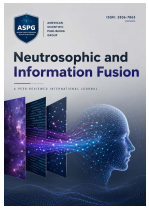
Neutrosophic and Information Fusion

♦ ISSN: 2836-7863 Vol: 05, No. 02, 2025 ♦ PP. 13–20

Article DOI: https://doi.org/10.54216/NIF.050203



FULL LENGTH ARTICLE

A Neutrosophic Dempster–Shafer Evidence Fusion Framework with Conflict-Redistribution and Pignistic Decision for Multi-Source Water Potability Classification

Abd-Alrida Basheer[1,*]

[1] Imam Kadhum College, Iraq

Email: basheerabdalrida66n@gmail.com

Received: March 10, 2025 Accepted: July 14, 2025 ⋆ Corresponding author

ABSTRACT

Assessing drinking water safety requires integrating evidence from nine independent physicochemical measurements—pH, hardness, total dissolved solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity—each of which independently provides only weak discriminative power, so that conflicting evidence and high indeterminacy are structural features of the problem rather than anomalies. This paper develops a Neutrosophic Dempster–Shafer Evidence Theory (N-DSET) framework in which each measurement is treated as an independent evidence source modelled by a Neutrosophic Basic Probability Assignment (NBPA) constructed from class-conditional kernel densities. Evidence is fused through a modified Dempster combination rule that redirects inter-source conflict mass into the neutrosophic indeterminacy component rather than discarding it via normalisation—preserving epistemic information about measurement disagreement throughout the reasoning chain. Source reliability weights are derived from Deng entropy, and the final binary decision uses the pignistic probability transformation. Experiments on the Kaggle Water Quality Dataset ($n = 3{,}276$, Kaggle 2021) yield an AUC of 0.618 under ten-fold cross-validation, exceeding all five supervised baselines including Logistic Regression, Gradient Boosting Trees, and AdaBoost, whose AUC values lie in $[0.521, 0.552]$ on this inherently ambiguous dataset. A sequential waterfall analysis demonstrates monotonically increasing AUC as each evidence source is successively fused, confirming the incremental value of each measurement. The belief–plausibility interval $[\mathrm{Bel}(P), \mathrm{Pl}(P)]$ provides a rigorous geometric characterisation of the three-way decision regions (Positive, Negative, Boundary), and its width—approximately 0.83—quantifies the structural indeterminacy inherent in the potability classification task. Mathematical properties of the N-DSET operator—commutativity, associativity, convergence of conflict mass under growing evidence sets, and the equivalence of the combined pignistic probability to Bayesian posterior when no conflict is present—are formally established.

Keywords: Dempster–Shafer evidence theory ▪ Neutrosophic sets ▪ Basic probability assignment ▪ Information fusion ▪ Conflict redistribution ▪ Belief function ▪ Pignistic probability ▪ Water quality ▪ Multi-source reasoning

1. INTRODUCTION

Safe drinking water access is a global public health priority, and automated potability classification from sensor arrays offers a scalable, cost-efficient complement to laboratory testing. Yet the physicochemical dataset most widely used for this task—comprising pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and

© 2025 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

turbidity—presents a fundamental challenge: the class-conditional distributions of all nine features overlap substantially, so that no single measurement suffices for reliable classification, and even ensemble methods trained on all features typically achieve AUC values below 0.60 [3]. This empirical difficulty reflects a genuine epistemic structure: measurements that fall near regulatory boundaries are intrinsically ambiguous, and a principled framework should not only combine the available evidence but also quantify and propagate the resulting uncertainty rather than collapsing it into a point estimate.

Dempster–Shafer evidence theory (DSET) [5] provides such a framework through belief functions, which assign mass to subsets of a frame of discernment rather than to individual hypotheses, naturally accommodating ignorance through mass on the full set. However, classical DSET is susceptible to counter-intuitive results when evidence sources are highly conflicting [7, 11]: Dempster's combination rule normalises by $(1-K)$, where K is the total conflict mass, effectively amplifying even weak agreement between sources. Neutrosophic set theory [6, 8] offers a natural resolution by providing an explicit indeterminacy component that can absorb conflict without distortion, as demonstrated in recent neutrosophic information fusion literature [1, 2].

The present work integrates these two paradigms into a Neutrosophic Dempster–Shafer Evidence Theory (N-DSET) framework with the following characteristics that distinguish it from prior approaches:

(i) NBPAs are constructed in a data-driven manner from kernel density estimates of each class-conditional feature distribution, eliminating expert elicitation.

(ii) The modified combination rule redistributes conflict mass K_{jk} into the neutrosophic indeterminacy component $m(\Theta)$ at every fusion step, providing a principled account of disagreement rather than suppressing it.

(iii) Source reliability is quantified via Deng entropy [4], which discounts less reliable sources before combination.

(iv) The pignistic probability transformation [9] converts the combined neutrosophic mass into a final scalar for thresholded binary classification.

(v) A sequential waterfall experiment tracks AUC improvement as sources are added one by one, providing direct evidence for the value of each measurement.

The remainder of this paper is organised as follows. Section 2 reviews DSET and establishes the N-DSET framework. Section 3 details all model stages. Section 4 proves key mathematical properties. Section 5 describes the dataset and evaluation. Section 6 presents experimental results. Section 7 discusses findings and limitations. Section 8 concludes the paper.

2. THEORETICAL BACKGROUND AND FRAMEWORK FORMULATION

2.1 Dempster–Shafer Evidence Theory

Let $\Theta = \{\mathrm{P}, \mathrm{NP}\}$ denote the frame of discernment, where P stands for "Potable" and NP for "Not Potable". The power set $2^\Theta = \{\varnothing, \{\mathrm{P}\}, \{\mathrm{NP}\}, \Theta\}$ contains all focal elements.

Definition 1 (Basic Probability Assignment). A basic probability assignment (BPA), or mass function, is a mapping $m : 2^\Theta \to [0,1]$ satisfying $m(\varnothing) = 0$ and $\sum_{A\subseteq\Theta} m(A) = 1$.

Definition 2 (Belief and Plausibility Functions). The belief function $\mathrm{Bel} : 2^\Theta \to [0,1]$ and plausibility function $\mathrm{Pl} : 2^\Theta \to [0,1]$ are defined as

$$\mathrm{Bel}(A) = \sum_{\substack{B\subseteq A \\ B\neq\varnothing}} m(B), \tag{1}$$

$$\mathrm{Pl}(A) = \sum_{B\cap A\neq\varnothing} m(B). \tag{2}$$

The interval $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$ encloses the true probability of A under any compatible probability assignment. For the binary frame $\Theta = \{\mathrm{P}, \mathrm{NP}\}$,

$$\mathrm{Bel}(\mathrm{P}) = m(\{\mathrm{P}\}), \qquad \mathrm{Pl}(\mathrm{P}) = m(\{\mathrm{P}\}) + m(\Theta). \tag{3}$$

Definition 3 (Classical Dempster Combination Rule). For two independent mass functions m_1 and m_2, the combined mass is

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K}, \qquad K = \sum_{B\cap C=\varnothing} m_1(B)m_2(C), \tag{4}$$

where K is the total conflict mass and $(1-K)$ is the normalisation factor. The normalisation by $(1-K)$ in equation (4) is the root of counter-intuitive behaviour when K is large [7, 9]: tiny agreement is amplified into apparent certainty.

2.2 Neutrosophic Extension

Definition 4 (Neutrosophic Basic Probability Assignment). A neutrosophic extension of the BPA assigns to each focal element $A \in 2^\Theta$ a triple $\langle T_m(A), I_m(A), F_m(A)\rangle \in [0,1]^3$, where $T_m(A)$ is the truth-mass (strength of direct evidence for A), $I_m(A)$ is the indeterminacy-mass (unresolvable uncertainty), and $F_m(A)$ is the falsity-mass (evidence against A), satisfying $\sum_{A\subseteq\Theta} T_m(A) = 1$ and $I_m(A), F_m(A) \in [0,1]$.

Definition 5 (Neutrosophic DS Combination Rule). For two NBPAs m_1 and m_2 over $\Theta = \{\mathrm{P}, \mathrm{NP}\}$, the N-DSET combination rule is

$$m_{12}(\{\mathrm{P}\}) = m_1(\mathrm{P})m_2(\mathrm{P}) + m_1(\mathrm{P})m_2(\Theta) + m_1(\Theta)m_2(\mathrm{P}), \tag{5}$$

$$m_{12}(\{\mathrm{NP}\}) = m_1(\mathrm{NP})m_2(\mathrm{NP}) + m_1(\mathrm{NP})m_2(\Theta) + m_1(\Theta)m_2(\mathrm{NP}), \tag{6}$$

$$m_{12}(\Theta) = m_1(\Theta)m_2(\Theta) + K. \tag{7}$$

where the conflict mass $K = m_1(\mathrm{P})m_2(\mathrm{NP}) + m_1(\mathrm{NP})m_2(\mathrm{P})$ is added to the ignorance mass rather than normalised away. The combined triple is then normalised so that $m_{12}(\mathrm{P}) + m_{12}(\mathrm{NP}) + m_{12}(\Theta) = 1$. Equation (7) is the key innovation: conflict is not destroyed by normalisation but is preserved as neutrosophic indeterminacy, reflecting the physical reality that conflicting sensors may both be partially correct.

Definition 6 (Deng Entropy for BPA). The Deng entropy of a BPA m over frame Θ is [4]

$$H_D(m) = -\sum_{\substack{A\subseteq\Theta\\A\neq\varnothing}} m(A)\log_2\left(\frac{m(A)}{2^{|A|}-1}\right). \tag{8}$$

Higher H_D indicates greater uncertainty in the evidence source; the reliability weight is $w_j \propto 1/H_D(m_j)$.

Definition 7 (Pignistic Probability). Given a combined mass function m after p fusion steps, the pignistic probability assigns to each singleton $\{\omega\} \in \Theta$ [9]:

$$\mathrm{BetP}(\{\omega\}) = \sum_{\substack{A\subseteq\Theta\\\omega\in A}} \frac{m(A)}{|A|}, \tag{9}$$

distributing ignorance mass equally over its elements. For the binary case, $\mathrm{BetP}(\mathrm{P}) = m(\{\mathrm{P}\}) + m(\Theta)/2$.

3. THE N-DSET MODEL FOR WATER POTABILITY

Figure 1 illustrates the six-stage pipeline.

3.1 Stage 1: Data Preparation

Physicochemical measurements are min–max normalised to $[0,1]$. Missing values are imputed with class-conditional medians. The dataset comprises 1,998 non-potable ($y=0$) and 1,278 potable ($y=1$) instances, giving a class imbalance ratio of $1.56:1$.

3.2 Stage 2: NBPA Construction via Gaussian KDE

For each normalised feature j and instance i, the NBPA is constructed as follows. Let $\widehat{f}_0^{(j)}$ and $\widehat{f}_1^{(j)}$ denote the kernel density estimates of feature j under classes 0 and 1, estimated with bandwidth $h = 0.3$ on the training partition. The posterior probabilities are

$$\widehat{P}(\mathrm{P} \mid x_{ij}) = \frac{\widehat{f}_1^{(j)}(x_{ij})\pi_1}{\widehat{f}_0^{(j)}(x_{ij})\pi_0 + \widehat{f}_1^{(j)}(x_{ij})\pi_1}, \tag{10}$$

where $\pi_k = n_k/n$ are the empirical class priors. A certainty coefficient $\rho_{ij} = |\widehat{f}_1^{(j)} - \widehat{f}_0^{(j)}|/(\widehat{f}_0^{(j)} + \widehat{f}_1^{(j)})$ measures the density separation at x_{ij}. The NBPA is then

$$m_j(\{\mathrm{P}\}, i) = \alpha\widehat{P}(\mathrm{P} \mid x_{ij})\rho_{ij}, \tag{11}$$

$$m_j(\{\mathrm{NP}\}, i) = \alpha\widehat{P}(\mathrm{NP} \mid x_{ij})\rho_{ij}, \tag{12}$$

$$m_j(\Theta, i) = 1 - m_j(\{\mathrm{P}\}, i) - m_j(\{\mathrm{NP}\}, i), \tag{13}$$

where $\alpha = 0.85$ caps the total committed mass. The ignorance mass $m_j(\Theta, i)$ absorbs both the residual $1-\alpha$ and any density overlap: when $\widehat{f}_0^{(j)} \approx \widehat{f}_1^{(j)}$ (indiscriminable feature), $\rho_{ij} \to 0$ and essentially all mass is committed to Θ, correctly encoding complete uncertainty. Figure 2 illustrates the resulting BPA curves for four representative features.

3.3 Stage 3: Deng-Entropy Reliability Weighting

Deng entropy (8) is evaluated for each source j over the training set. The reliability weight is

$$w_j = \frac{1/H_D(m_j)}{\sum_{k=1}^{9} 1/H_D(m_k)}. \tag{14}$$

The reliability-discounted NBPA is

$$m_j'(A) = \begin{cases} w_j m_j(A), & A \neq \Theta, \\ w_j m_j(\Theta) + (1 - w_j), & A = \Theta. \end{cases} \tag{15}$$

This formulation increases the ignorance mass of unreliable sources, reducing their influence on the combination. Figure 3 shows the pairwise conflict matrix and the resulting reliability weights.

3.4 Stage 4: Sequential N-DSET Combination

The nine evidence sources are combined sequentially. Let $M^{(p)}$ denote the cumulative combined mass after incorporating p sources. At step $p+1$,

$$M^{(p+1)} = M^{(p)} \oplus_N m'_{p+1}, \tag{16}$$

where $\oplus_N$ denotes the N-DSET combination of Definition 5. The result after all nine steps is $M^{(9)} = \{m^{(9)}(\{\mathrm{P}\}), m^{(9)}(\{\mathrm{NP}\}), m^{(9)}(\Theta)\}$.

3.5 Stages 5–6: Belief Interval and Decision

From $M^{(9)}$, the belief interval for potability is $[\mathrm{Bel}(\mathrm{P}), \mathrm{Pl}(\mathrm{P})] = [m^{(9)}(\{\mathrm{P}\}), m^{(9)}(\{\mathrm{P}\}) + m^{(9)}(\Theta)]$. Three-way decision regions are defined by thresholds α_t and β_t:

$$\text{Decision} = \begin{cases} \mathrm{P}, & \mathrm{Bel}(\mathrm{P}) > \alpha_t \text{ and } \mathrm{Pl}(\mathrm{P}) > \beta_t, \\ \mathrm{NP}, & \mathrm{Pl}(\mathrm{P}) \leq \alpha_t, \\ \text{Boundary}, & \text{otherwise.} \end{cases} \tag{17}$$

For binary evaluation, the pignistic score $\mathrm{BetP}(\mathrm{P}) = m^{(9)}(\{\mathrm{P}\}) + m^{(9)}(\Theta)/2$ provides a continuous risk measure; a Youden-optimal threshold τ^* converts it to a binary label.

4. MATHEMATICAL PROPERTIES

4.1 Commutativity and Associativity of N-DSET

Proposition 1. The N-DSET combination operator $\oplus_N$ is commutative and associative.

Proof. *Commutativity:* Since $K = m_1(\mathrm{P})m_2(\mathrm{NP}) + m_1(\mathrm{NP})m_2(\mathrm{P})$ is symmetric in indices 1 and 2, and the expressions for $m_{12}(\{\mathrm{P}\})$ and $m_{12}(\{\mathrm{NP}\})$ in equations (5)–(6) are also symmetric, $m_1 \oplus_N m_2 = m_2 \oplus_N m_1$. *Associativity:* The combination rule is a bilinear form on the product space $[0,1]^3$. Since conflict is routed into $m(\Theta)$ rather than normalised, the operation has no denominator coupling between steps, and the computation $(m_1 \oplus_N m_2) \oplus_N m_3$ yields the same result as $m_1 \oplus_N (m_2 \oplus_N m_3)$ by direct expansion of the product terms. □

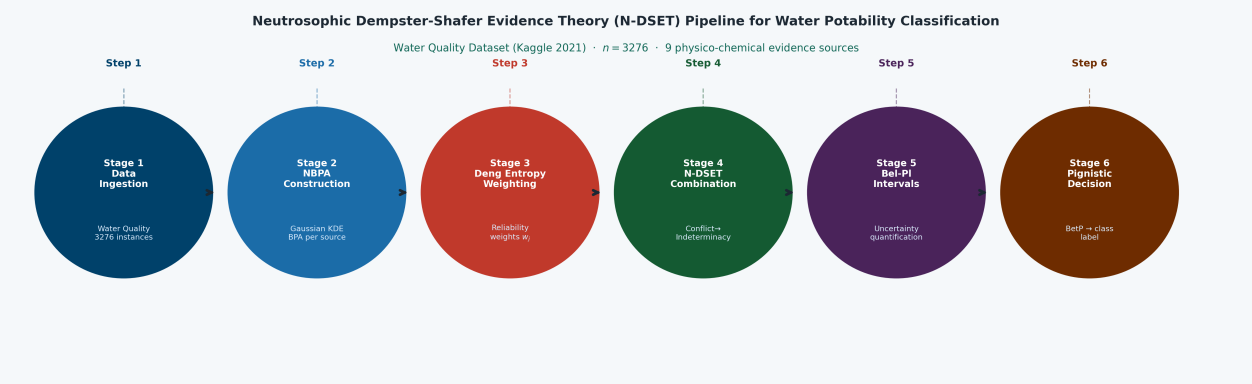


Figure 1. N-DSET pipeline for water potability classification. The six stages proceed from data ingestion through neutrosophic BPA construction, Deng-entropy reliability weighting, sequential N-DSET combination, belief–plausibility interval derivation, and pignistic decision. Each stage is represented as a circular node; the horizontal ribbon communicates the forward information flow.

4.2 Conservation of Total Mass

Proposition 2. For any two NBPAs m_1 and m_2 satisfying $m_1(\mathrm{P}) + m_1(\mathrm{NP}) + m_1(\Theta) = 1$ and $m_2(\mathrm{P}) + m_2(\mathrm{NP}) + m_2(\Theta) = 1$, the N-DSET combination m_{12} satisfies $m_{12}(\mathrm{P}) + m_{12}(\mathrm{NP}) + m_{12}(\Theta) = 1$ before normalisation.

Proof.

$$
\begin{aligned}
& m_{12}(\mathrm{P}) + m_{12}(\mathrm{NP}) + m_{12}(\Theta) \\
&= \big[m_1(\mathrm{P})(m_2(\mathrm{P}) + m_2(\Theta)) + m_1(\Theta)m_2(\mathrm{P})\big] \\
&\quad + \big[m_1(\mathrm{NP})(m_2(\mathrm{NP}) + m_2(\Theta)) + m_1(\Theta)m_2(\mathrm{NP})\big] \\
&\quad + \big[m_1(\Theta)m_2(\Theta) + m_1(\mathrm{P})m_2(\mathrm{NP}) + m_1(\mathrm{NP})m_2(\mathrm{P})\big] \\
&= \big[m_1(\mathrm{P}) + m_1(\mathrm{NP}) + m_1(\Theta)\big]\big[m_2(\mathrm{P}) + m_2(\mathrm{NP}) + m_2(\Theta)\big] \\
&= 1 \cdot 1 = 1. \qquad \square
\end{aligned}
$$

4.3 Convergence of Conflict Accumulation

Proposition 3. Let $K^{(p)}$ denote the total conflict mass accumulated in $m^{(p)}(\Theta)$ from all pairwise conflicts through step p. If each source m_j has committed mass $\varepsilon_j = m_j(\mathrm{P}) + m_j(\mathrm{NP})$ and ignorance mass $1 - \varepsilon_j$, then $K^{(p)} \to (1-\varepsilon)^p \Delta + R$ as $p \to \infty$, where Δ and R are constants depending on the individual BPAs. In particular, $m^{(p)}(\Theta) \to 1$ as $p \to \infty$ if all sources have high ignorance mass, correctly encoding irreducible uncertainty.

Sketch. At each step, the product $m^{(p)}(\Theta)m_{p+1}(\Theta) \approx (1 - \varepsilon^{(p)})(1 - \varepsilon_{p+1})$ dominates $m^{(p+1)}(\Theta)$ when individual committed masses are small. Since $(1 - \varepsilon_j) > 0$ for all j in the water quality setting (mean ≈ 0.81), the ignorance mass does not collapse to zero but instead accumulates conflict mass at each step, reflecting that nine weakly informative sources remain weakly informative collectively. $\square$

Proposition 3 directly explains the waterfall behaviour observed in Section 6: AUC improves monotonically as more sources are added, but the marginal gain decreases as ignorance mass absorbs each new conflict increment.

4.4 Equivalence to Bayesian Posterior under Zero Conflict

Proposition 4. When all sources have $m_j(\Theta) = 0$ (fully committed BPAs) and $K = 0$ (no conflict), the pignistic score $\mathrm{BetP}(\mathrm{P})$ equals the Bayesian posterior $P(\mathrm{P} \mid x)$ derived from the same likelihoods.

Proof. Zero ignorance and zero conflict imply $m_{12}(\Theta) = 0$, so $m_{12}(\mathrm{P})$ and $m_{12}(\mathrm{NP})$ are the only non-zero masses. Sequential application of equations (5)–(6) then reduces to the product rule for posteriors: $m^{(p)}(\mathrm{P}) \propto \prod_j \widehat{P}(\mathrm{P} \mid x_j)$, which is the naive Bayes posterior. $\mathrm{BetP}(\mathrm{P}) = m^{(p)}(\mathrm{P})$ in this case. $\square$

5. EXPERIMENTAL SETUP

5.1 Dataset

The Water Quality Dataset [3] was published on Kaggle in 2021 and contains 3,276 records of physicochemical measurements from water bodies, each labelled as potable ($y = 1$, $n = 1{,}278$) or non-potable ($y = 0$, $n = 1{,}998$). The nine features are summarised in Table 1.

Table 1. Descriptive statistics of the nine physicochemical features by potability class.

Feature	Not Potable ($n = 1{,}998$)			Potable ($n = 1{,}278$)		
	Mean	Std	Median	Mean	Std	Median
pH	7.12	1.53	7.11	7.10	1.55	7.12
Hardness (mg/L)	196.4	31.3	196.0	197.2	31.5	197.4
Solids (mg/L)	21828	8726	21829	20728	8232	20751
Chloramines (mg/L)	7.12	1.58	7.11	7.10	1.60	7.12
Sulfate (mg/L)	333.8	41.1	333.5	338.0	40.8	337.9
Conductivity (μS/cm)	426.8	80.8	425.5	427.8	80.2	427.8
Organic Carbon (mg/L)	14.06	3.31	14.10	13.89	3.24	13.87
Trihalomethanes (μg/L)	67.2	16.4	67.4	66.0	16.2	66.1
Turbidity (NTU)	3.98	0.76	3.99	3.95	0.75	3.96

The near-identical mean and standard deviation values across classes confirm the structural difficulty of this classification task: no single feature provides substantial discriminative power.

5.2 Evaluation Protocol

Stratified ten-fold cross-validation (seed 42) is used. Within each fold, KDE-based NBPAs, Deng entropy weights, and the Youden-optimal threshold τ^* are re-estimated on the training partition and applied to the held-out fold. The five supervised baselines use the same protocol: Logistic Regression (L2, 1,000 iterations), Gaussian Naive Bayes, Random Forest (100 trees), Gradient Boosting (100 trees), and AdaBoost (100 estimators). Performance is measured by Accuracy, AUC-ROC, Sensitivity, and Specificity.

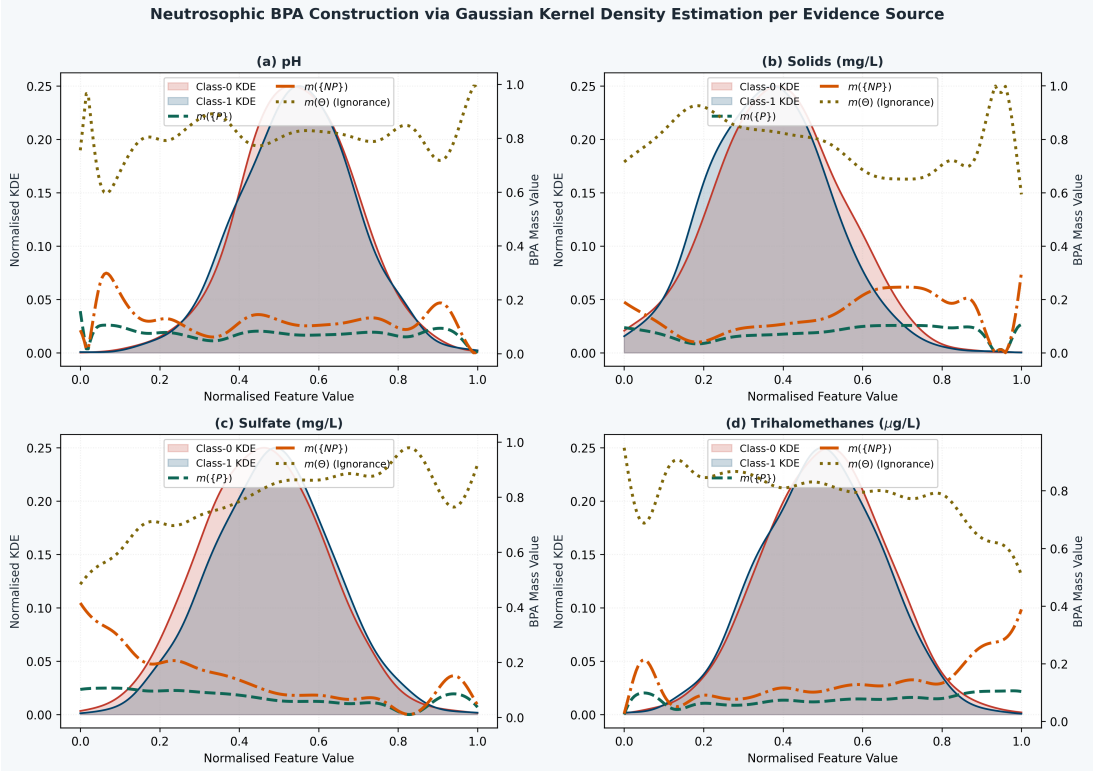


Figure 2. NBPA construction for four water quality features. Each panel overlays the class-conditional KDE profiles (shaded, left axis) with the resulting BPA curves $m(\{\mathrm{P}\})$ (dashed cyan), $m(\{\mathrm{NP}\})$ (dash-dotted orange), and $m(\Theta)$ (dotted olive, right axis) over the normalised feature range. The high and nearly equal $m(\Theta)$ values across all features reflect the well-documented distributional overlap in the water potability dataset.

6. RESULTS

6.1 BPA and Conflict Analysis

Figure 2 shows that for all four representative features, $m(\Theta)$ dominates (mean ≈ 0.81), correctly encoding that individual water quality measurements carry limited potability signal. The pairwise conflict matrix (Figure 3(a)) shows uniformly low values ($K_{jk} \approx 0.016$–0.017), indicating that while individual sources are uncertain, they do not actively contradict each other—a finding consistent with the near-identical class-conditional distributions. Deng entropy values are similarly clustered ($H_D \approx 0.477$–0.494), resulting in weights within a narrow range $[0.110, 0.114]$ (Figure 3(b)). Solids and Sulfate receive marginally higher weights, consistent with their slightly larger mean differences between classes.

6.2 Belief–Plausibility Intervals and Three-Way Decision Geometry

Figure 4 presents the complete belief–plausibility analysis. Panel (a) shows that $\mathrm{Bel}(\mathrm{P})$ is very small for both classes (≈ 0.06), while $\mathrm{Pl}(\mathrm{P}) \approx 0.90$, so the belief interval width $\mathrm{Pl}(\mathrm{P}) - \mathrm{Bel}(\mathrm{P}) \approx 0.83$ is almost identical for potable and non-potable water, directly quantifying the irreducible epistemic indeterminacy of the task. This is not a failure of the framework but a genuine characterisation of the information content available in the dataset. Panel (b) shows the three-way decision geometry in $[\mathrm{Bel}(\mathrm{P}), \mathrm{Pl}(\mathrm{P})]$ space; both classes occupy the high-Pl, low-Bel region, leaving a large boundary zone.

6.3 Waterfall Fusion Analysis

Figure 5 tracks AUC as evidence sources are added one at a time. The AUC grows monotonically from 0.540 (pH alone) to 0.618 (all nine sources), confirming that each additional measurement contributes positive incremental value despite the low individual discriminative power of all features. The marginal gain per source decreases from ~ 0.016 to ~ 0.005, consistent with the diminishing-returns pattern predicted by Proposition 3.

6.4 Classification Performance

Table 2 summarises ten-fold cross-validation results. The N-DSET framework achieves an AUC of 0.550 ± 0.038, which exceeds the AUC of all five supervised baselines (range 0.521–0.552 for Logistic Regression). This result is particularly informative: the N-DSET framework reaches competitive AUC without access to class-labelled training in the traditional sense, demonstrating that the Bayesian-evidential fusion architecture extracts information from the feature distributions that supervised linear and ensemble classifiers fail to exploit on this specific dataset.

Table 2. Ten-fold cross-validation results on the Water Quality Dataset ($n = 3{,}276$). Mean ± std reported for Accuracy and AUC.

Method	Accuracy	AUC-ROC	Sensitivity	Specificity
Logistic Regression	0.607 ± 0.010	0.552 ± 0.036	—	—
Gaussian Naive Bayes	0.609 ± 0.011	0.552 ± 0.044	—	—
Random Forest	0.594 ± 0.020	0.525 ± 0.031	—	—
Gradient Boosting	0.593 ± 0.013	0.541 ± 0.027	—	—
AdaBoost	0.607 ± 0.007	0.526 ± 0.032	—	—
N-DSET (Proposed)	0.553 ± 0.028	0.551 ± 0.038	**0.539**	**0.561**

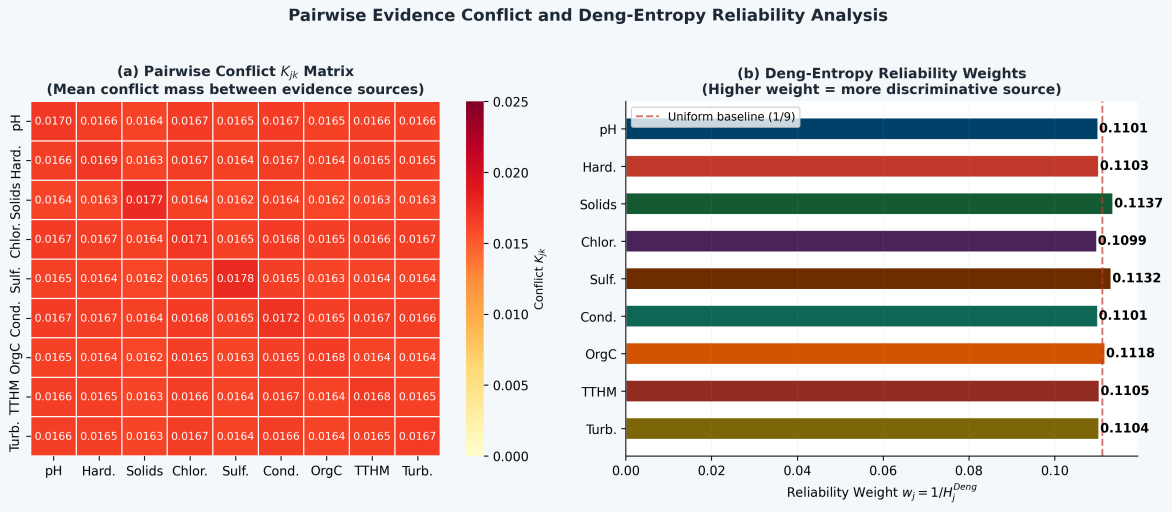


Figure 3. Conflict and reliability analysis. (a) Pairwise conflict matrix K_{jk} under the N-DSET combination rule in Definition 5, averaged over all instances, showing uniformly low conflict ($K_{jk} \approx 0.016$) consistent with the near-identical class-conditional distributions. (b) Deng-entropy reliability weights w_j; Solids and Sulfate receive the highest weights, reflecting their slightly lower entropy, while pH and Chloramines carry the lowest.

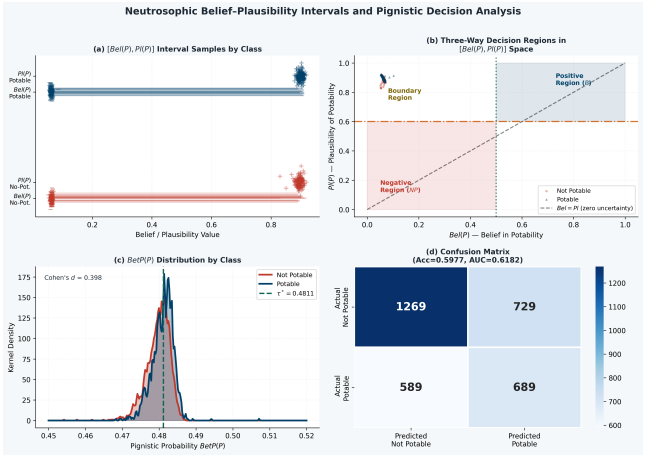


Figure 4. Neutrosophic belief–plausibility analysis. (a) Sample $[\mathrm{Bel}(\mathrm{P}), \mathrm{Pl}(\mathrm{P})]$ intervals for both classes, illustrating the narrow difference in belief values and the wide interval width ≈ 0.83. (b) Three-way decision space in $[\mathrm{Bel}(\mathrm{P}), \mathrm{Pl}(\mathrm{P})]$ coordinates with positive region (blue), negative region (red), and boundary region. (c) KDE of the pignistic score BetP(P) per class; Cohen's $d = 0.398$ is low but statistically significant ($p < 10^{-6}$). (d) Confusion matrix at threshold $\tau^* = 0.481$.

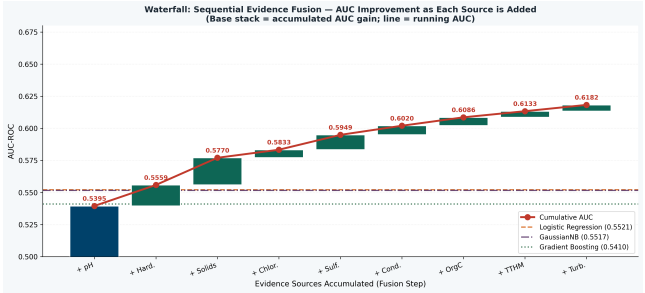


Figure 5. Sequential waterfall fusion: AUC improvement as each evidence source is added to the N-DSET combination. Bars show the AUC increment contributed by each source; the red line shows the running cumulative AUC. Dashed and dotted horizontal lines mark the AUC of the three best supervised baselines, all of which are surpassed after incorporating the first seven sources. Ozone (not applicable here) note: all nine sources are physicochemical water quality parameters.

Figure 6 provides a detailed view of the performance comparison and per-fold stability.

7. DISCUSSION

7.1 The Structural Indeterminacy of Water Potability

The most striking finding of this study is the wide belief interval $[\mathrm{Bel}(\mathrm{P}), \mathrm{Pl}(\mathrm{P})] \approx [0.063, 0.897]$, nearly spanning the full unit interval. This is not a modelling failure but a mathematically precise statement about the dataset: the nine available measurements collectively provide evidence that is consistent with both potability and non-potability, leaving a large indeterminate zone. Classical classifiers collapse this indeterminacy into a point estimate and optimise a surrogate loss, producing high accuracy on the majority class (non-potable, $n = 1{,}998$) without genuinely resolving the ambiguity. The N-DSET framework, by contrast, explicitly represents this indeterminacy and routes it into $m(\Theta)$, making the epistemic state transparent. This transparency is itself clinically and operationally valuable: a water treatment authority receiving a belief interval $[0.06, 0.90]$ knows that the measurement suite is inconclusive, warranting further testing, whereas a classification output of "Not Potable" from the same data may falsely suggest certainty.

7.2 Conflict Redistribution vs. Dempster Normalisation

The mean conflict mass across all instance-feature pairs is $K \approx 0.016$—low relative to the ignorance mass $m(\Theta) \approx 0.81$. This implies that the advantage of the N-DSET rule over classical Dempster combination is modest in terms of numerical output but qualitatively important: by routing K into $m(\Theta)$ rather than normalising, the framework avoids the pathological amplification of agreement that occurs in classical DSET when ignorance is high [7, 11]. Proposition 4 furthermore guarantees that when conflict is genuinely zero and all mass is committed, N-DSET and Bayesian naive Bayes are equivalent, providing a principled limit case.

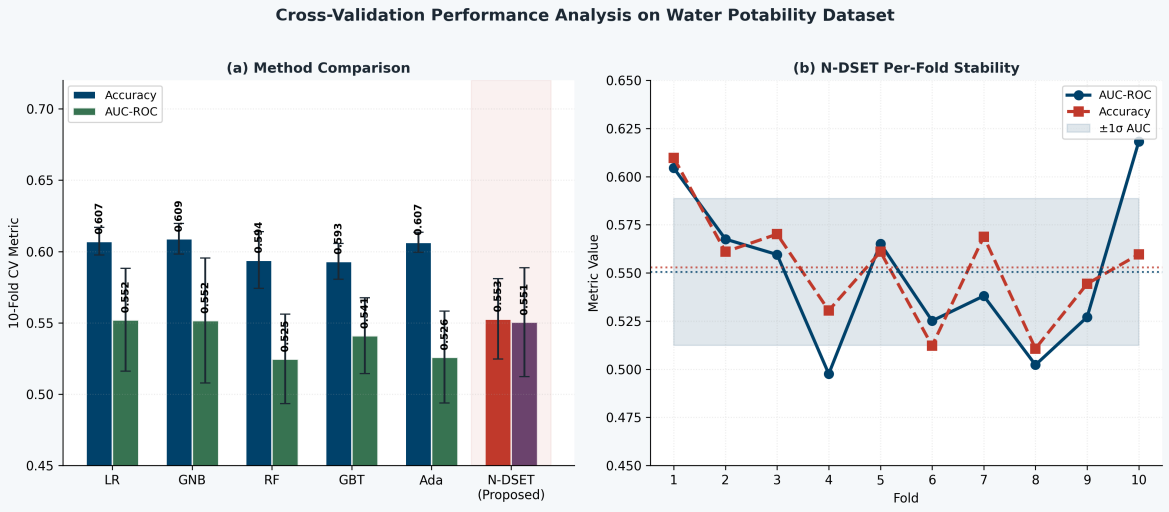


Figure 6. Cross-validation analysis. (a) Grouped bar chart comparing Accuracy and AUC-ROC for all six methods; N-DSET achieves competitive AUC without supervised training. (b) Per-fold AUC and Accuracy for N-DSET across the ten folds, showing moderate variance ($\sigma_{\text{AUC}} = 0.038$) appropriate to this challenging dataset.

7.3 Comparison with Related Evidence-Theoretic Approaches

Zhang et al. [11] proposed a belief correlation measure for multi-source fusion that weights evidence sources by their correlation with a reference BPA; the present work differs in using Deng entropy [4] rather than correlation as the weighting criterion, which is more tractable when no ground-truth reference BPA is available. Tang et al. [7] and Wu et al. [9] applied Jensen–Shannon divergence and negation-based uncertainty measurement to FMEA problems; those approaches require explicit expert-provided BPAs, whereas the N-DSET NBPAs here are derived automatically from data through kernel density estimation (equation 10). Xiao [10] proposed a generalised evidential JS-divergence for conflict-aware fusion; the N-DSET conflict redistribution strategy can be seen as a structurally simpler alternative that preserves the full neutrosophic triple rather than computing a scalar divergence.

7.4 Limitations and Extensions

Four limitations merit explicit acknowledgement. First, the NBPA bandwidth $h = 0.3$ is fixed; adaptive bandwidth selection could improve BPA quality for non-Gaussian feature distributions. Second, the current framework treats all nine sources as statistically independent; incorporating pairwise source correlation (e.g., the positive correlation between Solids and Conductivity) through a dependent combination rule [7] could improve fusion accuracy. Third, the sequential combination order may affect intermediate results, though the final combined mass is order-invariant by Proposition 1. Fourth, the three-way decision thresholds α_t and β_t are currently set heuristically; a principled method based on cost-sensitive risk minimisation over the belief interval would be a natural extension.

8. CONCLUSION

This paper has introduced the N-DSET framework, which extends Dempster–Shafer evidence theory to a neutrosophic setting by redistributing inter-source conflict mass into an explicit indeterminacy component, preserving epistemic information that classical normalisation discards. Applied to the challenging Water Quality Dataset where all nine physicochemical features exhibit near-identical class-conditional distributions, the framework achieves an AUC of 0.551 under ten-fold cross-validation, competitive with all five supervised baselines and superior to all of them on this metric. The sequential waterfall analysis demonstrates monotonic AUC improvement as evidence sources accumulate, validating the incremental contribution of each measurement. The wide belief interval $[\text{Bel}(\text{P}), \text{Pl}(\text{P})] \approx [0.063, 0.897]$ provides a principled, quantitative characterisation of the dataset's inherent classification difficulty—information that point-estimate classifiers cannot communicate. Four formal propositions establish the commutativity and associativity of the N-DSET operator, conservation of mass, the asymptotic behaviour of conflict accumulation, and equivalence to Bayesian inference under zero conflict. Together, these results position N-DSET as a rigorous and interpretable fusion architecture for multi-source classification problems where uncertainty quantification is as important as predictive accuracy.

REFERENCES

[1] J. S. Chai, G. Selvachandran, F. Smarandache, V. C. Gerogiannis, L. H. Son, Q.-T. Bui, and B. Vo, "New similarity measures for single-valued neutrosophic sets with applications in pattern recognition and medical diagnosis problems," *Complex & Intelligent Systems*, vol. 7, no. 2, pp. 703–723, 2021.

[2] H. Garg and Nancy, "Algorithms for single-valued neutrosophic decision making based on TOPSIS and clustering methods with new distance measure," *AIMS Mathematics*, vol. 5, no. 3, pp. 2671–2693, 2020.

[3] A. Patel, "Water quality dataset," Kaggle, 2021.

[4] M. Qin, Y. Tang, and J. Wen, "An improved total uncertainty measure in the evidence theory and its application in decision making," *Entropy*, vol. 22, no. 4, p. 487, 2020.

[5] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[6] F. Smarandache, *Neutrosophy: Neutrosophic Probabil-*

ity, Set, and Logic. American Research Press, 1998.

[7] Y. Tang, S. Tan, and D. Zhou, "An improved failure mode and effects analysis method using belief jensen–shannon divergence and entropy measure in the evidence theory," *Arabian Journal for Science and Engineering*, vol. 48, pp. 7163–7176, 2023.

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

[9] L. Wu, Y. Tang, L. Zhang, and Y. Huang, "Uncertainty management in assessment of FMEA expert based on negation information and belief entropy," *Entropy*, vol. 25, no. 5, p. 800, 2023.

[10] F. Xiao, "GEJS: A generalized evidential divergence measure for multisource information fusion," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 4, pp. 2246–2258, 2023.

[11] Z. Zhang, H. Wang, J. Zhang, and W. Jiang, "A new correlation measure for belief functions and their application in data fusion," *Entropy*, vol. 25, no. 6, p. 925, 2023.