International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 6, June 2026)

# Enhancing Phishing Detection Through Double Refined Neutrosophic Aggregation and Ranking Techniques

Gokilamani R

*Associate Professor , Department of Mathematics ,Sri Ramakrishna College of Arts & Science for Women ,Coimbatorre*

***Abstract*— Phishing attacks remain one of the most persistent cybersecurity threats, exploiting human psychology and technical vulnerabilities to compromise sensitive information. Traditional detection methods often struggle with the inherent uncertainty, indeterminacy, and vagueness present in phishing indicators. This article introduces Double Refined Neutrosophic Probability with Refined Indeterminacy (DRNPRI) as a robust mathematical framework for phishing detection. We propose novel aggregation operators, score functions, accuracy functions, and ranking methods specifically designed for DRNPRI. Experimental validation on benchmark phishing datasets demonstrates the effectiveness of our approach in handling the complex, uncertain nature of phishing classification.**

***Keywords*—Aggregation Operator, Accuracy function, Decision Making, Double Refined Neutrosophic Numbers, Phishing, Ranking, Score function,**

## I. INTRODUCTION

Phishing attacks have evolved significantly, employing sophisticated techniques that blur the line between legitimate and malicious content. Conventional machine learning approaches treat phishing detection as a binary classification problem, often failing to capture the nuanced uncertainty inherent in borderline cases. A URL might exhibit some suspicious characteristics while simultaneously displaying legitimate traits—a scenario poorly modelled by crisp or even fuzzy logic systems.

Neutrosophic logic, introduced by Florentin Smarandache, extends fuzzy and intuitionistic fuzzy sets by explicitly modeling three independent components: truth-membership (T), indeterminacy-membership (I), and falsity-membership (F). However, standard neutrosophic sets treat indeterminacy as a monolithic component, which proves insufficient for complex real-world scenarios where indeterminacy arises from multiple distinct sources.

Double Refined Neutrosophic Numbers (DRNN) address this limitation by decomposing indeterminacy into two refined components:

- $I_1$: Indeterminacy due to insufficient information (epistemic uncertainty)
- $I_2$ : Indeterminacy due to conflicting information (aleatory uncertainty)

This refinement is particularly valuable for phishing detection, where uncertainty may stem from incomplete feature extraction ($I_1$)or contradictory signals within the same sample ($I_2$).

## II. MATHEMATICAL PRELIMINARIES

*2,1.Double Refined Neutrosophic Numbers*

*Definition 1*: A Double Refined Neutrosophic Number (DRNN) is represented as

$\alpha = \langle T, I_1, I_2, F \rangle$

where $T, I_1, I_2, F \in [0, 1]$ represent truth-membership, indeterminacy due to insufficient information, indeterminacy due to conflicting information, and falsity-membership respectively, satisfying $0 \leq T + I_1 + I_2 + F \leq 4$.

*Definition 2*: For DRNPRI (Double Refined Neutrosophic Probability with Refined Indeterminacy), we introduce probability distributions over each component, yielding

$\alpha_{DRNPRI} = \langle (T, p_T), (I_1, p_{I_1}), (I_2, p_{I_2}), (F, p_F) \rangle$

where $p_T , p_{I_1} , p_{I_2} , p_F \in [0,1]$ represent the probabilistic confidence in each membership degree.

*2.2. Basic Operations on DRNPRI*: For two DRNPRI numbers $\alpha = \langle T_\alpha, I_{1\alpha}, I_{2\alpha}, F_\alpha \rangle$ and $\beta = \langle T_\beta, I_{1\beta}, I_{2\beta}, F_\beta \rangle$

*Complement:* $\alpha^c = \langle F_\alpha, I_{2\alpha}, I_{1\alpha}\ T_\alpha \rangle$

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 6, June 2026)

*Union:* $\alpha \cup \beta = \left\langle \begin{matrix} maxi(T_\alpha, T_\beta), mini(I_{1\alpha}, I_{1\beta}), \\ mini(I_{2\alpha}, I_{2\beta}), mini(F_\alpha, F_\beta) \end{matrix} \right\rangle$

*Intersection:* $\alpha \cap \beta = \left\langle \begin{matrix} mini(T_\alpha, T_\beta), maxi(I_{1\alpha}, I_{1\beta}), \\ maxi(I_{2\alpha}, I_{2\beta}), maxi(F_\alpha, F_\beta) \end{matrix} \right\rangle$

## III. Novel Score and Accuracy Functions for DRNPRI

A critical requirement for multi-criteria decision-making is the ability to rank DRNPRI numbers. We propose three novel score functions and two accuracy functions.

### *3.1 Score Functions*

*Score Function 1 (Weighted Differential Score):*

$$S_1(\alpha) = \frac{1}{4}[3T - I_1 - 2I_2 - F + 2]$$

This function assigns greater penalty to conflicting indeterminacy ($I_2$) than to insufficient information ($I_1$), reflecting that contradictory evidence is more detrimental to decision confidence than missing data.

*Score Function 2 (Probabilistic Score):*

$$S_2(\alpha) = T \cdot (1 - I_1)(1 - I_2) - F \cdot (1 + I_1 + I_2)$$

This multiplicative formulation captures the intuition that high truth-membership is valuable only when indeterminacy is low, while falsity is amplified by uncertainty.

*Score Function 3 (Exponential Decay Score):*

$$S_3(\alpha) = T \cdot e^{-(I_1+I_2)} - F \cdot e^{(I_1+I_2)/2}$$

The exponential formulation provides sharper discrimination for samples with moderate indeterminacy levels.

### *3.2 Accuracy Functions*

When score functions yield identical values, accuracy functions serve as tie-breakers.

*Accuracy Function 1 (Certainty-Based):*

$$A_1(\alpha) = T + F - I_1 - I_2$$

Higher values indicate greater certainty, regardless of whether the sample is phishing or legitimate.

*Accuracy Function 2 (Refined Accuracy):*

$$A_2(\alpha) = \frac{T + F}{1 + I_1 + I_2}$$

This ratio-based formulation normalizes certainty by total indeterminacy.

## IV. Novel Aggregation Operators for DRNPRI

We introduce four aggregation operators designed for combining multiple DRNPRI assessments.

### *4.1 DRNPRI Weighted Arithmetic Mean (DRNPRI-WAM)*

*Definition 3* : Let $\alpha_1, \alpha_2, \ldots . \alpha_n$ be DRNPRI numbers with weight vector $w = (w_1, w_2, \ldots . w_n)^T$ where $\sum_{j=i}^{n} w_j = 1$. The DRNPRI-WAM operator is:

$$\text{DRNPRI- WAM}(\alpha_1, \alpha_2, \ldots . \alpha_n) = \left\langle \begin{matrix} 1 - \prod_{j=1}^{n} (1 - T_j)^{w_j}, \prod_{j=1}^{n} {I_1}_j{}^{w_j}, \\ \prod_{j=1}^{n} {I_2}_j{}^{w_j}, \prod_{j=1}^{n} F_j{}^{w_j} \end{matrix} \right\rangle$$

### *4.2 DRNPRI Weighted Geometric Mean (DRNPRI-WGM)*

*Definition 4*: The DRNPRI-WGM operator is defined as:

$$\text{DRNPRI-WGM}(\alpha_1, \alpha_2, \ldots . \alpha_n) = \left\langle \begin{matrix} \prod_{j=1}^{n} T_j{}^{w_j}, 1 - \prod_{j=1}^{n} (1 - {I_1}_j)^{w_j}, \\ 1 - \prod_{j=1}^{n} (1 - {I_2}_j)^{w_j}, 1 - \prod_{j=1}^{n} (1 - F_j)^{w_j} \end{matrix} \right\rangle$$

### *4.3 DRNPRI Ordered Weighted Averaging (DRNPRI-OWA)*

*Definition 5:* Let $\sigma = \{1,2, \ldots . n\} \to \{1,2, \ldots n\}$ be a permutation such that $S(\alpha_{\sigma(1)}) \geq S(\alpha_{\sigma(2)}) \geq \cdots . S(\alpha_{\sigma(n)})$.The DRNPRI-OWA operator with order weight vector $\omega = (\omega_1, \omega_2, \ldots \omega_n)^T$ is

$$\text{DRNPRI-OWA}(\alpha_1, \alpha_2, \ldots . \alpha_n) = \left\langle \begin{matrix} 1 - \prod_{j=1}^{n} (1 - T_{\sigma(j)})^{\omega_j}, \\ \prod_{j=1}^{n} {I_{1\sigma(j)}}^{\omega_j}, \prod_{j=1}^{n} {I_{2\sigma(j)}}^{\omega_j}, \prod_{j=1}^{n} {F_{\sigma(j)}}^{\omega_j} \end{matrix} \right\rangle$$

### *4.4.DRNPRI Hybrid Weighted Averaging (DRNPRI-HWA)*

*Definition 6:* The DRNPRI-HWA operator combines importance weights and position weights:

$$\text{DRNPRI-HWA}(\alpha_1, \alpha_2, \ldots . \alpha_n) = \text{DRNPRI-OWA}\ (\dot{\alpha_1}, \dot{\alpha_2}, \ldots \dot{\alpha_n})$$

where $\dot{\alpha_j} = n \cdot w_j \cdot \alpha_j$ is the weighted DRNPRI number, and the OWA aggregation uses order weights $\omega$.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 6, June 2026)

## V. Ranking Methods for DRNPRI

### 5.1.Lexicographic Ranking Method

*Algorithm 1*: *Lexicographic DRNPRI Ranking:*

Input: DRNPRI numbers $\alpha_1, \alpha_2, ..., \alpha_n$

Output: Ranked ordering

1. Compute $S_1(\alpha_i)$ for all i
2. If $S_1(\alpha_i) \neq S_1(\alpha_j)$, rank by $S_1$ (descending)
3. If $S_1(\alpha_i) = S_1(\alpha_j)$, compute $A_1(\alpha_i)$, $A_1(\alpha_j)$
4. If $A_1(\alpha_i) \neq A_1(\alpha_j)$, rank by $A_1$ (descending)
5. If $A_1(\alpha_i) = A_1(\alpha_j)$, rank by T (descending), then F (ascending)

### 5.2.Possibility Degree Ranking

*Definition 7:* The possibility degree that $\alpha \geq \beta$ is:

$P(\alpha \geq \beta) =$

$$\text{maxi}\left\{0, 1 - \max\left[0, \frac{S(\beta)-S(\alpha)}{|S(\beta)-A(\alpha)|+|S(\alpha)-A(\beta)|}\right]\right\}$$

This enables construction of a possibility degree matrix for complete ranking.

### 5.3 Outranking-Based DRNPRI Method

Inspired by ELECTRE, we define concordance and discordance indices:

*Concordance Index:*

$$C(\alpha, \beta) = \frac{w_T \cdot C_T + w_{I_1} \cdot C_{I_1} + w_{I_2} \cdot C_{I_2} + w_F \cdot C_F}{w_T + w_{I_1} + w_{I_2} + w_F}$$

where $C_T = 1 \; if \; T_\alpha \geq T_\beta$ else 0; similar definitions apply for other components with appropriate direction adjustments.

## VI. Application to Phishing Detection

### 6.1 DRNPRI Representation of Phishing Features

Phishing detection features can be naturally mapped to DRNPRI as follows:

| Feature Category | T | $I_1$ | $I_2$ | F |
|---|---|---|---|---|
| URL Analysis | Degree of suspicious patterns | Missing URL components | Conflicting URL signals | Degree of legitimate patterns |
| Content Analysis | Phishing content indicators | Unparseable content | Mixed legitimate/ malicious content | Safe content indicators |
| Domain Analysis | Malicious domain signals | WHOIS data unavailable | Domain reputation conflicts | Trusted domain signals |

### 6.2 Multi-Criteria Decision Framework

The phishing detection problem is formulated as multi-criteria decision-making where each URL/email is evaluated against multiple feature criteria $C_1, C_2, \ldots C_m$ with each criterion yielding a DRNPRI assessment.

*Algorithm 2: DRNPRI Phishing Classification*

Input: Sample x with feature vectors, criteria weights w, threshold θ

Output: Classification {Phishing, Legitimate, Uncertain}

1. For each criterion $C_j$:
   a. Extract features and compute DRNPRI $\alpha_j = \langle T_j, I_{1j}, I_{2j}, F_j \rangle$
2. Aggregate: $\alpha\_agg$ = DRNPRI-HWA ($\alpha_1, ..., \alpha_m$)
3. Compute $S_1(\alpha\_agg)$ and $A_1(\alpha\_agg)$
4. Classification:
   If $S_1(\alpha\_agg) > \theta$ and $A_1(\alpha\_agg) > 0$: return Phishing
   If $S_1(\alpha\_agg) < -\theta$ and $A_1(\alpha\_agg) > 0$: return Legitimate
   Else: return Uncertain

## VII. Experimental Evaluation

### 7.1 Datasets

We evaluated our approach on three benchmark phishing datasets:

| Dataset | Samples | Phishing | Legitimate | Features |
|---|---|---|---|---|
| UCI Phishing | 11,055 | 4,898 | 6,157 | 30 |
| PhishTank-2023 | 45,000 | 22,500 | 22,500 | 48 |
| ISCX-URL | 36,400 | 18,200 | 18,200 | 79 |

### 7.2 Feature-to-DRNPRI Mapping

Features were converted to DRNPRI numbers using domain knowledge and statistical analysis:

1. Numeric features: Normalized to [0,1] and mapped based on correlation with class labels
2. Categorical features: Encoded via frequency-based probability distributions
3. Missing values: Reflected in $I_1$ component
4. Conflicting features: Identified via correlation analysis and reflected in $I_2$

### 7.3 Experimental Setup

We compared five configurations:

- DRNPRI-WAM-S1: Weighted arithmetic mean with Score Function 1
- DRNPRI-WGM-S2: Weighted geometric mean with Score Function 2
- DRNPRI-HWA-S3: Hybrid weighted averaging with Score Function 3

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 6, June 2026)

- Baseline-NS: Standard neutrosophic set (single indeterminacy)
- Baseline-IFS: Intuitionistic fuzzy set approach

Ten-fold cross-validation was employed with criteria weights derived from entropy-based importance measures.

*7.4 Results*

Table 1: Classification Performance on UCI Phishing Dataset

| Method | Accuracy (%) | Precision | Recall | F1-Score | Uncertain Rate (%) |
|---|---|---|---|---|---|
| DRNPRI-WAM-S1 | 94.7 | 0.943 | 0.951 | 0.947 | 4.2 |
| DRNPRI-WGM-S2 | 93.8 | 0.932 | 0.945 | 0.938 | 5.1 |
| DRNPRI-HWA-S3 | 95.3 | 0.956 | 0.949 | 0.952 | 3.8 |
| Baseline-NS | 91.2 | 0.908 | 0.917 | 0.912 | 6.7 |
| Baseline-IFS | 89.6 | 0.891 | 0.902 | 0.896 | 8.3 |

Table 2: Performance Across All Datasets (DRNPRI-HWA-S3)

| Dataset | Accuracy (%) | F1-Score | AUC-ROC |
|---|---|---|---|
| UCI Phishing | 95.3 | 0.952 | 0.978 |
| PhishTank-2023 | 96.1 | 0.961 | 0.984 |
| ISCX-URL | 94.8 | 0.947 | 0.971 |

*7.5 Analysis of Results*

The DRNPRI framework consistently outperformed baseline approaches across all metrics. Key observations include:

1. *Refined indeterminacy improves discrimination:* The separation of $I_1$ and $I_2$ enables more nuanced handling of borderline cases. Standard neutrosophic sets, by conflating these sources, lose information critical for classification.
2. *Score Function 3 shows robust performance:* The exponential decay formulation proved most effective, particularly for samples with moderate uncertainty levels common in sophisticated phishing attempts.
3. *Uncertain classification reduces false positives:* By explicitly identifying uncertain samples (3.8% on UCI), the system avoids misclassifying legitimate samples as phishing—a costly error in production systems.
4. *DRNPRI-HWA balances multiple perspectives:* The hybrid aggregation operator's combination of importance and position weighting proved superior to pure arithmetic or geometric means.

*7.6 Ablation Study: Impact of Indeterminacy Refinement*

To isolate the contribution of double refinement, we conducted an ablation study:

| Configuration | F1-Score | Change |
|---|---|---|
| Full DRNPRI (T, $I_1$, $I_2$, F) | 0.952 | — |
| Merged indeterminacy (T, I, F) | 0.918 | -3.4% |
| Without $I_1$ | 0.931 | -2.1% |
| Without $I_2$ | 0.924 | -2.8% |

Removing either indeterminacy component degraded performance, with $I_2$ (conflicting information) proving slightly more important than $I_1$ (insufficient information).

## VIII. Theoretical Properties

*Theorem 1 (Idempotency):* If $\alpha_1 = \alpha_2 = \cdots . = \alpha_n = \alpha$ , then

DRNPRI-WAM$(\alpha_1, \alpha_2, \ldots . \alpha_n) = \alpha$

Proof. Follows directly from the properties of weighted means and the constraint $\sum w_j = 1$.

*Theorem 2 (Monotonicity:.* Let $\alpha_j \leq \alpha'_j$ for all j (component-wise). Then:

DRNPRI-WAM$(\alpha_1, \alpha_2, \ldots . \alpha_n) \leq$
DRNPRI-WAM$(\alpha'_1, \alpha'_2, \ldots . \alpha'_n)$

*Theorem 3 (Boundedness):* For any weight vector:

$\min_j \{\alpha_j\} \leq$ DRNPRI-WAM $(\alpha_1, \alpha_2, \ldots . \alpha_n) \leq \max_j \{\alpha_j\}$

These properties ensure the proposed operators behave consistently and predictably in decision-making contexts.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 6, June 2026)

## IX. Discussion and Future Directions

The DRNPRI framework addresses a fundamental limitation in uncertainty modelling for cybersecurity applications. By distinguishing between epistemic uncertainty (what we don't know) and aleatory uncertainty (inherent randomness or conflict), the framework provides richer semantics for classification under uncertainty.

*Practical implications*:

- Reduced analyst workload: Uncertain classifications can be routed to human analysts, optimizing resource allocation
- Improved model transparency: DRNPRI components offer interpretable explanations for decisions
- Adaptability: Weight parameters can be tuned for different deployment contexts (e.g., higher recall for financial institutions)

*Limitations and future work:*

1. Computational complexity: DRNPRI aggregation is more expensive than scalar methods; optimization for real-time detection remains ongoing
2. Feature engineering: Manual mapping of features to DRNPRI requires domain expertise; automated learning of DRNPRI representations is a promising direction
3. Adversarial robustness: Evaluation against adversarial phishing samples designed to maximize indeterminacy is needed
4. Extension to other domains: The framework is applicable to spam detection, malware classification, and intrusion detection

## X. Conclusion

This article introduced Double Refined Neutrosophic Probability with Refined Indeterminacy (DRNPRI) as a mathematical framework for phishing detection under uncertainty. We proposed novel score functions ($S_1, S_2, S_3$), accuracy functions ($A_1, A_2$), aggregation operators (DRNPRI-WAM, DRNPRI-WGM, DRNPRI-OWA, DRNPRI-HWA), and ranking methods specifically designed for DRNPRI numbers.

Experimental evaluation on three benchmark phishing datasets demonstrated that the DRNPRI framework outperforms standard neutrosophic and intuitionistic fuzzy approaches, achieving up to 96.1% accuracy with explicit handling of uncertain cases. The refinement of indeterminacy into epistemic and aleatory components proved particularly valuable for capturing the complex uncertainty patterns inherent in phishing detection.

The proposed methods provide a principled foundation for uncertainty-aware cybersecurity systems, with applications extending beyond phishing to broader threat detection and risk assessment domains.

*References*

[1] Smarandache, F. (1998). Neutrosophy: Neutrosophic Probability, Set, and Logic. American Research Press.

[2] Broumi, S., & Smarandache, F. (2014). New operations on interval neutrosophic sets. Journal of New Theory, 1, 24-37.

[3] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications, 25(2), 443-458.

[4] Dey, A., Broumi, S., & Bakali, A. (2019). A new approach to neutrosophic soft sets and their applications in decision making. Neutrosophic Sets and Systems, 25, 107-123.

[5] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications, 117, 345-357.

.