conflict redistribution, already exist [32, 33, 34, 35, 36]. What the present framework adds is not the idea of fusing sources into a triple but two evaluation criteria that the fusion literature does not apply: a non-redundancy audit of the aggregated I against the aggregated (T, F), and source-separated provenance, so that an I produced by disagreement between sources is reported as conflict in the polar plane and not as indeterminacy.

Ensemble- and stability-based neutrosophication. Here T and F are substantive evidence for and against the predicted proposition and I is the instability of that evidence across independently trained models, bootstrap refits, architectures or data perturbations: variance of a deep ensemble, disagreement rate, conformal ambiguity [37, 38, 39], and outside neutrosophy deep ensembles, aleatoric–epistemic decompositions and conformal prediction supply the disagreement statistics [40, 26, 41]. The warning of Section 4 applies directly: if T = p and $F = 1 -$ p come from the same ensemble probability and I = 4p(1 − p), the triple is Type A in disguise and I is a function of (T, F). An evidence-enriched ensemble computes I from resampling that (T, F) do not see, for instance I(x) = $P_{bootstrap}$(decision changes) or a normalized predictive variance; the resulting I may be statistically dependent on (T, F), which is allowed, but it must not be recoverable from them, which is testable. Neutrosophic ensemble classifiers exist [37, 38, 39], and outside neutrosophy deep ensembles, aleatoric–epistemic decompositions and conformal prediction supply the disagreement statistics [40, 26, 41]; the earliest define I from the disagreement between the truth and falsity outputs of paired networks [37], a published configuration to which the polar-fibre criterion applies directly, and none that we found audits the recoverability of I. The external validation of Section 4.4 is the test such a design would have to pass, and the K-Means I fails it.

Temporal neutrosophication. For time-dependent evidence, T_t and F_t summarise current support and opposition and I_t records temporal instability: change-point uncertainty, disagreement between recent and historical evidence, forecast instability, concept drift. Neutrosophic time-series and forecasting models exist [42, 43, 72]; one of them already ties I to the information entropy of high-order fluctuations, a statistic distinct from the trend statistics that give T and F [43], which makes it the closest published instance of a separately grounded I. Whether I_t is grounded in drift detection [44] is open; the design question, again, is whether I_t is grounded in a temporal phenomenon or derived from the current polar pair.

Multi-sensor and multimodal neutrosophication. Each sensor or modality should keep its signal direction, reliability and missingness before fusion; T and F aggregate reliable evidence for and against a state, I aggregates measurement ambiguity, degradation and absent modalities, and disagreement between modalities is reported as conflict. Sensor fusion with neutrosophic sets is established [32, 33, 34, 35, 36], as is neutrosophic image–text fusion [45]; evidential multi-view classification keeps a per-view uncertainty mass before combination [46], which is what this design asks of modalities. The contribution proposed here is source-separated non-redundancy auditing, not the fusion itself.

Expert and crowd neutrosophication. In Delphi panels and clinical committees each expert supplies support, opposition, abstention and a reliability weight. Neutrosophic Delphi and group decision methods exist [47, 48, 49]. The design rule that follows from this paper is that disagreement between confident experts is conflict (high T, high F), whereas I should be reserved for explicit "unknown", abstention, lack of expertise and insufficient information.

Retrieval-augmented generation. For a claim q and retrieved documents d_i, T_i, F_i and I_i record whether d_i supports, refutes or does not settle q, with source reliability kept separately; aggregation across documents should preserve support, opposition, indeterminacy, evidence magnitude and source diversity. This is an ongoing, unpublished application of independent evidence coordinates by the present authors, with a pre-registered experiment and public code [50], and it is mentioned here only as an instance of the design principle; the present clinical paper contributes nothing to it. A published neutrosophic assessment of fact-checking outcomes derives a triple from supporting, refuting and unresolved evidence in the same spirit, without source reliability or a redundancy audit [51]. The literature on knowledge conflicts in retrieval [52, 53] is the relevant external prior art.

LLM-elicited neutrosophication. Unrestricted elicitation of (T, I, F) from language models exists, including work by the present authors on the behaviour of elicited triples [54], and refined neutrosophic stance detection [55]. An elicited I is a stated coordinate, not a calibrated probability and not a latent internal state, and the elicitation pipeline must itself be audited with the questions of this paper: can the elicited I be predicted from the elicited (T, F); does unconstrained prompting add expressed information; is I stable across prompts and models; does it track an external uncertainty signal; does it change decisions. A recent independent study derives the triple from calibrated class probabilities and routes high-I cases to a human [39]; the authors' own unpublished predictive-validity experiments find that verbally elicited triples behave as a scalar with three names, a negative result that must be reported when published.

Multi-agent evidence aggregation. When agents receive distinct evidence, one supporting, one opposing, one inconclusive, a majority vote destroys the evidential structure. A pipeline that keeps per-agent coordinates, aggregates them with source awareness and maps the resulting state to one of five behaviours (answer, refute, report conflict, abstain, request more evidence) preserves it. Multi-agent debate and evidence aggregation with language models are active areas [56, 57, 58]; the only neutrosophic multi-agent system found combines JADE agents with neutrosophic AHP and has no language models [59]; we found no work that audits the aggregated indeterminacy for recoverability, and we do not claim priority for the pipeline.

Active evidence acquisition. The most useful operational role for I is to decide when to acquire more information. Insufficiency (low T, low F, low magnitude) calls for another measurement, document or source; conflict (high T, high F) calls for an independent adjudicating source; a robust state stops acquisition. This is the value-of-information logic of decision analysis and active learning [60, 61, 62] applied to the evidence state rather than to a scalar; it gives I an operational meaning that can be validated by whether acquisition decisions improve outcomes.

Probabilistic and credal semantics. If T = P(E_T), I = P(E_I) and F = P(E_F) are marginal probabilities of overlapping evidence events, the sum need not be one and the triple does not determine a unique joint distribution on the eight evidence states of $\{0,1\}^3$; the marginals induce a set of compatible joint distributions, which may be read as a credal set. This is a motivated direction, not a result: it must be compared with neutrosophic probability, probabilistic Belnap–Dunn logic, imprecise probability, Dempster–Shafer theory and subjective logic [63, 64, 65, 66, 22, 23, 67] before any claim is made, and it belongs to a separate theoretical paper.

Provenance-aware representation. A triple without provenance hides where its coordinates came from. Retaining (T, I, F) together with a provenance record (source, timestamp, reliability, modality, model, retrieval rank, instrument) keeps the triple as the evidence summary while making conflict investigation and audit possible, as in provenance standards for data and workflows [68, 69, 70]. Provenance should not be forced into a fourth membership coordinate.

Table 7 summarises the families, their inputs, the meaning of each coordinate, whether they add evidence beyond x, their redundancy risk and their status.

Table 7: Taxonomy of neutrosophication families. Status: AUDITED HERE, EXISTS IN LITERATURE, ONGOING AUTHOR WORK, PROPOSED DIRECTION (prior-art audit of September 2026; classes A = established, B = exists without non-redundancy audit, C = analogue in another uncertainty framework, D = open).

Family	Inputs	T	F	I	Evidence beyond x?	Redundancy risk	Status / prior art	Example
1 Scalar-derived (parabolic, threshold, KDE, fuzzy)	one scaled value x	position of x	1 − T	shape function of x	no	certain (I = g(T))	AUDITED HERE; A [3, 6, 15]	attribute membership
2 Clustering-based (K-Means + sigmoid, NCM)	x + fitted centroids	high-cluster logistic	low-cluster logistic	middle-cluster proximity	no	high; non-redundant only via saturation	AUDITED HERE; A [7, 8]	attribute membership
3 Reliability-aware	x + reliability q	position of x	position of x	g(q): instrument precision, CI width, credibility	yes (q)	low if q ⊥ x	PROPOSED; B [29–31]	assay with known precision
4 Multi-source	m sources (s_i, r_i, u_i, w_i)	aggregated support	aggregated opposition	aggregated unresolved evidence	yes (m sources)	medium: I may track disagreement (conflict)	EXISTS (fusion) B/C [32–36]; audit PROPOSED	multi-sensor recognition
5 Ensemble / stability	evidence + resamples/models	evidence for proposition	evidence against	instability across refits	yes (resampling)	high if I = h(p) [37]	EXISTS B [37–39]; C [40, 26, 41]; audit PROPOSED	bootstrap decision flip rate
6 Temporal	series up to t	current support	current opposition	drift, change-point, fluctuation entropy	yes (history)	medium	EXISTS B [42, 43, 72]; drift-grounded I D [44]	monitoring, finance
7 Multi-sensor	sensor readings + quality	reliable evidence for A	reliable evidence for ¬A	degradation, ambiguity	yes	medium	EXISTS B/C [32–34]; source-separated audit PROPOSED	activity recognition

8 Multimodal	modalities + missingness	modalities supporting	modalities opposing	absent / low-quality modality	yes	medium	EXISTS B [45]; C [46]; audit PROPOSED	medical multimodal AI
9 Expert / crowd	experts (support, oppose, abstain, w_i)	weighted support	weighted opposition	abstention, declared unknown	yes	high if disagreement → I	EXISTS A/B [47–49]; conflict rule PROPOSED	Delphi panels
10 RAG evidence	retrieved documents + reliability	documents supporting q	documents refuting q	documents that do not settle q	yes	medium	ONGOING AUTHOR WORK [50]; related [51]; C [52, 53]	claim verification
11 LLM-elicited	prompt + model	elicited support	elicited opposition	elicited indeterminacy	elicited, not measured	high (stated triple may be a scalar)	EXISTS B [54, 55, 39]; audit PROPOSED	epistemic elicitation
12 Multi-agent LLM	per-agent evidence	agents' support	agents' opposition	agents' inconclusive evidence	yes	medium	C [56–58]; neutrosophic version D [59]; PROPOSED	answer / refute / conflict / abstain / acquire
13 Probabilistic / credal	joint evidence events	P(E_T)	P(E_F)	P(E_I)	n/a (semantics)	n/a	C [63–67, 22, 23]; credal reading D; FUTURE PAPER	theory
14 Provenance-aware	triple + provenance record	as in 3–12	as in 3–12	as in 3–12	yes (metadata)	n/a	C [68–70]; neutrosophic version D; PROPOSED	regulated AI, evidence synthesis

5.5 Design principles and an indeterminacy validation checklist

We name the general form Evidence-Grounded Neutrosophication (EGN): a neutrosophication is evidence-grounded when each coordinate is computed from an evidence channel with a stated meaning and a stated provenance, and when the indeterminacy coordinate passes the audit of this paper. Ten design principles follow from the results. P1, proposition grounding: define the proposition A before defining T, I and F. P2, semantic separation: T, F and I must have explicitly stated evidential meanings. P3, provenance: identify which observations or sources generate each coordinate. P4, non-redundancy audit: test whether I is recoverable from (T, F). P5, resolution awareness: run the audit at the numerical precision at which the data will be stored. P6, conflict separation: do not collapse simultaneous support and opposition into I. P7, external validity: validate I against the phenomenon it claims to represent. P8, downstream utility: test whether I affects a task. P9, stability: test seeds, parameters, scaling and source perturbations. P10, no forced normalization: normalize only when the application requires composition rather than magnitude [10]. The principles apply to Type A and Type B alike; the audited methods fail P1–P3 and P7–P8 and satisfy P10 vacuously.

The Indeterminacy Validation Checklist asks of any proposed I: (1) an explicit semantic interpretation; (2) provenance or information not trivially contained in (T, F); (3) empirical non-redundancy; (4) stability; (5) external validation; (6) downstream relevance; (7) an appropriate response to missing or ambiguous evidence; (8) distinction from support–opposition conflict; (9) robustness to numerical resolution; (10) reproducibility. A researcher with a new transformation can run items (3), (4), (5), (6) and (9) with the scripts released with this paper; items (1), (2), (7) and (8) are answered by the design; item (10) by the release.

5.6 Limitations

One data set with 299 records and six attributes, one downstream task and three classifier families. The transformations are univariate; multivariate neutrosophication was not examined, and no evidence-enriched design was implemented or tested here, so Section 5.4 is a research agenda supported by the audit, not by new experiments. The K-Means implementation audited is the published one; other K-means-based definitions of I exist and only two alternatives were tested. Out-of-sample R^2 depends on the learners tried and the resolution test on the chosen rounding; the nearest-neighbour tolerances were pre-declared but are conventions. Percentile ranges over correlated folds are descriptive. The polar-fibre criterion and its corollary are elementary restatements of the companion's results, not new theory. The clinical propositions that would give T, I and F a medical meaning were not defined, so no medical interpretation is offered. The discrete vessel count is a limitation of the data for

all five methods. The prior-art audit behind Section 5.4 was conducted in September 2026 with bibliographic databases and cannot exclude unindexed work.

6 Conclusions

Not every three-component representation contains three independent dimensions of information. On the Cleveland heart-disease attributes, four of five neutrosophication methods derive their indeterminacy analytically from T and are fuzzy pairs with an added, redundant coordinate; the proposed K-Means method produces a coordinate that is non-redundant at practical resolution only because its truth and falsity sigmoids saturate, and that coordinate adds little to a downstream classifier and tracks external uncertainty weakly. Non-normalization is not independence, non-complementarity is not independence, and a univariate transformation cannot add information about its input. The results do not make K-Means an ideal method, nor a bad one; they make it a reparameterization whose extra coordinate must be justified by a task.

The constructive consequence is a change of direction. The next methodological step is not to search for increasingly complex functions of a single scalar x, but to design evidence-grounded neutrosophication procedures in which support, opposition and indeterminacy are tied to explicit, separately auditable sources of information, and to evaluate every such procedure by the information and semantics of its coordinates rather than by their algebraic non-normalization. The polar-fibre criterion, the resolution-aware redundancy test, the same-classifier ablation and the validation checklist released here are the tools for that evaluation; the families reviewed in Section 5.4 are where it should be applied next.

Declarations

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest. The authors are the developers of the neutrolab library audited here.

Data and Code Availability: UCI Cleveland heart-disease data (processed.cleveland.data). Scripts analysis/01_neutrosophication_audit.py and analysis/02_resolution_redundancy.py, results tables (results/*.csv, METHOD_AUDIT.md) and figures accompany the manuscript; neutrolab 1.4.0 is on PyPI.

Use of AI: Claude (Anthropic) assisted with the design and coding of the re-analysis, the equation audit and the drafting under the authors' supervision; all computations were executed by the accompanying scripts and all references were verified against Crossref/DOI records. The authors are responsible for interpretations and the final text.

References

[1] Smarandache, F. A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability; American Research Press: Rehoboth, 1999.

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

[3] Smarandache, F. Introduction to Neutrosophic Statistics; Sitech & Education Publishing: Craiova, 2014. https://doi.org/10.48550/arXiv.1406.2000

[4] Guo, Y.; Cheng, H.D. New neutrosophic approach to image segmentation. Pattern Recognition 2009, 42(5), 587–595. https://doi.org/10.1016/j.patcog.2008.10.002

[5] Guo, Y.; Cheng, H.D.; Zhang, Y. A new neutrosophic approach to image denoising. New Mathematics and Natural Computation 2009, 5(3), 653–662. https://doi.org/10.1142/S1793005709001490

[6] Zhang, M.; Zhang, L.; Cheng, H.D. A neutrosophic approach to image segmentation based on watershed method. Signal Processing 2010, 90(5), 1510–1517. https://doi.org/10.1016/j.sigpro.2009.10.021

[7] Qureshi, M.N.; Ahamad, M.V. An improved method for image segmentation using K-means clustering with neutrosophic logic. Procedia Computer Science 2018, 132, 534–540. https://doi.org/10.1016/j.procs.2018.05.006

[8] Guo, Y.; Sengur, A. NCM: Neutrosophic c-means clustering algorithm. Pattern Recognition 2015, 48(8), 2710–2724. https://doi.org/10.1016/j.patcog.2015.02.018

[9] Leyva Vázquez, M.Y.; Cevallos-Torres, L.; Mar Cornelio, O.; Smarandache, F. A comparative analysis of data-driven and model-based neutrosophication methods: advancing true neutrosophic logic in medical data transformation. engrXiv preprint, 2025. https://doi.org/10.31224/5968

[10] Leyva Vázquez, M.Y.; Smarandache, F. Independent evidence coordinates: representation, normalization, and inference in paraconsistent logic. Manuscript, 2026 (companion paper).

[11] Detrano, R.; Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Schmid, J.-J.; Sandhu, S.; Guppy, K.H.; Lee, S.; Froelicher, V. International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology 1989, 64(5), 304–310. https://doi.org/10.1016/0002-9149(89)90524-9

[12] Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O. Leakage in data mining: formulation, detection, and avoidance. ACM Transactions on Knowledge Discovery from Data 2012, 6(4), 15. https://doi.org/10.1145/2382577.2382579
[13] Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006, 7, 91. https://doi.org/10.1186/1471-2105-7-91
[14] Leyva-Vázquez, M.Y.; Smarandache, F. NeutroLab: a unified library for neutrosophic learning, configuration analysis, and logical machines, version 1.4.0. PyPI, 2025. https://pypi.org/project/neutrolab/
[15] Zadeh, L.A. Fuzzy sets. Information and Control 1965, 8(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
[16] Nadeau, C.; Bengio, Y. Inference for the generalization error. Machine Learning 2003, 52(3), 239–281. https://doi.org/10.1023/A:1024068626366
[17] Bouckaert, R.R.; Frank, E. Evaluating the replicability of significance tests for comparing learning algorithms. In Advances in Knowledge Discovery and Data Mining (PAKDD 2004), LNCS 3056; Springer, 2004; pp. 3–12. https://doi.org/10.1007/978-3-540-24775-3_3
[18] Arthur, D.; Vassilvitskii, S. k-means++: the advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07); SIAM, 2007; pp. 1027–1035.
[19] Rousseeuw, P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 1987, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
[20] Belnap, N.D. A useful four-valued logic. In Modern Uses of Multiple-Valued Logic; Dunn, J.M., Epstein, G., Eds.; Reidel: Dordrecht, 1977; pp. 5–37. https://doi.org/10.1007/978-94-010-1161-7_2
[21] da Costa, N.C.A.; Subrahmanian, V.S.; Vago, C. The paraconsistent logics Pτ. Zeitschrift für mathematische Logik und Grundlagen der Mathematik 1991, 37, 139–148. https://doi.org/10.1002/malq.19910370903
[22] Jøsang, A. Subjective Logic: A Formalism for Reasoning Under Uncertainty; Springer, 2016. https://doi.org/10.1007/978-3-319-42337-1
[23] Shafer, G. A Mathematical Theory of Evidence; Princeton University Press, 1976. https://doi.org/10.1515/9780691214696
[24] Atanassov, K.T. Intuitionistic fuzzy sets. Fuzzy Sets and Systems 1986, 20(1), 87–96. https://doi.org/10.1016/S0165-0114(86)80034-3
[25] Cuong, B.C. Picture fuzzy sets. Journal of Computer Science and Cybernetics 2014, 30(4), 409–420. https://doi.org/10.15625/1813-9663/30/4/5032
[26] Hüllermeier, E.; Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine Learning 2021, 110(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3
[27] Smarandache, F. Degree of dependence and independence of the (sub)components of fuzzy set and neutrosophic set. Neutrosophic Sets and Systems 2016, 11, 95–97. https://doi.org/10.5281/zenodo.571359
[28] Yao, Y. Three-way decisions with probabilistic rough sets. Information Sciences 2010, 180(3), 341–353. https://doi.org/10.1016/j.ins.2009.09.021
[29] Du, S.; Ye, J.; Yong, R.; Zhang, F. Some aggregation operators of neutrosophic Z-numbers and their multicriteria decision making method. Complex & Intelligent Systems 2021, 7(1), 429–438. https://doi.org/10.1007/s40747-020-00204-w
[30] Aslam, M. Neutrosophic analysis of variance: application to university students. Complex & Intelligent Systems 2019, 5, 403–407. https://doi.org/10.1007/s40747-019-0107-2
[31] Zadeh, L.A. A note on Z-numbers. Information Sciences 2011, 181(14), 2923–2932. https://doi.org/10.1016/j.ins.2011.02.022
[32] Gong, Y.; Ma, Z.; Wang, M.; Deng, X.; Jiang, W. A new multi-sensor fusion target recognition method based on complementarity analysis and neutrosophic set. Symmetry 2020, 12(9), 1435. https://doi.org/10.3390/sym12091435
[33] Jiang, W.; Zhong, Y.; Deng, X. A neutrosophic set based fault diagnosis method based on multi-stage fault template data. Symmetry 2018, 10(8), 346. https://doi.org/10.3390/sym10080346
[34] Dong, Y.; Li, X.; Dezert, J.; Khyam, M.O.; Noor-A-Rahim, M.; Ge, S.S. Dezert–Smarandache theory-based fusion for human activity recognition in body sensor networks. IEEE Transactions on Industrial Informatics 2020, 16(11), 7138–7149. https://doi.org/10.1109/TII.2020.2976812
[35] Dezert, J.; Smarandache, F. DSmT: a new paradigm shift for information fusion. arXiv 2006, cs/0610175. https://doi.org/10.48550/arXiv.cs/0610175
[36] Deng, Y. Generalized evidence theory. Applied Intelligence 2015, 43, 530–543. https://doi.org/10.1007/s10489-015-0661-2
[37] Kraipeerapun, P.; Fung, C.C. Binary classification using ensemble neural networks and interval neutrosophic sets. Neurocomputing 2009, 72(13–15), 2845–2856. https://doi.org/10.1016/j.neucom.2008.07.017
[38] Kavitha, B.; Karthikeyan, S.; Maybell, P.S. An ensemble design of intrusion detection system for handling uncertainty using neutrosophic logic classifier. Knowledge-Based Systems 2012, 28, 88–96. https://doi.org/10.1016/j.knosys.2011.12.004
[39] Abdel-Aziz, N.M.; Ibrahim, M.; Eldrandaly, K.A. Responsible AI for text classification: a neutrosophic approach combining classical models and BERT. Neutrosophic Sets and Systems 2025, 94, 49–59. https://doi.org/10.5281/zenodo.17041894

[40] Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems 30 (NIPS 2017); Curran Associates, 2017. https://papers.nips.cc/paper_files/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html
[41] Angelopoulos, A.N.; Bates, S. Conformal prediction: a gentle introduction. Foundations and Trends in Machine Learning 2023, 16(4), 494–591. https://doi.org/10.1561/2200000101
[42] Abdel-Basset, M.; Chang, V.; Mohamed, M.; Smarandache, F. A refined approach for forecasting based on neutrosophic time series. Symmetry 2019, 11(4), 457. https://doi.org/10.3390/sym11040457
[43] Guan, H.; Dai, Z.; Guan, S.; Zhao, A. A neutrosophic forecasting model for time series based on first-order state and information entropy of high-order fluctuation. Entropy 2019, 21(5), 455. https://doi.org/10.3390/e21050455
[44] Gama, J.; Žliobaitė, I.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. A survey on concept drift adaptation. ACM Computing Surveys 2014, 46(4), 1–37. https://doi.org/10.1145/2523813
[45] Wajid, M.A.; Zafar, A.; Terashima-Marín, H.; Wajid, M.S. Neutrosophic-CNN-based image and text fusion for multimodal classification. Journal of Intelligent & Fuzzy Systems 2023, 45(1), 1039–1055. https://doi.org/10.3233/JIFS-223752
[46] Han, Z.; Zhang, C.; Fu, H.; Zhou, J.T. Trusted multi-view classification. In International Conference on Learning Representations (ICLR 2021), 2021. https://openreview.net/pdf?id=OOsR8BzCnl5
[47] Smarandache, F.; Estupiñán Ricardo, J.; González Caballero, E.; Leyva Vázquez, M.Y.; Batista Hernández, N. Delphi method for evaluating scientific research proposals in a neutrosophic environment. Neutrosophic Sets and Systems 2020, 34, 204–213. https://doi.org/10.5281/zenodo.3820430
[48] Vafadarnikjoo, A.; Mishra, N.; Govindan, K.; Chalvatzis, K. Assessment of consumers' motivations to purchase a remanufactured product by applying fuzzy Delphi method and single valued neutrosophic sets. Journal of Cleaner Production 2018, 196, 230–244. https://doi.org/10.1016/j.jclepro.2018.06.037
[49] Abdel-Basset, M.; Mohamed, M.; Smarandache, F. An extension of neutrosophic AHP–SWOT analysis for strategic planning and decision-making. Symmetry 2018, 10(4), 116. https://doi.org/10.3390/sym10040116
[50] Leyva-Vázquez, M.Y.; Smarandache, F. Beyond scalar uncertainty: independent neutrosophic evidence coordinates for conflict-aware retrieval-augmented generation. Manuscript in preparation, 2026; pre-registered experiment and code at https://github.com/mleyvaz/evidence-coordinates-rag
[51] Voinea, D.V. Quantifying indeterminacy in news: a neutrosophic method for assessing fact-checking outcomes. Neutrosophic Sets and Systems 2025, 90. https://doi.org/10.5281/zenodo.16123010
[52] Xu, R.; Qi, Z.; Guo, Z.; Wang, C.; Wang, H.; Zhang, Y.; Xu, W. Knowledge conflicts for LLMs: a survey. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024); ACL, 2024; pp. 8541–8565. https://doi.org/10.18653/v1/2024.emnlp-main.486
[53] Chen, H.-T.; Zhang, M.; Choi, E. Rich knowledge sources bring complex knowledge conflicts: recalibrating models to reflect conflicting evidence. In Proceedings of EMNLP 2022; ACL, 2022; pp. 2292–2307. https://doi.org/10.18653/v1/2022.emnlp-main.146
[54] Leyva-Vázquez, M.Y.; Smarandache, F. Breaking the chains of probability: neutrosophic logic as a new framework for epistemic uncertainty in large language models. Neutrosophic Sets and Systems 2026, 99, article 19. https://doi.org/10.5281/zenodo.19954583
[55] Alejo Machado, O.J.; Estupiñán Sera, A.M.; Leyva Vázquez, M.Y.; Smarandache, F. Modeling ambiguity in AI-enhanced learning: a neutrosophic approach to stance detection and causal evaluation. International Journal of Neutrosophic Science 2025, 26(4), 298–308. https://doi.org/10.54216/IJNS.260426
[56] Du, Y.; Li, S.; Torralba, A.; Tenenbaum, J.B.; Mordatch, I. Improving factuality and reasoning in language models through multiagent debate. arXiv 2023, 2305.14325. https://doi.org/10.48550/arXiv.2305.14325
[57] Liang, T.; He, Z.; Jiao, W.; Wang, X.; Wang, Y.; Wang, R.; Yang, Y.; Shi, S.; Tu, Z. Encouraging divergent thinking in large language models through multi-agent debate. In Proceedings of EMNLP 2024; ACL, 2024. https://doi.org/10.18653/v1/2024.emnlp-main.992
[58] Baltaji, R.; Hemmatian, B.; Varshney, L.R. Conformity, confabulation, and impersonation: persona inconstancy in multi-agent LLM collaboration. In Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP (C3NLP, ACL 2024); ACL, 2024. https://doi.org/10.18653/v1/2024.c3nlp-1.2
[59] Meziani, A.; Bourouis, A.; Chebout, M.S. NeutroMAS4SCRM: a combined multi-agent system with neutrosophic data analytic hierarchy process framework for supply chain risk management. Journal of Intelligent & Fuzzy Systems 2022. https://doi.org/10.3233/JIFS-222305
[60] Howard, R.A. Information value theory. IEEE Transactions on Systems Science and Cybernetics 1966, 2(1), 22–26. https://doi.org/10.1109/TSSC.1966.300074
[61] Ren, P.; Xiao, Y.; Chang, X.; Huang, P.-Y.; Li, Z.; Gupta, B.B.; Chen, X.; Wang, X. A survey of deep active learning. ACM Computing Surveys 2022, 54(9), 1–40. https://doi.org/10.1145/3472291
[62] Jiang, Z.; Xu, F.F.; Gao, L.; Sun, Z.; Liu, Q.; Dwivedi-Yu, J.; Yang, Y.; Callan, J.; Neubig, G. Active retrieval augmented generation. In Proceedings of EMNLP 2023; ACL, 2023. https://doi.org/10.18653/v1/2023.emnlp-main.495
[63] Smarandache, F. Introduction to Neutrosophic Measure, Neutrosophic Integral, and Neutrosophic Probability; Sitech & Education Publisher: Craiova, 2013. arXiv 1311.7139.

[64] Klein, D.; Majer, O.; Rafiee Rad, S. Probabilities with gaps and gluts. Journal of Philosophical Logic 2021, 50(5), 1107–1141. https://doi.org/10.1007/s10992-021-09592-x

[65] Rodrigues, A.; Bueno-Soler, J.; Carnielli, W. Measuring evidence: a probabilistic approach to an extension of Belnap–Dunn logic. Synthese 2021, 198(S22), 5451–5480. https://doi.org/10.1007/s11229-020-02571-w

[66] Bílková, M.; Frittella, S.; Kozhemiachenko, D.; Majer, O.; Nazari, S. Reasoning with belief functions over Belnap–Dunn logic. Annals of Pure and Applied Logic 2024, 175(9), 103338. https://doi.org/10.1016/j.apal.2023.103338

[67] Augustin, T.; Coolen, F.P.A.; de Cooman, G.; Troffaes, M.C.M. (Eds.) Introduction to Imprecise Probabilities; Wiley: Chichester, 2014. https://doi.org/10.1002/9781118763117

[68] Moreau, L.; Missier, P. (Eds.) PROV-DM: The PROV Data Model. W3C Recommendation, 30 April 2013. https://www.w3.org/TR/prov-dm/

[69] Sikos, L.F.; Philp, D. Provenance-aware knowledge representation: a survey of data models and contextualized knowledge graphs. Data Science and Engineering 2020, 5, 293–316. https://doi.org/10.1007/s41019-020-00118-0

[70] Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Daumé III, H.; Crawford, K. Datasheets for datasets. Communications of the ACM 2021, 64(12), 86–92. https://doi.org/10.1145/3458723

[71] Smarandache, F. Introduction to Neutrosophic Sociology (Neutrosociology); Pons: Brussels, 2019. https://digitalrepository.unm.edu/math_fsp/28

[72] Tanuwijaya, B.; Selvachandran, G.; Son, L.H.; Abdel-Basset, M.; Huynh, H.X.; Pham, V.-H.; Ismail, M. A novel single valued neutrosophic hesitant fuzzy time series model: applications in Indonesian and Argentinian stock index forecasting. IEEE Access 2020, 8, 60126–60141. https://doi.org/10.1109/ACCESS.2020.2982825

[73] Chakraborty, A.; Mondal, S.P.; Ahmadian, A.; Senu, N.; Alam, S.; Salahshour, S. Different forms of triangular neutrosophic numbers, de-neutrosophication techniques, and their applications. Symmetry 2018, 10(8), 327. https://doi.org/10.3390/sym10080327