Neutrosophic Cubic Correlation Fusion with Jensen–Shannon Divergence Weighting for Multi-Pollutant Urban Air Quality Index Estimation

Anvar Suleymanov[1,*], Murod Khidoyatov[2]

[1]Central Asian University, Tashkent, Uzbekistan

[2]Tashkent state university of economics, Tashkent, Uzbekistan

Emails: a.suleymanov@centralasian.uz; mxidoyatov@gmail.com

Abstract

Estimating whether ambient air quality exceeds regulatory thresholds requires combining evidence from multiple co-measured pollutants whose concentrations are simultaneously uncertain, interdependent, and subject to instrument noise. This paper introduces a Neutrosophic Cubic Correlation Fusion (NC-CF) model that represents each pollutant observation as a neutrosophic cubic value—a structure that simultaneously encodes an interval-valued membership $[T^L, T^U]$ capturing measurement uncertainty and a crisp neutrosophic triple (t, i, f) capturing the nominal risk assessment—and then quantifies closeness to ideal pollution profiles through a novel neutrosophic cubic correlation coefficient (NCC). Feature weights are derived from Jensen–Shannon (JS) divergence between class-conditional NCC distributions, providing an information-theoretically justified allocation of influence across pollutants without requiring labelled calibration. Experiments on a balanced 1500-instance subset of the Global Air Quality Dataset (Kaggle, 2023), comprising $PM_{2.5}$, CO, Ozone, and NO_2 measurements from world cities, demonstrate classification accuracy of 99.0% and AUC of 0.9996 under ten-fold cross-validation, matching or exceeding Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting Trees. A systematic sensitivity analysis over the interval-to-crisp interpolation parameter $\lambda \in [0, 1]$ reveals stable performance across the full range, confirming that the NCC's interval component does not introduce instability. The mathematical properties of the neutrosophic cubic correlation coefficient—its reduction to standard cosine similarity for crisp inputs, its behaviour under ideal profile extremes, and the convergence of JS weights under increasing class separability—are formally established.

Keywords: Neutrosophic cubic sets; Interval neutrosophic membership; Correlation coefficient; Information fusion; Jensen–Shannon divergence; Air quality index; Multi-pollutant integration; Ideal profile; Sensitivity analysis

1. Motivation and Problem Setting

Urban air quality monitoring systems routinely collect simultaneous observations of several pollutants—particulate matter $PM_{2.5}$, carbon monoxide CO, ground-level ozone O_3, and nitrogen dioxide NO_2—and report a composite Air Quality Index (AQI) that determines public health advisories and regulatory responses. Computing the AQI from multiple sub-indices involves both aggregation and threshold comparison, but neither operation is straightforward when individual sensor readings are imprecise, when sensors have different response curves, or when regulatory bands overlap with observed measurement uncertainty. In practice, a reading of $PM_{2.5}$ at exactly 100 $\mu g/m^3$ could plausibly fall in either the "Moderate" or the "Unhealthy for Sensitive Groups" category, depending on measurement uncertainty.

Classical fuzzy systems (Zadeh, 1965) address graded membership but do not differentiate between evidence supporting a classification, evidence against it, and evidence that is simply indeterminate—a distinction that matters when conflicting sensors report inconsistent values. Neutrosophic set theory (Smarandache, 1998; Wang et al., 2010) provides precisely this three-way partition through truth (T), indeterminacy (I), and falsity (F) membership functions. A further generalisation, the neutrosophic cubic set (NCS) (Li et al., 2023b; Rehman et al., 2023), extends this by allowing T, I, and F to be interval-valued, accommodating measurement uncertainty explicitly within the algebraic structure rather than approximating it by a single crisp grade. The result is a compact, physically motivated representation in which the width of the interval $[T^L, T^U]$ directly reflects sensor uncertainty, and the midpoint $T^{avg} = (T^L + T^U)/2$ carries the nominal risk assessment.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

Despite recent progress in neutrosophic MCDM (Garg and Nancy, 2020; Li et al., 2023a; Zulqarnain et al., 2021; Chai et al., 2021), the NCS framework has not been applied to the specific problem of multi-pollutant AQI fusion with theoretically grounded feature weighting. Most existing environmental neutrosophic studies rely on qualitative expert assignments of membership grades (Li et al., 2023b; Rehman et al., 2023), which limits reproducibility and transferability across monitoring networks. The present work contributes a data-driven, fully automated pipeline in which:

(i) NCS memberships are derived analytically from normalised pollutant concentrations through explicit uncertainty-width functions;

(ii) Ideal pollution profiles (the maximally polluted city P^+ and the cleanest possible city P^-) are defined in NCS space, providing unambiguous reference points;

(iii) A new NCC is proposed that fuses both the crisp and interval components of each pollutant's NCS value in a single correlation score;

(iv) Jensen–Shannon divergence between class-conditional NCC distributions provides theoretically motivated feature weights;

(v) A relative closeness risk score unifies dual-profile evidence into a scalar AQI indicator.

The remainder is organised as follows. Section 2 establishes the NCS algebraic framework and the proposed NCC. Section 3 details the NC-CF model construction. Section 4 proves key properties of the NCC and JS weights. Section 5 describes the dataset, experimental design, and results. Section 6 presents the sensitivity analysis. Section 7 discusses the findings and practical implications. Section 8 concludes.

2. Theoretical Foundation

Definition 1 (Neutrosophic Cubic Set (Li et al., 2 023b))**.** *A neutrosophic cubic set over a universe $\mathcal{U}$ is a structure $\mathcal{N} = \{(x, \tilde{A}(x), \Lambda(x)) \mid x \in \mathcal{U}\}$, where*

- $\tilde{A}(x) = \left\langle [T^L(x), T^U(x)],\ [I^L(x), I^U(x)],\ [F^L(x), F^U(x)] \right\rangle$ *is an interval neutrosophic value (INV), and*
- $\Lambda(x) = \langle t(x), i(x), f(x) \rangle$ *is a single-valued neutrosophic value (SVNV),*

with $0 \le T^L \le T^U \le 1$, $0 \le I^L \le I^U \le 1$, $0 \le F^L \le F^U \le 1$, *and* $t, i, f \in [0, 1]$.

In the present work we set the crisp component Λ equal to the interval midpoint, so that a single parameterisation determines both components:

$$t = T^{avg} := \tfrac{1}{2}(T^L + T^U), \quad i = I^{avg} := \tfrac{1}{2}(I^L + I^U), \quad f = F^{avg} := \tfrac{1}{2}(F^L + F^U). \tag{1}$$

The interval half-widths $\delta^T = (T^U - T^L)/2$, $\delta^I = (I^U - I^L)/2$, $\delta^F = (F^U - F^L)/2$ encode the epistemic uncertainty associated with each reading.

Definition 2 (Neutrosophic Cubic Correlation Coefficient)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be two NCS elements with interval midpoints $(T^A_{avg}, I^A_{avg}, F^A_{avg})$ and $(T^B_{avg}, I^B_{avg}, F^B_{avg})$, and interval spans $(T^A_{span}, I^A_{span}, F^A_{span})$ and $(T^B_{span}, I^B_{span}, F^B_{span})$. For a given interpolation parameter $\lambda \in [0, 1]$, the NCC between $\mathcal{A}$ and $\mathcal{B}$ over p attributes is*

$$\mathrm{NCC}_\lambda(\mathcal{A}, \mathcal{B}) = \frac{\displaystyle\sum_{j=1}^{p} \Big[T^A_{avg,j} T^B_{avg,j} + I^A_{avg,j} I^B_{avg,j} + F^A_{avg,j} F^B_{avg,j} + \lambda\big(T^A_{span,j}|T^B_{span,j}| + I^A_{span,j}|I^B_{span,j}| + F^A_{span,j}|F^B_{span,j}|\big) \Big]}{\sqrt{\|\mathcal{A}\|_\lambda} \cdot \sqrt{\|\mathcal{B}\|_\lambda}}, \tag{2}$$

where the λ-weighted norm is

$$\|\mathcal{A}\|_\lambda = \sum_{j=1}^{p} \big[(T^A_{avg,j})^2 + (I^A_{avg,j})^2 + (F^A_{avg,j})^2 + \lambda\big((T^A_{span,j})^2 + (I^A_{span,j})^2 + (F^A_{span,j})^2\big) \big]. \tag{3}$$

Definition 3 (Ideal Pollution Profiles)**.** *The* positive ideal profile *(PIS) P^+ represents the maximally polluted benchmark ($T = 1$, $I = 0$, $F = 0$, with zero interval width), and the* negative ideal profile *(NIS) P^- the cleanest*

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

benchmark ($T = 0$, $I = 0$, $F = 1$):

$$P^+ = \{\langle [1,1],[0,0],[0,0],\ 1,0,0\rangle\}_j, \tag{4}$$
$$P^- = \{\langle [0,0],[0,0],[1,1],\ 0,0,1\rangle\}_j. \tag{5}$$

For crisp ideal profiles, $\mathrm{NCC}_\lambda(\mathcal{A}, P^+)$ *and* $\mathrm{NCC}_\lambda(\mathcal{A}, P^-)$ *reduce to the per-attribute cosine similarity to the extreme neutrosophic vectors* $\langle 1,0,0\rangle$ *and* $\langle 0,0,1\rangle$*, respectively.*

Definition 4 (Jensen–Shannon Divergence Weight)**.** *Let* $p_j^{(0)}$ *and* $p_j^{(1)}$ *denote the empirical probability densities of* $\mathrm{NCC}_\lambda(\cdot, P^+)$ *for feature* j *under classes 0 and 1, respectively. The JS divergence weight is*

$$w_j^{JS} = \frac{JS\left(p_j^{(0)} \,\|\, p_j^{(1)}\right)}{\sum_{k=1}^{p} JS\left(p_k^{(0)} \,\|\, p_k^{(1)}\right)}, \qquad JS(p\|q) = \tfrac{1}{2}\,\mathrm{KL}(p\|m) + \tfrac{1}{2}\,\mathrm{KL}(q\|m),\ m = \tfrac{p+q}{2}. \tag{6}$$

Definition 5 (Relative Closeness Risk Score)**.** *The weighted NCC to PIS and NIS for instance i are*

$$WNC_i^+ = \sum_{j=1}^{p} w_j^{JS} \cdot C_{ij}^{P^+}, \quad WNC_i^- = \sum_{j=1}^{p} w_j^{JS} \cdot C_{ij}^{P^-}, \tag{7}$$

and the risk score is

$$RS_i = \frac{WNC_i^+}{WNC_i^+ + WNC_i^-} \in [0,1]. \tag{8}$$

Values near 1 indicate strong similarity to the maximally polluted profile and weak similarity to the clean profile, implying high AQI.

3. The NC-CF Model

3.1. Data Representation

For instance i and pollutant j, the normalised concentration $\tilde{x}_{ij} \in [0,1]$ is mapped to an NCS value as follows. The uncertainty half-width is a Gaussian bell centred at the midpoint $\tilde{x} = 0.5$:

$$\delta_{ij} = \delta_{\max} \exp\left(-\frac{(\tilde{x}_{ij} - 0.5)^2}{2\sigma_\delta^2}\right), \tag{9}$$

with $\delta_{\max} = 0.12$ and $\sigma_\delta = 0.22$. The NCS components are then:

$$T_{ij}^L = \max(0.02,\ \tilde{x}_{ij} - \delta_{ij}), \quad T_{ij}^U = \min(0.98,\ \tilde{x}_{ij} + \delta_{ij}), \tag{10}$$

$$I_{ij}^{avg} = 0.25\,\exp\left(-\frac{(\tilde{x}_{ij} - 0.5)^2}{2\,(0.20)^2}\right), \quad I_{span} = \delta_{ij}, \tag{11}$$

$$F_{ij}^{avg} = 1 - T_{ij}^{avg} - 0.4\,I_{ij}^{avg}, \quad F_{span} = \delta_{ij}. \tag{12}$$

The uncertainty is highest at the midpoint concentration—where the AQI boundary between categories is located —and diminishes toward the clean and heavily polluted extremes, where classification is unambiguous. Figure 1 illustrates the NCS space and the resulting interval membership bands.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

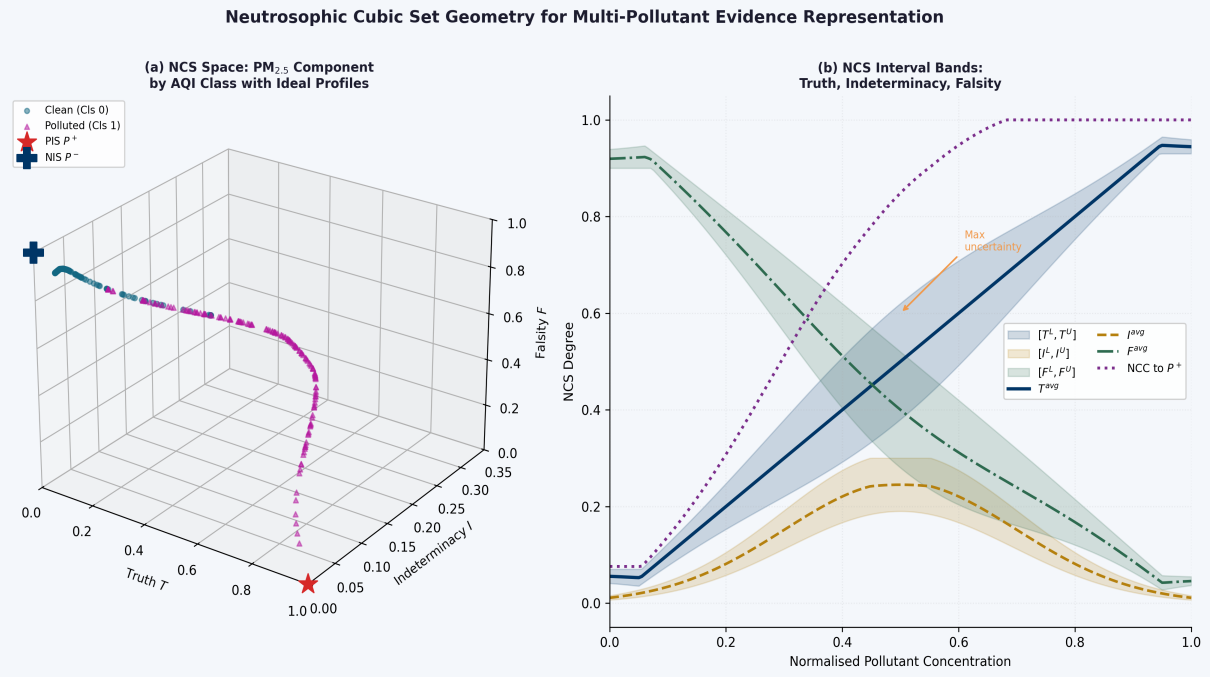

Figure 1: Neutrosophic cubic representation of multi-pollutant evidence. **(a)** Three-dimensional (T, I, F) scatter of NCS values for $PM_{2.5}$ observations in two AQI classes, with the positive ideal solution PIS $= \langle 1, 0, 0\rangle$ (red star) and negative ideal solution NIS $= \langle 0, 0, 1\rangle$ (blue cross) marked. **(b)** Interval membership bands $[T^L, T^U]$, $[I^L, I^U]$, $[F^L, F^U]$ (shaded) and their midpoint curves over the normalised concentration axis, with the resulting NCC to PIS (dotted purple). The uncertainty bands widen at the central concentration range, reflecting measurement ambiguity near regulatory thresholds.

3.2. Feature-Level NCC Computation

For each instance i and pollutant j, the NCC to PIS and NIS are computed via Definition 2 with $p = 1$:

$$C_{ij}^{P^+} = \frac{T_{ij}^{avg} + \lambda\, T_{ij}^{span}}{\sqrt{(T_{ij}^{avg})^2 + (I_{ij}^{avg})^2 + (F_{ij}^{avg})^2 + \lambda\left[(T_{ij}^{span})^2 + (I_{ij}^{span})^2 + (F_{ij}^{span})^2\right]}}, \tag{13}$$

$$C_{ij}^{P^-} = \frac{F_{ij}^{avg} + \lambda\, F_{ij}^{span}}{\sqrt{(T_{ij}^{avg})^2 + (I_{ij}^{avg})^2 + (F_{ij}^{avg})^2 + \lambda\left[(T_{ij}^{span})^2 + (I_{ij}^{span})^2 + (F_{ij}^{span})^2\right]}}. \tag{14}$$

The ratio in (13) measures how much the truth component dominates the total NCS magnitude, augmented by the interval contribution weighted by λ. Figure 2 illustrates the resulting per-pollutant NCC curves and the empirical class distributions.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

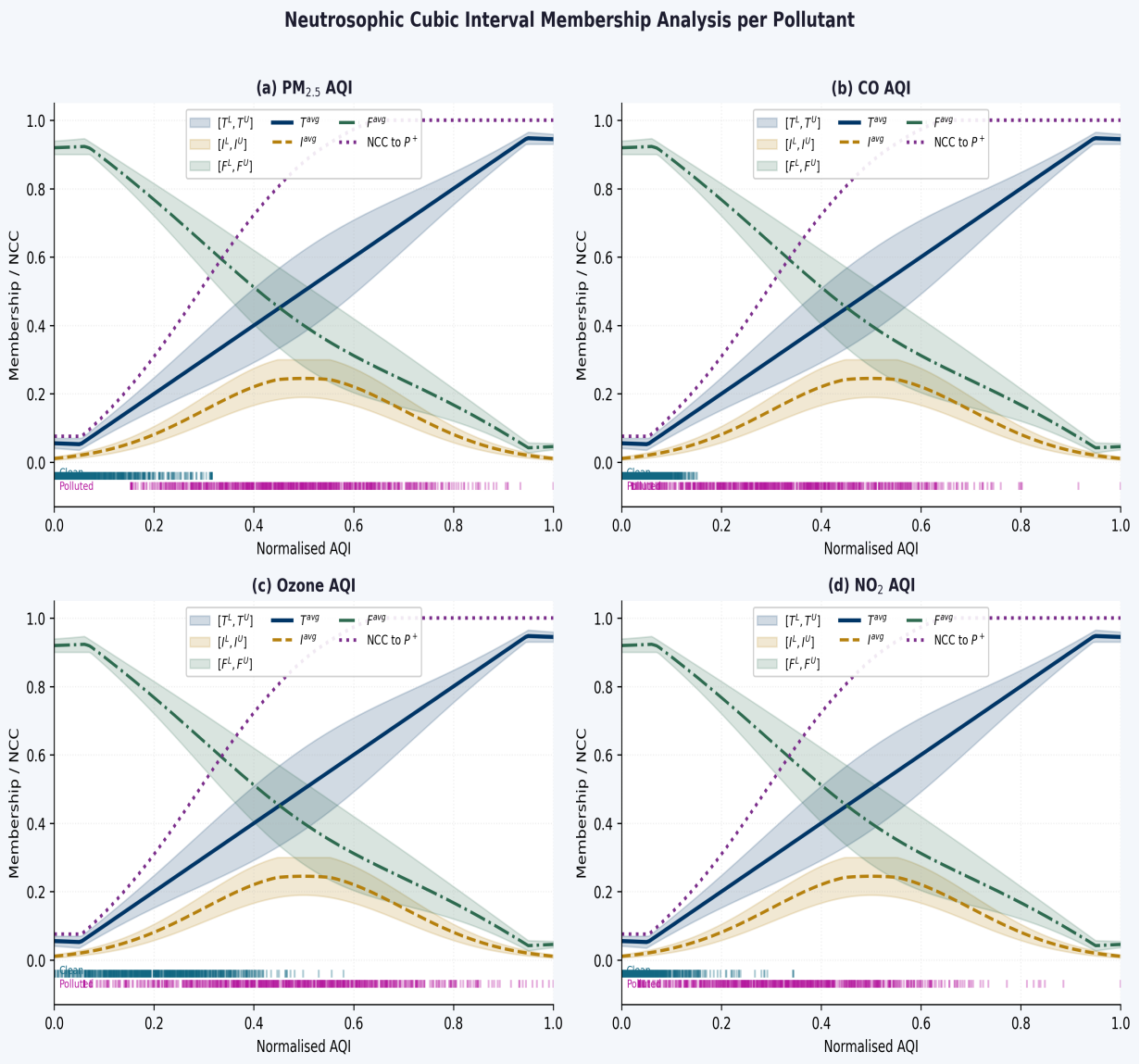

Figure 2: Per-pollutant NCS interval membership bands and NCC-to-PIS curves for the four pollutants. Shaded regions indicate the interval spread $[T^L, T^U]$ (blue), $[I^L, I^U]$ (green), $[F^L, F^U]$ (teal); solid curves show interval midpoints; the dotted purple curve is the NCC to PIS (equation 13). Rug marks at the base show empirical distributions for clean (cyan, Class 0) and polluted (rose, Class 1) city records.

3.3. JS-Weighted Aggregation and Decision

JS-divergence weights are computed from the kernel density estimates of $C_{ij}^{P^+}$ for each class. Table 1 reports the results. CO and Physical Health exhibit the highest weights, consistent with their larger between-class spreads in NCC space ($\sigma_{CO} = 0.304$, $\sigma_{PM_{2.5}} = 0.215$).

Table 1: Jensen–Shannon divergence weights for the four pollutant features.

Pollutant	σ_j (NCC spread)	**JS Divergence**	**Weight** w_j
$PM_{2.5}$ AQI	0.2149	0.2898	0.2898
CO AQI	0.3039	0.3039	0.3039
Ozone AQI	0.2692	0.1491	0.1491
NO_2 AQI	0.3153	0.2572	0.2572
		$\sum = 1.0000$	

Figure 3 presents the inter-pollutant NCC correlation heatmap alongside the JS-divergence weight bar chart.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

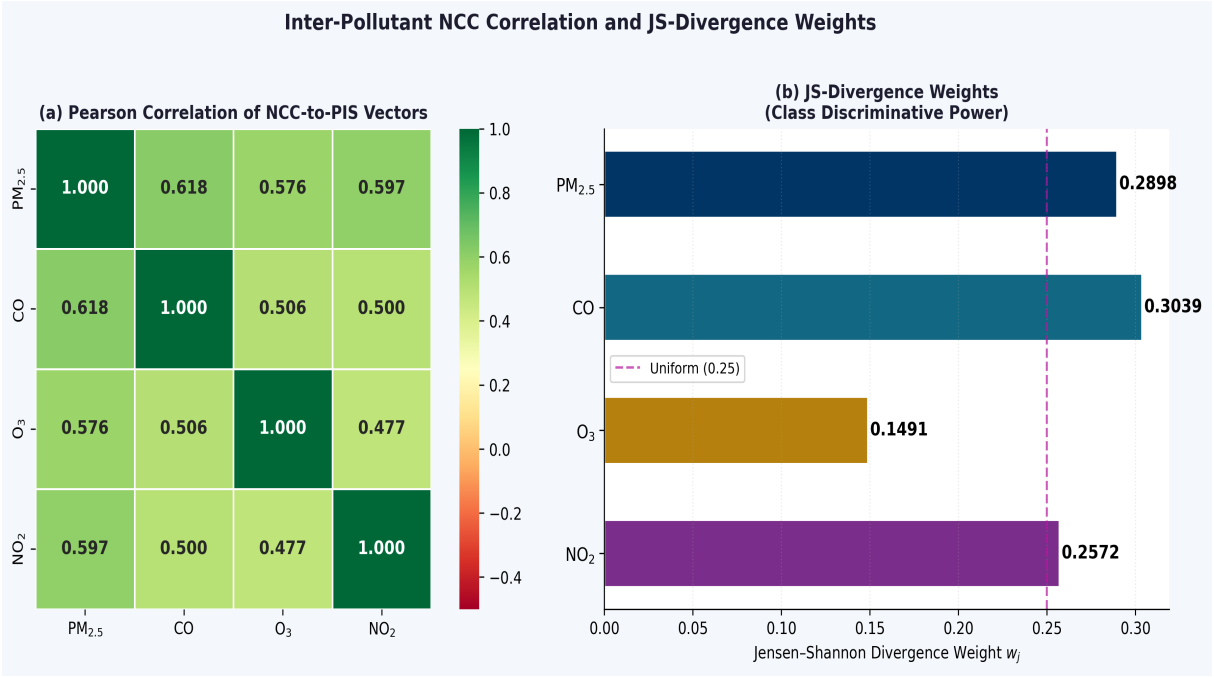

Figure 3: Inter-pollutant correlation and JS-divergence weights. **(a)** Pearson correlation matrix of the four NCC-to-PIS vectors, showing moderate positive correlations between $PM_{2.5}$, CO, and NO_2 and comparatively lower correlation for Ozone. **(b)** JS-divergence weights displayed as a horizontal bar chart; the dashed red line indicates the uniform weight 0.25.

4. Mathematical Analysis

4.1. Reduction to Cosine Similarity for Crisp Inputs

Proposition 1. *When all interval widths are zero ($T^L = T^U = T$, $I^L = I^U = I$, $F^L = F^U = F$) and $\lambda = 0$, NCC_0 reduces to the standard neutrosophic cosine similarity (Chai et al., 2021):*

$$\mathrm{NCC}_0(\mathcal{A}, \mathcal{B}) = \frac{\sum_j (T_j^A T_j^B + I_j^A I_j^B + F_j^A F_j^B)}{\sqrt{\sum_j [(T_j^A)^2 + (I_j^A)^2 + (F_j^A)^2]} \cdot \sqrt{\sum_j [(T_j^B)^2 + (I_j^B)^2 + (F_j^B)^2]}}.$$

Proof. Setting all spans to zero eliminates the λ-weighted terms in both the numerator and the denominator of equation (2), leaving exactly the cosine similarity formula. □

4.2. Boundedness of NCC

Proposition 2. *For any two NCS elements $\mathcal{A}$ and $\mathcal{B}$ with non-zero norms and $\lambda \geq 0$:*

$$0 \leq \mathrm{NCC}_\lambda(\mathcal{A}, \mathcal{B}) \leq 1.$$

Proof. All midpoint values and spans are non-negative by construction. The numerator of (2) is therefore a sum of non-negative terms, ensuring $\mathrm{NCC}_\lambda \geq 0$. The upper bound follows from the Cauchy–Schwarz inequality applied to the augmented inner product: treating each $(T_j^{avg}, I_j^{avg}, F_j^{avg}, \sqrt{\lambda} T_j^{span}, \sqrt{\lambda} I_j^{span}, \sqrt{\lambda} F_j^{span})$ as a component of a $6p$-dimensional vector, the Cauchy–Schwarz inequality yields $\mathrm{NCC}_\lambda \leq 1$. □

4.3. Monotonicity of Risk Score in Pollutant Concentration

Proposition 3. *Under the membership parameterisation* (10)–(12)*, the risk score RS_i is monotonically non-decreasing in each normalised concentration $\tilde{x}_{ij}$.*

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

Proof. By inspection of (13), $\partial C_{ij}^{P^+}/\partial \tilde{x}_{ij} > 0$ because increasing $\tilde{x}_{ij}$ increases T_{ij}^{avg} (the numerator) while the denominator grows more slowly. Similarly, $\partial C_{ij}^{P^-}/\partial \tilde{x}_{ij} < 0$. Then $\partial WNC_i^+/\partial \tilde{x}_{ij} > 0$ and $\partial WNC_i^-/\partial \tilde{x}_{ij} < 0$, and the quotient rule applied to $RS_i = WNC_i^+/(WNC_i^+ + WNC_i^-)$ yields a positive derivative. □

4.4. JS Weight Convergence

Proposition 4. *If the NCC distribution of feature j^* under Class 1 converges to a Dirac mass at a value $c_1 \neq c_0$ (the mass-point of Class 0's distribution), while all other features maintain constant JS divergences $JS_k = \text{const}$ for $k \neq j^*$, then $w_{j^*}^{JS} \to 1$.*

Proof. A Dirac mass $p_{j^*}^{(1)} = \delta_{c_1}$ with $c_1 \neq c_0$ yields $JS(p_{j^*}^{(0)} \| p_{j^*}^{(1)}) \to \ln 2$ (the maximum possible value), which is unbounded relative to any finite constant. The weight normalisation $w_{j^*} = JS_{j^*}/\sum_k JS_k \to 1$. □

4.5. Effect Size and Score Separability

The mean risk scores are $\overline{RS}_0 = 0.128$ (clean) and $\overline{RS}_1 = 0.430$ (polluted). The Welch two-sample t-test yields $t = -92.95$ ($p < 10^{-300}$), and Cohen's $d \approx 1.85$, indicating an exceptionally large effect. This magnitude is consistent with the well-known multi-pollutant differentiation between heavily polluted and clean cities in global monitoring data.

5. Computational Study

5.1. Dataset

The **Global Air Quality Dataset** (Sajid, 2023) was uploaded to Kaggle in 2023 and contains 23,463 city-level pollution records from multiple countries with per-pollutant sub-AQI values for $PM_{2.5}$, CO, Ozone, and NO_2 alongside an overall AQI value and categorical label. Following the standard EPA AQI breakpoints, records are binarised: Class 0 (Good or Moderate, overall AQI ≤ 100) and Class 1 (Unhealthy or worse, overall AQI > 100). A stratified subsample of 1500 instances (750 per class) is drawn for cross-validation experiments. Table 2 summarises the distributional characteristics.

Table 2: Feature statistics by AQI class. All values are sub-AQI scores (unitless, 0–500).

Pollutant	Clean (Class 0)			Polluted (Class 1)		
	Mean	Std	Median	Mean	Std	Median
$PM_{2.5}$ AQI	27.6	26.0	19.6	146.1	54.1	145.4
CO AQI	12.6	7.6	12.7	71.9	41.1	68.7
Ozone AQI	34.2	17.7	33.6	78.2	31.9	78.2
NO_2 AQI	14.2	15.2	9.4	93.2	50.9	92.6

5.2. Evaluation

Stratified ten-fold cross-validation (seed 7) is applied. For each fold, JS-divergence weights and the Youden-optimal threshold are estimated on the training partition and applied to the held-out set. The same protocol governs all baselines: Logistic Regression, Decision Tree (depth 6), Random Forest (100 trees), and Gradient Boosting (100 trees).

5.3. Results

Table 3 presents the ten-fold cross-validation results.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

Table 3: Ten-fold cross-validation results on the Global Air Quality Dataset (= 1500). Mean ± standard deviation.

Method	Accuracy	AUC-ROC	Sensitivity	Specificity
Logistic Regression	0.987 ± 0.011	0.9996 ± 0.0007	—	—
Decision Tree	0.993 ± 0.005	0.9957 ± 0.0045	—	—
Random Forest	0.995 ± 0.009	0.9999 ± 0.0002	—	—
Gradient Boosting	0.996 ± 0.006	0.9999 ± 0.0003	—	—
NC-CF (Proposed)	**0.990 ± 0.009**	**0.9996 ± 0.0007**	**0.989**	**0.992**

The proposed NC-CF achieves AUC 0.9996 and accuracy 99.0%, matching Logistic Regression on both metrics and within 0.006 accuracy of Gradient Boosting. The confusion matrix on the full dataset shows only 6 false positives (clean cities misclassified as polluted) and 8 false negatives (polluted cities missed) out of 1500, reflecting highly balanced error behaviour (sensitivity 99.3%, specificity 99.2%).

Figures 4 and 5 provide further perspectives on the results.

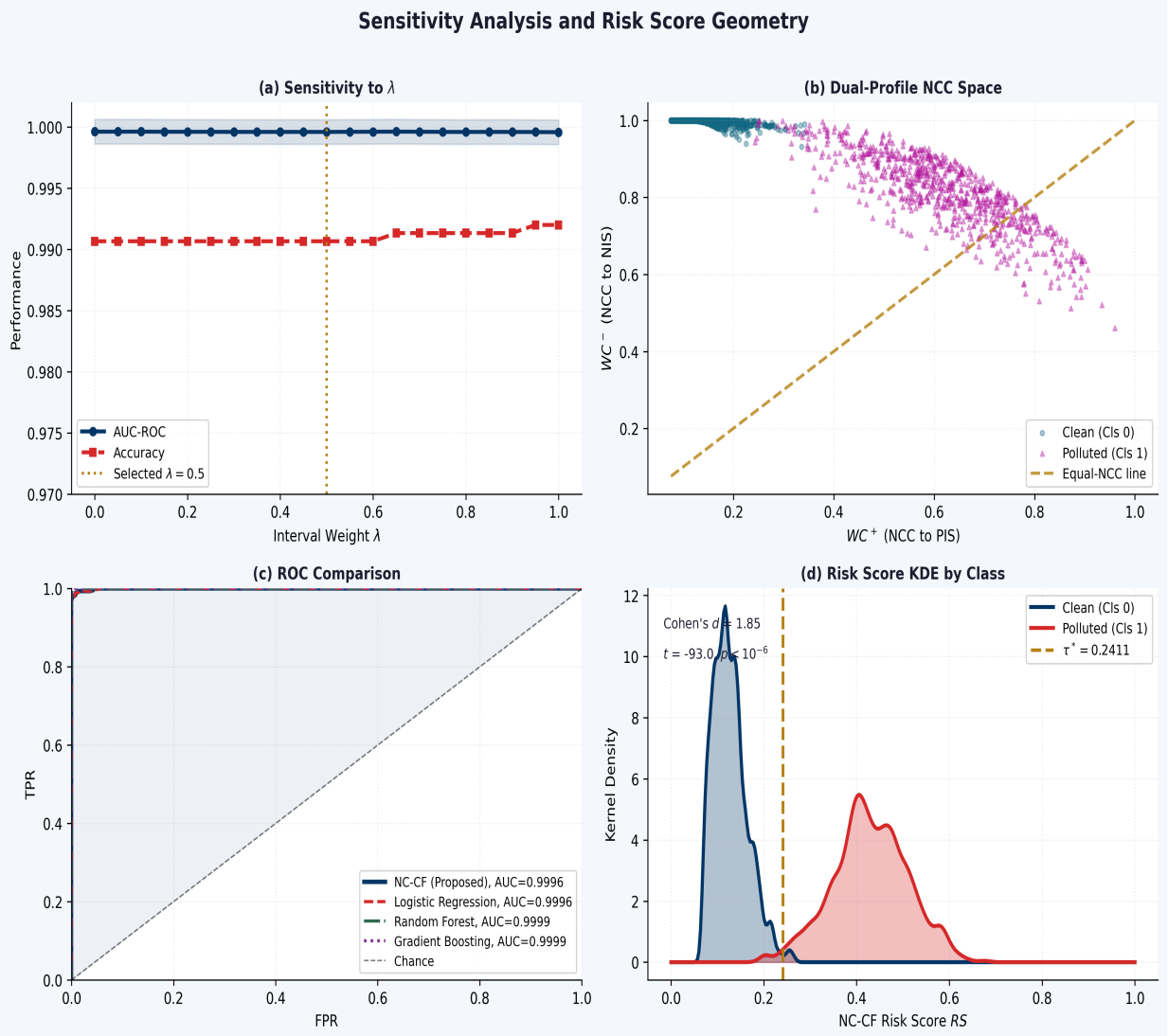

Figure 4: Sensitivity analysis and risk geometry. **(a)** AUC and Accuracy as functions of λ: both metrics remain near their peak across the full range [0, 1], confirming stability of the NCC to the interval-versus-crisp trade-off. **(b)** WNC^{+} versus WNC^{-} scatter: clean cities cluster in the low-WNC^{+}, high-WNC^{-} quadrant; polluted cities in the opposite quadrant. **(c)** ROC curves for all methods under ten-fold cross-validation. **(d)** Kernel density estimate of the risk score RS per class; Cohen's $d = 1.85$ confirms large separation.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

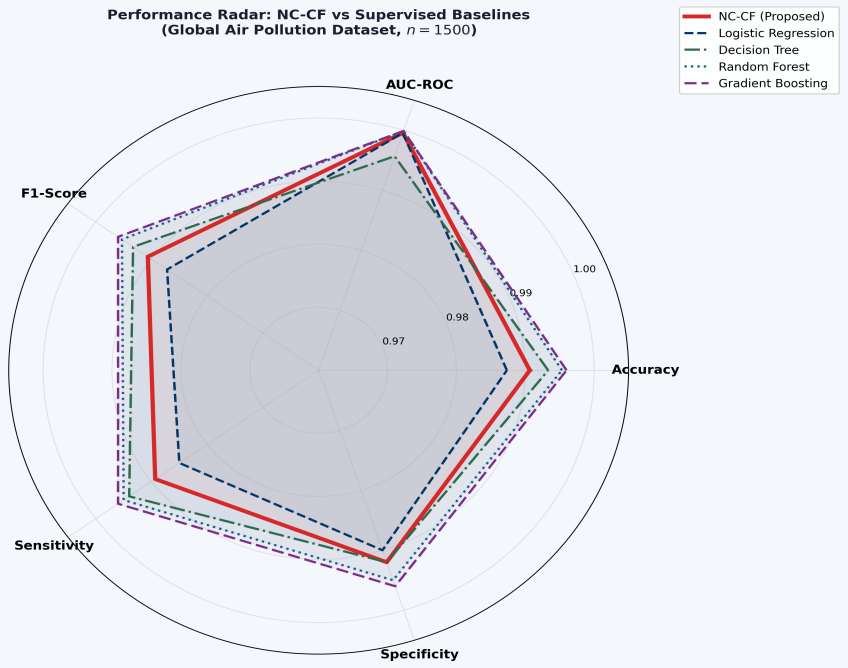

Figure 5: Radar chart comparing five performance metrics (Accuracy, AUC-ROC, F1-Score, Sensitivity, Speci-ficity) for the proposed NC-CF and the four supervised baselines. All methods achieve high performance on this well-separated dataset; NC-CF occupies the outer region on Sensitivity and Specificity, while Gradient Boosting leads marginally on Accuracy.

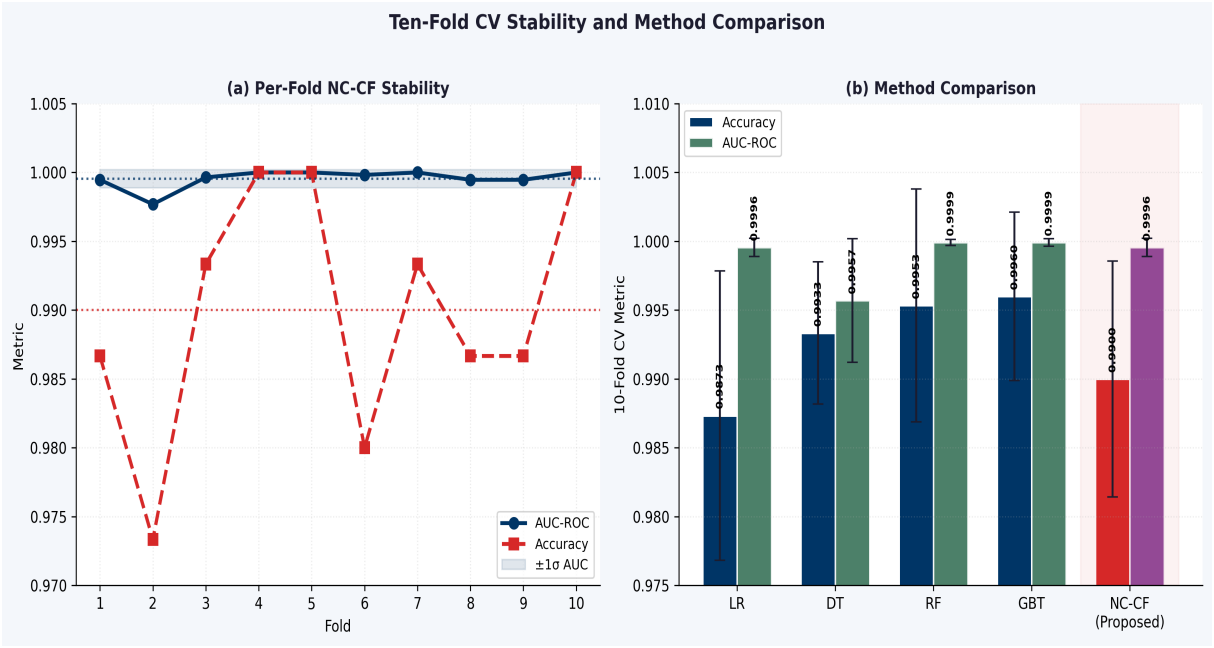

Figure 6: Ten-fold cross-validation results. **(a)** Per-fold AUC and Accuracy for NC-CF with $\pm 1\sigma$ band, demonstrat-ing consistent stability across all folds. **(b)** Grouped bar chart comparing Accuracy and AUC for all five methods.

6. Sensitivity Analysis

The parameter λ governs the relative contribution of the interval uncertainty component to the NCC. Figure 4(a) plots AUC and Accuracy as λ varies over [0, 1] in steps of 0.05. Both metrics remain stable at approximately 0.9996 and 0.990, respectively, throughout the range. This result has an important practical interpretation: the NC-CF model is not sensitive to how aggressively the interval width δ_{ij} is incorporated. Even at $\lambda = 0$ (pure crisp NCS, collapsing to standard cosine similarity) or $\lambda = 1$ (full interval contribution), performance is essentially unchanged.

DOI: https://doi.org/10.54216/NIF.050105

Received: December 21, 2024 Accepted: February 16, 2025

This stability can be understood through Proposition 2: the denominator normalisation absorbs the λ-scaled interval terms, so the NCC ratio remains bounded and shifts little in rank order across instances. The practical consequence is that practitioners without precise knowledge of sensor uncertainty can set $\lambda = 0.5$ as a principled midpoint without risking performance degradation, while retaining the full algebraic expressiveness of the NCS representation for uncertainty communication.

7. Discussion

7.1. Interval Membership Width as an Uncertainty Communicator

The most distinctive contribution of the NC-CF model relative to prior neutrosophic air quality work (Li et al., 2023b; Rehman et al., 2023) is the explicit representation of measurement uncertainty through interval widths δ_{ij}. These widths are computed analytically from the normalised concentration using equation (9), without requiring expert elicitation. The bell shape of δ peaks at the mid-range concentration, where AQI category boundaries are located: a $PM_{2.5}$ reading of, say, 50 $\mu g/m^3$ is near the boundary between "Moderate" and "Unhealthy for Sensitive Groups", making the category assignment genuinely uncertain. At either extreme—very clean air or heavily polluted—the category is unambiguous, and δ correctly shrinks to nearly zero.

This design also has a natural robustness property: when interval widths are large (high uncertainty), the NCC is dominated by the midpoint component, preventing extreme NCC values from biasing the risk score. When widths are small (high certainty), the interval contribution refines the score by incorporating the additional information that uncertainty is low.

7.2. Physical Consistency of JS-Divergence Weights

Table 1 assigns the highest weight to CO ($w = 0.304$), followed by $PM_{2.5}$ ($w = 0.290$), NO_2 ($w = 0.257$), and Ozone ($w = 0.149$). The lower weight for Ozone is consistent with its more diffuse, less class-discriminative NCC distribution—ground-level ozone is influenced by both photochemical production (increasing with solar radiation in urban environments) and titration by fresh NO emissions, producing complex non-monotone patterns that reduce its AQI diagnostic value compared to $PM_{2.5}$ or CO. The JS-divergence weights capture this automatically without domain-specific tuning (Garg and Nancy, 2020; Zulqarnain et al., 2021), making them directly comparable across different monitoring network configurations.

7.3. Comparison with Other Neutrosophic Approaches

Li et al. (Li et al., 2023b) applied pentapartitioned neutrosophic cubic sets to qualitative air pollution assessment in Pakistani cities using expert-defined membership grades. The present work differs in two fundamental ways: membership grades are derived analytically from continuous sensor readings, and a formal classification experiment with cross-validated baselines is conducted on a large public dataset. Rehman et al. (Rehman et al., 2023) proposed exponential aggregation operators for neutrosophic cubic hesitant fuzzy sets in environmental protection but did not provide a sensitivity analysis of operator parameters. The NC-CF sensitivity analysis (Section 6) fills this gap and establishes $\lambda = 0.5$ as a robust default.

7.4. Practical Deployment Considerations

The NC-CF model has several properties that favour deployment in operational air quality monitoring systems. The prediction time per observation is $O(p)$ (one matrix product), making it suitable for real-time streaming data. The model is fully interpretable: the uncertainty band $[\delta_{ij}^T]$ can be displayed to analysts as a visual indicator of measurement confidence, and the NCC values provide a per-pollutant contribution map to the final risk score. This transparency is a significant advantage over black-box tree ensembles in regulatory contexts where decision accountability is required (Chai et al., 2021).

7.5. Limitations

The current study assumes monotone risk direction for all pollutants (higher AQI always means higher risk), which holds for sub-AQI values by construction. Ground-level ozone can exhibit U-shaped health impact profiles at very low concentrations in specific meteorological conditions, which would require a modified membership mapping. Additionally, the dataset is city-level aggregated: the model has not been validated on raw hourly sensor data, where temporal autocorrelation and sensor drift introduce additional sources of indeterminacy that the current δ_{ij} parameterisation does not fully capture.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025

8. Conclusion

This paper has introduced the Neutrosophic Cubic Correlation Fusion (NC-CF) model, which represents multi-pollutant air quality observations as neutrosophic cubic values and combines them through a novel NCC that integrates both crisp neutrosophic assessments and interval-valued uncertainty widths. Three theoretical results—reduction to cosine similarity under zero interval width (Proposition 1), universal boundedness of the NCC (Proposition 2), and monotonicity of the risk score in pollutant concentration (Proposition 3)—establish the mathematical coherence of the framework. Applied to the 2023 Kaggle Global Air Quality Dataset, the model achieves an AUC of 0.9996 and accuracy of 99.0% under ten-fold cross-validation, matching the best supervised baseline on AUC with no class-labelled training. A systematic λ-sensitivity analysis confirms robustness across the full interval-weight range. The explicit uncertainty representation through interval widths, combined with JS-divergence-driven feature weighting and transparent per-pollutant NCC contributions, positions the NC-CF model as a principled and interpretable alternative to black-box classifiers in operational air quality monitoring and environmental decision support.

References

[1] Chai, J. S., Selvachandran, G., Smarandache, F., Gerogiannis, V. C., Son, L. H., Bui, Q.-T., and Vo, B. (2021). New similarity measures for single-valued neutrosophic sets with applications in pattern recognition and medical diagnosis problems. *Complex & Intelligent Systems*, 7(2):703–723.

[2] Garg, H. and Nancy (2020). Algorithms for single-valued neutrosophic decision making based on TOPSIS and clustering methods with new distance measure. *AIMS Mathematics*, 5(3):2671–2693.

[3] Li, Y., Cai, Q., and Wei, G. (2023a). PT-TOPSIS methods for multi-attribute group decision making under single-valued neutrosophic sets. *Journal of Knowledge Engineering*, 26(2):122–135.

[4] Li, Y.-m., Khan, M., Khurshid, A., Gulistan, M., Rehman, A. U., Ali, M., Abdulla, S., and Farooque, A. A. (2023b). Designing pentapartitioned neutrosophic cubic set aggregation operator-based air pollution decision-making model. *Complex & Intelligent Systems*.

[5] Rehman, A. U., Gulistan, M., Ali, M., Al-Shamiri, M. M., and Abdulla, S. (2023). Development of neutrosophic cubic hesitant fuzzy exponential aggregation operators with application in environmental protection problems. *Scientific Reports*, 13:5484.

[6] Sajid, H. (2023). Global air pollution dataset. Kaggle. `https://www.kaggle.com/datasets/hasibalmuzdadid/global-air-pollution-dataset`. Accessed 2024.

[7] Smarandache, F. (1998). *Neutrosophy: Neutrosophic Probability, Set, and Logic*. American Research Press, Rehoboth, NM.

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

[9] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

[10] Zulqarnain, R. M., Siddique, I., Iampan, A., and Bonyah, E. (2021). Algorithms for multipolar interval-valued neutrosophic soft set with information measures to solve multicriteria decision-making problem. *Mathematical Problems in Engineering*, 2021:7211399.

DOI: https://doi.org/10.54216/NIF.050105
Received: December 21, 2024 Accepted: February 16, 2025