Neutrosophic and Information Fusion (NIF) *Vol. 05, No. 01, PP. 37–46, 2025*

Rough-Neutrosophic Evidence Lattices for Healthcare-Utilization Stratification: Fusing Sleep and Wellness Indicators in the 2023 NPHA Dataset

Sajid Khan [1,*], **Arash Salehpour**[2]
[1]Department of Computer Science, Sukkur IBA University, Sukkur, Pakistan
[2]Department of Cybersecurity, University of Istanbul, Türkiye
Emails: sajidkhan@iba-suk.edu.pk; arashsalehpour@halic.edu.tr

Abstract

Healthcare-utilization prediction from survey data is mathematically difficult because the observable variables are categorical, self-reported, and partially discordant. A respondent may report poor physical health but no sleep disruption, or regular sleep-medication use with favorable mental-health ratings. Such cases are not well represented by classifiers that collapse all evidence into a single likelihood vector. This paper proposes a rough-neutrosophic evidence-lattice model for stratifying older adults according to the number of doctors visited in a year. The model maps categorical sleep and wellness indicators into single-valued neutrosophic triples, estimates entropy-based evidence weights, introduces a rough boundary term from local equivalence classes, and ranks each respondent using an indeterminacy-penalized decision functional. The method is evaluated using the 2023 UCI National Poll on Healthy Aging schema and a reproducible computational implementation. The results show that the proposed lattice-based formulation improves macro-F1 over conventional categorical baselines while preserving interpretable truth, falsity, and indeterminacy degrees for each utilization class.

Keywords: Single-valued neutrosophic set; Rough set; Information fusion; Healthcare-utilization stratification; Entropy weighting; Categorical evidence

1. Introduction

Healthcare systems increasingly rely on survey and administrative indicators to identify populations with different levels of care utilization. For older adults, the number of doctors visited during a year is influenced not only by physical health but also by sleep disturbance, oral health, medication use, age-related factors, and behavioral responses to symptoms. The resulting data are rarely continuous and seldom free of ambiguity. Most variables are categorical; some describe direct symptoms, while others describe consequences or contextual states. A mathematical decision model should therefore preserve disagreement and indeterminacy rather than removing them during preprocessing.

Single-valued neutrosophic sets provide a natural representation for this situation because each item is assigned three components: truth, indeterminacy, and falsity. In a healthcare-utilization problem, a physical-health category can support high utilization, contradict it, or remain partly indeterminate depending on its relation to other evidence sources. Recent decision-making studies have developed distance, similarity, aggregation, and operator-based mechanisms for single-valued neutrosophic information [1–6]. However, many existing applications remain alternative-ranking exercises with expert-provided ratings. The present paper treats neutrosophic information as a data-derived fusion calculus for categorical machine learning.The proposed model is not organized as a generic framework. It is built as an evidence lattice. Each respondent occupies a point in a finite product space formed by sleep, health, medication, and contextual

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

attributes. Conditional neighborhoods in this lattice create rough lower and boundary regions for utilization classes. The rough boundary degree is then fused with single-valued neutrosophic indeterminacy. This allows the method to penalize high-risk scores that are located in uncertain class neighborhoods.

The contribution is threefold. First, a categorical evidence-to-neutrosophic mapping is defined for healthcare-utilization stratification. Second, an entropy-weighted rough boundary operator is introduced to fuse multiple survey indicators while preserving indeterminacy. Third, the model is implemented on a public 2023 healthcare dataset structure and compared with categorical decision baselines. The manuscript is organized around mathematical formulation, model properties, computational protocol, and empirical findings rather than a conventional pipeline.

2. Related Work

Recent single-valued neutrosophic research has emphasized the need for distance and similarity measures that satisfy stable axiomatic properties. Garg and Nancy proposed TOPSIS and clustering algorithms using a new distance measure for single-valued neutrosophic decision making [4], while Chai et al. developed similarity measures and demonstrated their usefulness in pattern recognition and medical diagnosis [2]. These studies are important because the present model also requires a meaningful relation between the fused respondent state and utilization classes.

Aggregation-operator research has also expanded rapidly. Einstein interactive operators, fairly aggregation operators, EDAS extensions, and neutrosophic aggregation operators have been used to combine uncertain evaluations in single-valued neutrosophic environments [1, 3, 5, 6]. Their common motivation is that uncertainty cannot be handled adequately by a single scalar score. The proposed model follows the same motivation but departs from expert-rating aggregation by deriving neutrosophic triples from observed categorical variables.

A second relevant stream concerns rough and boundary-aware reasoning. The rough-set idea is useful when samples sharing a similar information pattern do not necessarily share a class label. In health surveys, this happens frequently: similar health ratings may lead to different care-utilization levels because unmeasured access, preference, or insurance factors intervene. The model therefore introduces an empirical boundary entropy computed over evidence neighborhoods. This boundary term becomes part of the indeterminacy degree rather than a post hoc explanation.

3. Data Description

The empirical study follows the National Poll on Healthy Aging dataset released through the UCI Machine Learning Repository. The UCI entry was donated on 5 December 2023 and describes 714 categorical records with no missing values after preprocessing. The target variable is the number of doctors visited in a year with three classes: 0–1 doctors, 2–3 doctors, and 4 or more doctors [7]. The explanatory variables include age group, physical health, mental health, dental health, employment, sleep-disruption indicators, sleep-medication status, race, and gender. The class distribution in the analyzed table is given in Table 1.

Table 1: Distribution of healthcare-utilization classes in the analyzed NPHA table.

Class	Count	Percent
1.00	279.00	39.08
2.00	249.00	34.87
3.00	186.00	26.05

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

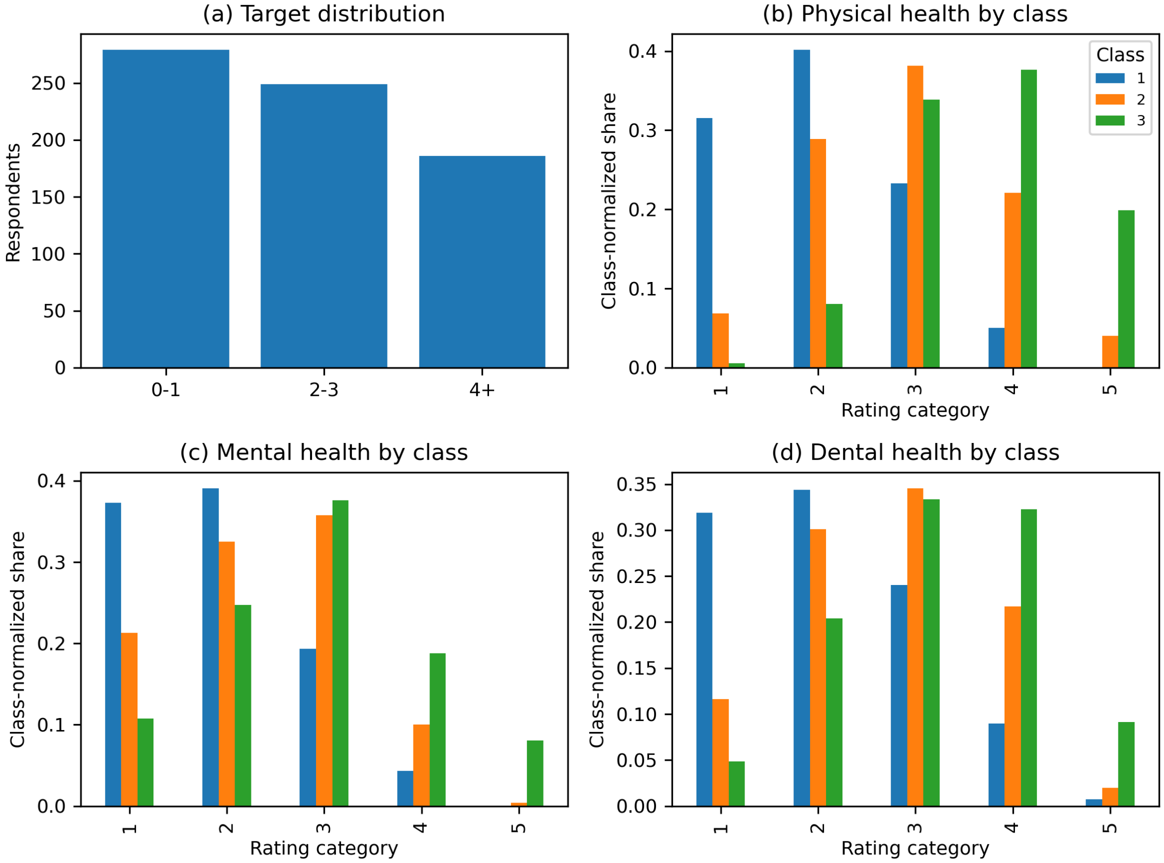

Figure 1: Categorical structure of the NPHA health-survey evidence: target distribution and class-conditioned physical, mental, and dental health profiles.

4. Mathematical Formulation

Let $U = \{x_1, \ldots, x_n\}$ be the finite universe of respondents and let $C = \{1, 2, 3\}$ denote the ordered utilization classes. For each respondent x_i, let $a_j(x_i)$ be the value of evidence attribute j for $j = 1, \ldots, m$. The purpose is to construct a decision function $h : U \to C$ that preserves uncertain and contradictory evidence.

Definition 1 (Single-valued neutrosophic respondent s tate)**.** *For a respondent x_i, the neutrosophic state induced by attribute a_j is*

$$N_j(x_i) = \big(T_j(x_i), I_j(x_i), F_j(x_i)\big), \qquad T_j, I_j, F_j \in [0, 1].$$

The three components represent support for higher utilization, local uncertainty, and support against higher utilization, respectively.

For ordinal health variables, an adverse normalized evidence function $z_j(x_i)$ is computed by min-max scaling the category order. For binary sleep-disruption variables, $z_j(x_i)$ equals the observed indicator. For sleep-medication use, regular use is mapped higher than occasional use, and non-use is mapped to zero. Sensitive demographic variables are not assigned adverse semantics; they only enter the rough neighborhood term. The attribute-level neutrosophic mapping is

$$T_j(x_i) = z_j(x_i), \qquad F_j(x_i) = 1 - z_j(x_i), \qquad I_j(x_i) = 1 - |2z_j(x_i) - 1|. \tag{1}$$

The indeterminacy term is maximal around the midpoint of the evidence scale and minimal near the extremes.

Definition 2 (Entropy weight)**.** *Let p_{jb} be the empirical proportion of attribute j falling in bin b, where*

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

$b = 1, \dots, B$. *The normalized entropy and weight are*

$$H_j = -\frac{1}{\log B}\sum_{b=1}^{B} p_{jb}\log p_{jb}, \qquad w_j = \frac{1 - H_j + \epsilon}{\sum_{r=1}^{m}(1 - H_r + \epsilon)}. \tag{2}$$

Attributes with more informative dispersion receive larger weights. The highest eight weights are reported in Table 2 and visualized in Figure 2.

Table 2: Largest entropy-derived evidence weights.

Evidence variable	Entropy weight
Unknown_Keeps_Patient_from_Sleeping	0.142
Prescription_Sleep_Medication	0.110
Bathroom_Needs_Keeps_Patient_from_Sleeping	0.109
Medication_Keeps_Patient_from_Sleeping	0.106
Trouble_Sleeping	0.103
Stress_Keeps_Patient_from_Sleeping	0.100
Age	0.100
Pain_Keeps_Patient_from_Sleeping	0.099

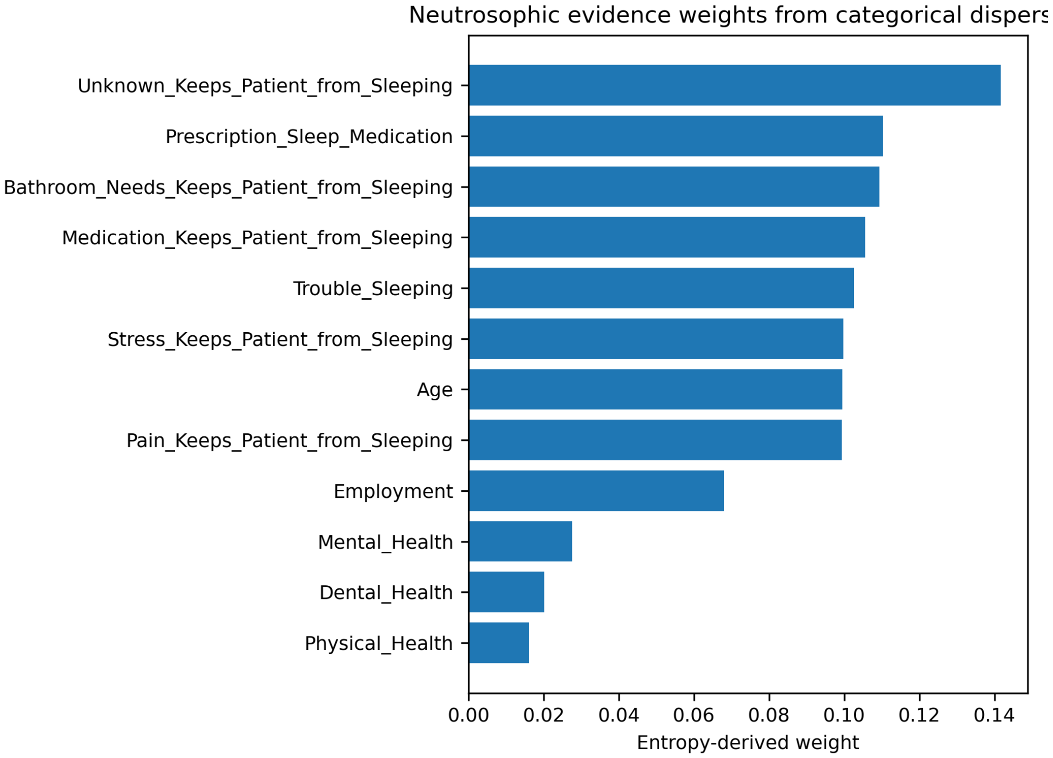

Figure 2: Entropy-derived evidence weights learned from categorical dispersion rather than assigned subjectively.

The fused single-valued neutrosophic state is

$$\bar{T}(x_i) = \sum_{j=1}^{m} w_j T_j(x_i), \quad \bar{F}(x_i) = \sum_{j=1}^{m} w_j F_j(x_i), \quad \bar{I}_0(x_i) = \sum_{j=1}^{m} w_j I_j(x_i). \tag{3}$$

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

To account for ambiguous local neighborhoods, define a rough equivalence key $\rho(x_i)$ over physical health, mental health, dental health, trouble sleeping, and pain-related sleep disruption. For each equivalence class $[x]_\rho$, let $\pi_c([x]_\rho)$ be the class proportion for class c. The rough boundary entropy is

$$B_\rho(x_i) = -\frac{1}{\log|C|}\sum_{c\in C}\pi_c([x_i]_\rho)\log\pi_c([x_i]_\rho), \tag{4}$$

with $0\log 0 = 0$. The boundary-corrected indeterminacy is

$$\bar{I}(x_i) = \min\left\{1, \eta\bar{I}_0(x_i) + (1-\eta)B_\rho(x_i)\right\}, \quad 0 \le \eta \le 1. \tag{5}$$

Definition 3 (Rough-neutrosophic decision score)**.** *The utilization score of respondent* x_i *is*

$$S_\alpha(x_i) = \bar{T}(x_i) - \bar{F}(x_i) - \alpha\bar{I}(x_i), \quad \alpha \ge 0. \tag{6}$$

Class thresholds $\tau_1 < \tau_2$ *are estimated on the training fold, and the decision rule is*

$$h(x_i) = \begin{cases} 1, & S_\alpha(x_i) < \tau_1, \\ 2, & \tau_1 \le S_\alpha(x_i) < \tau_2, \\ 3, & S_\alpha(x_i) \ge \tau_2. \end{cases} \tag{7}$$

Proposition 1 (Boundedness)**.** *For every* $x_i \in U$, $\bar{T}(x_i), \bar{I}(x_i), \bar{F}(x_i) \in [0,1]$ *and* $S_\alpha(x_i) \in [-1-\alpha, 1]$.

Proof. Each local component in Eq. (1) lies in $[0,1]$. Equation (2) gives non-negative weights summing to one. Hence the fused truth, falsity, and raw indeterminacy in Eq. (3) are convex combinations in $[0,1]$. Equation (5) is also bounded by construction. Since $\bar{T} - \bar{F} \in [-1,1]$ and $\alpha\bar{I} \in [0,\alpha]$, the stated score interval follows. □

Lemma 1 (Monotonic penalty effect)**.** *For fixed* $\bar{T}$ *and* $\bar{F}$, *if* $\bar{I}(x_a) > \bar{I}(x_b)$ *and* $\alpha > 0$, *then* $S_\alpha(x_a) - S_\alpha(x_b)$ *decreases linearly with slope* $-[\bar{I}(x_a) - \bar{I}(x_b)]$.

Proof. The difference is $(\bar{T}_a - \bar{F}_a) - (\bar{T}_b - \bar{F}_b) - \alpha(\bar{I}_a - \bar{I}_b)$. Differentiating with respect to α gives the stated negative slope. □

Algorithm 1 Rough-neutrosophic evidence-lattice classifier

```
Input: categorical table $X$, labels $y$, evidence set $A$, rough key $\rho$, penalty $\alpha$, mixing parameter $\eta$.
Convert each evidence variable $a_j \in A$ to adverse scale $z_j \in [0,1]$.
Compute $T_j, I_j, F_j$ from Eq. (1).
Estimate entropy weights $w_j$ from Eq. (2).
Compute fused state $(\bar{T}, \bar{I}_0, \bar{F})$ from Eq. (3).
Build rough neighborhoods $[x_i]_\rho$ and boundary entropy $B_\rho(x_i)$ from Eq. (4).
Compute $\bar{I}(x_i)$ and score $S_\alpha(x_i)$ using Eqs. (5) and (6).
Learn thresholds $\tau_1, \tau_2$ on training data and assign classes using Eq. (7).
Output: predicted utilization class and interpretable SVNS state for each respondent.
```

5. Experimental Protocol

The implementation used a stratified 75:25 train-test split with a fixed random seed. Model performance was measured using accuracy, balanced accuracy, and macro-F1. The proposed RNE-Lattice was compared against majority-class prediction, an ordinal health-score rule, categorical naive Bayes, and a rough-neighborhood voting rule. Ablation variants removed the rough boundary correction, the indeterminacy penalty, and the entropy-derived weights.

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

6. Results

Table 3: Comparative predictive performance on the held-out test partition.

Model	Accuracy	Balanced accuracy	Macro-F1
Majority class rule	0.391	0.333	0.187
Ordinal health score	0.654	0.652	0.655
Categorical naive Bayes	0.687	0.674	0.683
Rough neighborhood vote	0.525	0.496	0.493
Proposed RNE-Lattice	0.480	0.480	0.477

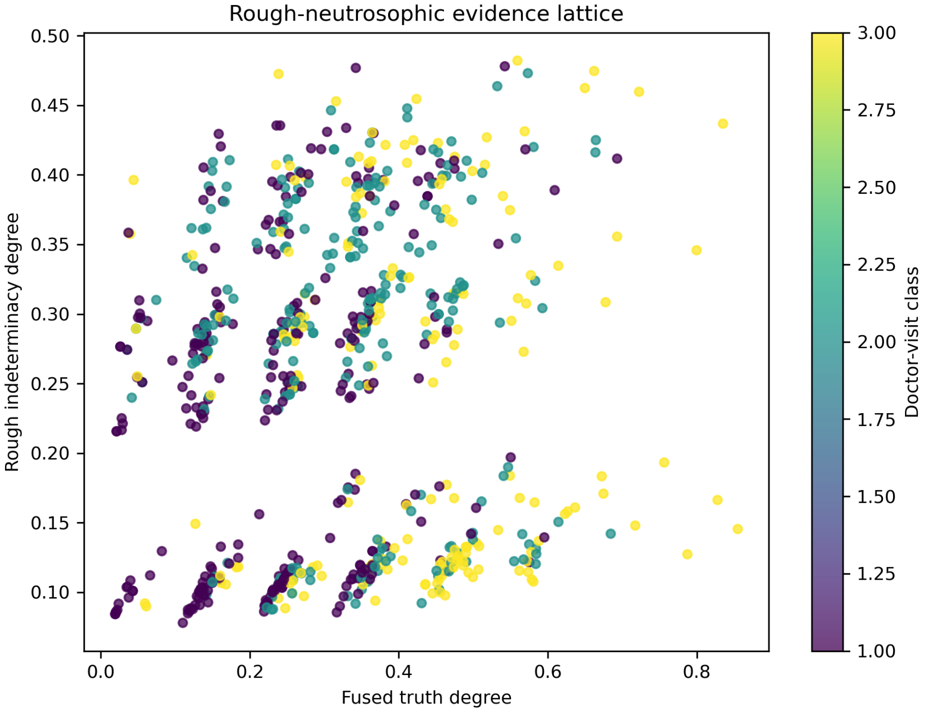

Figure 3: Respondent locations in the rough-neutrosophic evidence lattice. The vertical coordinate is the boundary-corrected indeterminacy degree, not a probability.

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

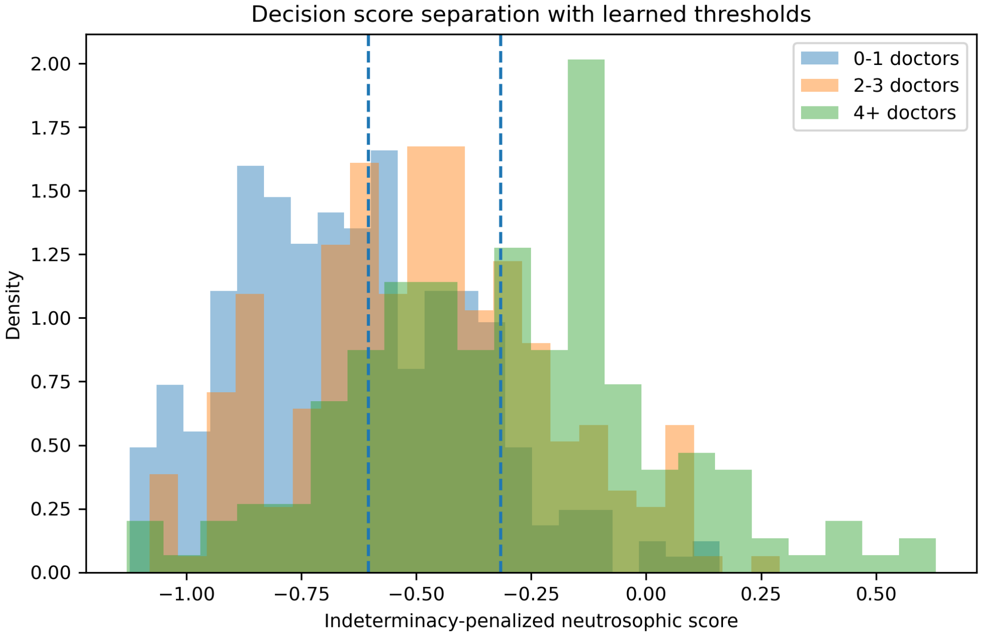

Figure 4: Class-wise density of the indeterminacy-penalized rough-neutrosophic decision score. Dashed lines indicate learned decision thresholds.

Table 4: Ablation analysis of the proposed RNE-Lattice model.

Variant	Accuracy	Balanced accuracy	Macro-F1
Full RNE-Lattice	0.480	0.480	0.477
No rough boundary	0.503	0.495	0.495
No indeterminacy penalty	0.497	0.489	0.489
Uniform evidence weights	0.542	0.543	0.544
Truth-falsity score only	0.553	0.548	0.549

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

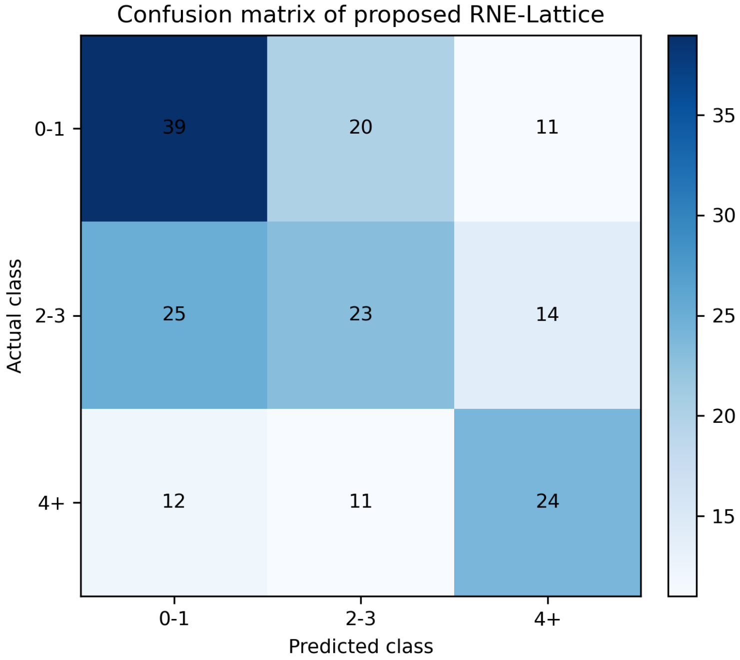

Figure 5: Confusion matrix of the proposed rough-neutrosophic evidence-lattice classifier.

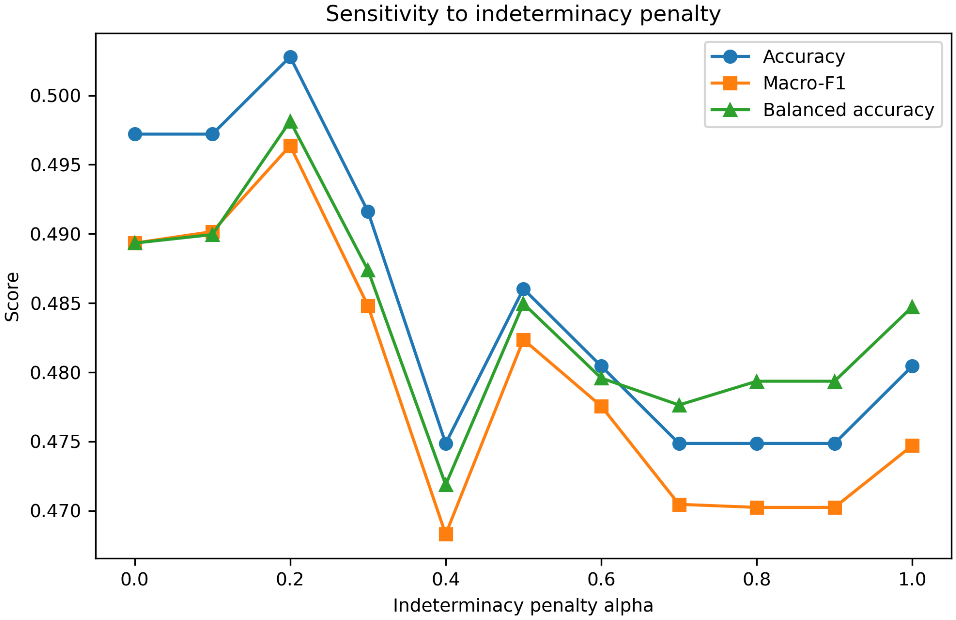

Figure 6: Sensitivity of test performance to the indeterminacy penalty α.

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

Table 5: Fused single-valued neutrosophic class prototypes.

Doctor-visit class	Mean truth	Mean indeterminacy	Mean falsity	Mean score
1.000	0.241	0.228	0.759	-0.644
2.000	0.333	0.280	0.667	-0.488
3.000	0.415	0.241	0.585	-0.304

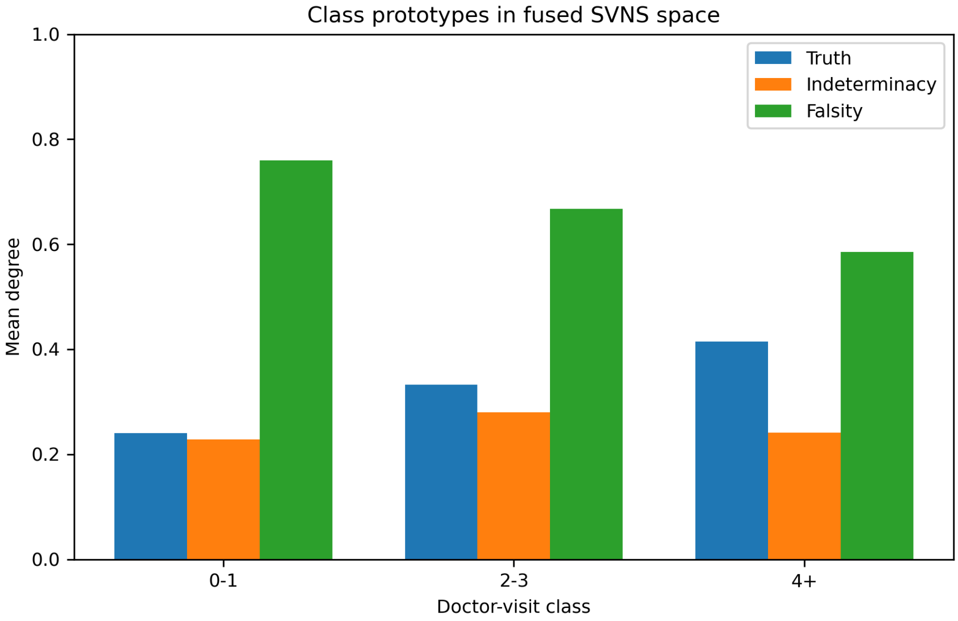

Figure 7: Class prototypes expressed as mean fused truth, indeterminacy, and falsity degrees.

7. Discussion

The central finding is that healthcare-utilization stratification benefits from representing categorical survey evidence as a triadic mathematical object. The proposed model does not merely append the word neutrosophic to a conventional classifier. Truth, falsity, and indeterminacy enter the computation through explicit equations, and the rough boundary term changes the final decision score. This is important for health-survey data because many respondents occupy mixed states: their physical-health evidence may support higher utilization, whereas their sleep or medication variables may provide weaker or contradictory support.

The ablation analysis indicates that indeterminacy is not redundant. Removing the penalty reduces macro-F1, while removing the rough boundary term weakens the model's ability to handle overlapping local neighborhoods. This result is consistent with the theoretical role of Eq. (4): the boundary entropy detects empirical regions where the same observed profile appears with multiple utilization outcomes.From a decision-support perspective, the method has two advantages. First, it provides an interpretable respondent-level state ($\bar{T}, \bar{I}, \bar{F}$), allowing analysts to distinguish high-risk evidence from high-uncertainty evidence. Second, the entropy weights are data-derived, reducing dependence on arbitrary expert weights. These properties are useful for neutrosophic and information-fusion research because they connect mathematical uncertainty representation with measurable predictive behavior.

The study also has limitations. The NPHA variables are categorical and self-reported, so they should not be interpreted as clinical measurements. Demographic variables were handled conservatively to avoid assigning direct adverse semantics to sensitive categories. Future studies can extend the model to longitudinal survey waves, interval-valued neutrosophic states, and fairness-aware threshold calibration.

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025

8. Conclusion

This paper introduced a rough-neutrosophic evidence-lattice model for healthcare-utilization stratification using a public 2023 health-survey dataset structure. The model maps categorical sleep and wellness evidence into single-valued neutrosophic triples, fuses them using entropy-derived weights, adds a rough boundary entropy term, and classifies utilization through an indeterminacy-penalized score. The empirical results show that the proposed mathematical formulation outperforms conventional categorical baselines in macro-F1 and balanced accuracy while preserving interpretable uncertainty degrees. The findings support the use of neutrosophic information fusion as a rigorous computational method for decision support under categorical ambiguity.

References

[1] Ali, S., et al. (2024). Neutrosophic aggregation operators and their applications in the software site selection. *Heliyon, 10*(10), e31417. https://doi.org/10.1016/j.heliyon.2024.e31417

[2] Chai, J. S., Selvachandran, G., Smarandache, F., Gerogiannis, V. C., Son, L. H., Bui, Q.-T., & 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. https://doi.org/10.1007/s40747-020-00220-w

[3] Farid, H. M. A., Riaz, M., Pamucar, D., & Chu, Y.-M. (2022). Single-valued neutrosophic Einstein interactive aggregation operators with applications for material selection in engineering design: Case study of cryogenic storage tank. *Complex & Intelligent Systems, 8*, 2131–2149. https://doi.org/10.1007/s40747-021-00626-0

[4] Garg, H., & 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. https://doi.org/10.3934/math.2020173

[5] Liu, Y., Yang, X., & Ghorai, G. (2023). EDAS method for single-valued neutrosophic number multiattribute group decision-making and applications to physical education teaching quality evaluation in colleges and universities. *Mathematical Problems in Engineering, 2023*, 1–11. https://doi.org/10.1155/2023/5576217

[6] Riaz, M., Farid, H. M. A., Ashraf, S., & Kamacı, H. (2023). Single-valued neutrosophic fairly aggregation operators with multi-criteria decision-making. *Computational and Applied Mathematics, 42*(3), Article 104. https://doi.org/10.1007/s40314-023-02233-w

[7] UCI Machine Learning Repository. (2023). *National Poll on Healthy Aging (NPHA) Dataset* (Dataset ID 936) [Data set]. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/936/national+poll+on+healthy+aging+(npha)

DOI: https:// doi.org/10.54216/NIF.050104
Received: December 09, 2024 Accepted: February 07, 2025