NSS

Neutrosophic Sets and Systems, Vol. 88, 2025

University of New Mexico

Lung Cancer Prediction Using an Enhanced Neutrosophic Set Combined with a Machine Learning Approach

Vakeel A. Khan[a,*], Asheesh Kumar Yadav[a], Mohammad Arshad[a,b] and Nadeem Akhtar[c]

[a]Department of Mathematics, Aligarh Muslim University, Aligarh–202002, India;

[b]Department of Computer Applications, Meerut Institute of Technology, Meerut–250103, India;

[c] Interdisciplinary Centre for Artificial Intelligence, Zakir Hussain College of Engineering and Technology, Aligarh, India

[*]Corresponding author: Vakeel A. Khan, E-mail: vakhanmaths@gmail.com

Abstract. Lung cancer (LC) remains one of the most lethal diseases globally, necessitating the development of advanced predictive models for early detection and accurate diagnosis. Traditional classification techniques often struggle with uncertainty and indeterminacy in medical data, which can lead to misdiagnosis and reduced diagnostic reliability. To address this issue, we propose an Enhanced Neutrosophic Set (ENS) framework integrated with machine learning algorithms to improve the prediction accuracy of lung cancer. Neutrosophic Set (NS) theory extends classical and fuzzy logic by introducing three independent membership components: truth, indeterminacy, and falsity, which enable more effective modeling of uncertainty in clinical datasets. The proposed ENS model enhances decision-making by optimizing feature selection and minimizing ambiguity in patient data representation. We apply machine learning classifiers including Logistic Regression (LR), K-Nearest Neighbors (KNN), and Random Forest (RF) to evaluate the performance of the ENS-transformed dataset in predicting lung cancer risk. Experimental results indicate that the ENS-based models outperform traditional approaches in terms of classification accuracy, sensitivity, and specificity. This study demonstrates the effectiveness of neutrosophic-based AI frameworks in medical diagnostics and highlights their potential in developing reliable, early detection systems for lung cancer and other critical diseases.

keywords: Lung Cancer Prediction, Enhanced Neutrosophic Set, Machine Learning, Medical Diagnosis, Uncertainty Modeling, Random Forest, Logistic Regression, K-Nearest Neighbors, Neutrosophic Logic, Clinical Data Analysis

1. Introduction

Lung cancer is one of the most critical health concerns worldwide, ranking as the second most commonly diagnosed cancer and the leading cause of cancer-related deaths [1,2]. Despite

significant improvements in public health, tobacco cessation programs, and medical interventions, lung cancer continues to account for millions of deaths annually. Smoking is the primary risk factor, responsible for 80% to 90% of cases, yet a significant portion of patients are non-smokers. If considered separately, lung cancer in non-smokers would rank as the 11^{th} most common cancer and the 7^{th} leading cause of cancer-related death globally [4].

Early diagnosis is key to improving survival outcomes. Technologies such as low-dose computed tomography (LDCT) have shown promise in detecting tumors at an early stage [3], and targeted therapies and immunotherapies have contributed to better disease management [8,20]. However, several challenges remain. Diagnosis often relies on imaging, biopsy, and clinical judgment, all of which are prone to uncertainties due to noise, image artifacts, and subjective interpretation. As a result, patients are often diagnosed at a later stage, which drastically reduces survival chances.

Traditional diagnostic tools and classical machine learning models often fail to handle the vagueness and ambiguity inherent in clinical data. Medical datasets commonly contain incomplete, imprecise, or inconsistent information, which affects classification accuracy. Standard fuzzy and intuitionistic fuzzy systems improve uncertainty handling to an extent but are limited in expressing higher-order indeterminacy [30].

To overcome such challenges, Neutrosophic Set Theory (NST), introduced by Smarandache [28, 29, 32, 41], offers a powerful extension of fuzzy and intuitionistic fuzzy set theories. It introduces three independent membership functions: truth (T), indeterminacy (I), and falsity (F), allowing for a more comprehensive modeling of uncertain and conflicting information. These properties make neutrosophic sets especially suitable for medical diagnosis, where decisions often involve ambiguity and incomplete knowledge [13, 14, 16, 42, 43].

In this study, we propose a novel Enhanced Neutrosophic Set (ENS)-based methodology combined with machine learning (ML) to improve lung cancer diagnosis. The originality of our approach lies in transforming the conventional lung cancer dataset into a neutrosophic dataset (N-dataset), where each attribute is expressed using the (T, I, F) structure. This transformation enables machine learning algorithms to learn from data more effectively by preserving and utilizing uncertainty rather than discarding it. To validate this method, we apply popular classifiers—Random Forest (RF) [27], Logistic Regression (LR), and K-Nearest Neighbors (KNN)—to both the original and transformed datasets.

The integration of Enhanced Neutrosophic Sets with machine learning addresses a critical need in current medical data science—managing imprecise, uncertain, and inconsistent clinical data. As lung cancer diagnosis becomes increasingly dependent on complex data (e.g., medical images, patient records), the ability to accurately model uncertainty is essential for developing

reliable diagnostic tools. The proposed ENS-based system offers improved classification performance, ultimately supporting early diagnosis, reducing misdiagnoses, and enhancing patient outcomes.

Key Contributions

The main contributions of this paper are as follows:

- We develop an Enhanced Neutrosophic Set (ENS) framework to transform traditional lung cancer datasets into uncertainty-aware representations (N-dataset).
- We integrate ENS with standard ML classifiers including Logistic Regression, Random Forest, and KNN, and evaluate their performance.
- We analyze and compare the results of classification on the original dataset versus the neutrosophic-transformed dataset.
- We demonstrate the superiority of the proposed approach in terms of accuracy, sensitivity, and specificity for early lung cancer prediction.

This study establishes a robust hybrid framework combining uncertainty modeling and machine learning, contributing to the advancement of intelligent diagnostic systems for lung cancer and potentially other medical applications.

2. **Methods and Experiments**

The following section presents a comprehensive overview of the materials and methods employed in the study.

2.1. *Proposed methodology*

The objective of this study is to enhance lung cancer prediction by integrating Neighborhood Selection (NS) with machine learning (ML) [35, 36] algorithms. The NCRP dataset (ID) was initially obtained from the Kaggle Machine Learning Repository [15, 37] and subsequently underwent careful preprocessing (ODpp) to ensure data quality and consistency.

$$I_D = LC \quad (1)$$

$$O_{Dpp} = f_{pp}(I_D) \quad (2)$$

The dataset was then converted into an N-representation (ON), in which each data point was defined not only by its original attributes but also by associated degrees of Truth (T), Indeterminacy (I), and Falsity (F). This method offers a more sophisticated representation of the uncertainty and variability typically present in medical datasets.

$$O_N = f_{T,I,F}(O_{Dpp}) \quad (3)$$

Subsequent to the transformation, the N-dataset ($O_{N_{(s)}}$) was partitioned into training and

testing subsets.

$$O_{N_s} = s(f_{O_N}) = s(X_{tain}, X_{test}, Y_{train}, Y_{test})$$

The N-dataset was normalized ($O_{N_{Nor}}$) to a range of 0 to 1 using a Min-Max scaler.

$O_{N_{Nor}} = f_{Nor}(O_{N_s})$ (5)

The normalized N-training dataset was utilized to train machine learning classifiers ($O_{N_{(ML)}}$), including K-nearest neighbors, Random Forest (RF), and Logistic Regression. These classifiers were chosen for their ability to manage complex feature interactions and detect subtle patterns, thereby supporting accurate lung cancer diagnosis.

$$O_{N_{ML}} = f_{ML}(O_{N_{(Nor)}}) = O_{ML}(DT_{O_{(Nor)}}, RF_{O_{(Nor)}}, AB_{O_{(Nor)}})$$

The primary performance metrics, denoted as $O_{N_{metrics}}$ including accuracy, precision, recall, and F1 score, were employed to evaluate the classifier's effectiveness and offer a comprehensive assessment of its predictive capabilities.

$$O_{N_{Metrics}} = O_{N_{(ML)}}(Metrics_{A_{cc}, P_r, R_c, F_1})$$

Finally, a comparative analysis (O_{CA}) was conducted between the N-representation dataset and the original dataset to evaluate the effect of integrating Neighborhood Selection (NS) on the accuracy and reliability of lung cancer prediction models

$$O_{CA} = CA(I_D; O_N)$$

The workflow is depicted in Figure 1.

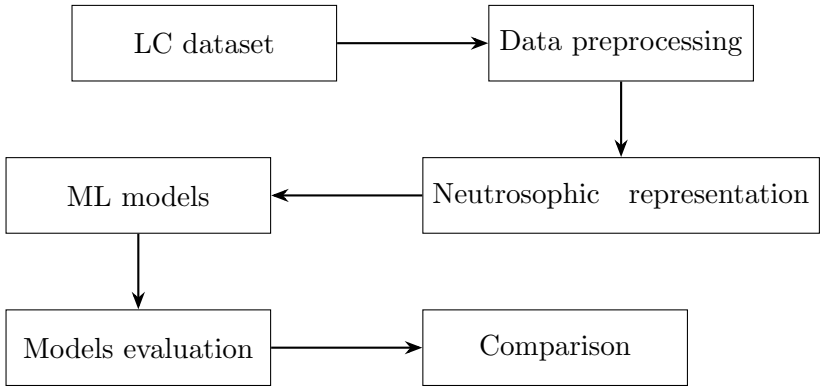

FIGURE 1. The flowchart of LC dataset prediction.

2.2. *Dataset description*

The Lung Cancer (LC) dataset, a widely utilized resource in machine learning and statistical analysis, was compiled by the University of Wisconsin–Madison. It serves as a valuable tool for lung cancer prediction and classification. Derived from digitized images of lung tumor samples, the dataset contains a range of attributes, including radius, texture, area, perimeter, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. It comprises 310 instances, each labeled as either malignant (indicating the presence of cancer) or benign (indicating its absence). This dataset provides a strong basis for developing accurate models that can effectively differentiate between malignant and benign lung cancer cases based on comprehensive tumor features.

2.3. *Preprocessing of data*

The lung cancer dataset underwent preprocessing to facilitate effective machine learning analysis. Initially, it was separated into feature attributes describing tumor characteristics and a categorical target variable indicating whether cases were benign or malignant. The categorical target was then encoded into numerical values to enhance model interpretability. To ensure consistency and improve prediction accuracy, all feature values were normalized using Min-Max scaling. This preprocessing step significantly enhanced the performance of the machine learning models in diagnosing lung cancer while maintaining data integrity.

2.4. *Neutrosophic Sets*

Let X be a universal set comprising various elements or entities, and let $x \in X$ be any arbitrary element. A **Neutrosophic Set (NS)** [18, 20–22, 24, 25, 28, 33, 40] is described by three separate membership functions assigned to each element $x \in X$:

- *Truth-membership function*: $T_A(x)$
- *Indeterminacy-membership function*: $I_A(x)$
- *Falsity-membership function*: $F_A(x)$

These functions are defined as mappings into the extended non-standard interval $\left]0^-, 1^+\right[$, specifically:

$$T_A : X \to \left]0^-, 1^+\right[, \quad I_A : X \to \left]0^-, 1^+\right[, \quad F_A : X \to \left]0^-, 1^+\right[$$

In this context, the values $T_A(x)$, $I_A(x)$, and $F_A(x)$ represent the respective degrees of truth, indeterminacy, and falsity associated with the element x. These values are **mutually independent**, and collectively satisfy the condition:

$$0 \leq T_A(x) + I_A(x) + F_A(x) \leq 3$$

This generalized structure allows Neutrosophic Sets to handle *incomplete*, *ambiguous*, and *conflicting* information more effectively than classical sets, fuzzy sets, or intuitionistic fuzzy sets. Such capability is particularly beneficial in practical applications involving uncertainty, such as medical diagnosis, multi-criteria decision-making, and pattern recognition.

Example: Medical Diagnosis with Conflicting Evidence

Consider a patient being evaluated for lung cancer. Based on several test results and expert opinions, the diagnosis is not straightforward:

- A senior radiologist strongly believes the CT scan indicates cancer, assigning a high truth-membership: $T_A(x) = 0.95$
- However, due to inconclusive biopsy and family history, there is considerable uncertainty: $I_A(x) = 0.40$
- Another specialist believes the symptoms could be due to a lung infection rather than cancer, assigning some falsity: $F_A(x) = 0.30$

Hence, the neutrosophic evaluation of the diagnosis is:

$$A(x) = \langle 0.95,\ 0.40,\ 0.30 \rangle$$

The sum:

$$T_A(x) + I_A(x) + F_A(x) = 0.95 + 0.40 + 0.30 = 1.65$$

This indicates the presence of overlapping evidence—high support for the cancer diagnosis, yet considerable uncertainty and partial opposition—something classical logic or fuzzy logic cannot effectively model. Neutrosophic Sets thus offer a powerful framework to handle such complex and contradictory scenarios.

2.5. *Neutrosophic Dataset Formation*

To effectively address the uncertainties associated with the Lung Cancer (LC) dataset in binary classification, a Neutrosophic Dataset (N-dataset) was introduced as a more generalized and inclusive representation [11, 31]. This formulation extends the traditional binary classification scheme by incorporating a neutral perspective in addition to the usual positive and negative categories. Each element in the dataset is expressed in the form $\langle T_A, I_A, F_A \rangle$, where the domain $X = (x_1, x_2, x_3, \ldots, x_n)$ consists of data points, and each $x \in X$ is represented as a triplet $x(t, i, f)$, with t, i, and f being real-valued degrees corresponding to truth, indeterminacy, and falsity, respectively.

To incorporate this structure, an additional neutrosophic component is integrated into the original dataset. The first step involves computing the mean vectors of the entire training set,

the positive class, and the negative class, denoted by ρ^{all}, ρ^{+}, and ρ^{-}, respectively. These are defined as follows:

$$\rho^{\text{all}} = \frac{1}{n}\sum_{k=1}^{n^{\text{all}}} x_k, \quad \rho^{+} = \frac{1}{n}\sum_{k=1}^{n^{+}} x_k, \quad \rho^{-} = \frac{1}{n}\sum_{k=1}^{n^{-}} x_k \tag{1}$$

Using the above mean vectors, the degrees of truth T, indeterminacy I, and falsity F for any given data point x are computed using the following expressions:

$$T = 1 - \frac{\|x - \rho^{+}\|}{\max\left(\|X_{\text{train}} - \rho^{+}\|\right)} \tag{2}$$

$$I = 1 - \frac{\|x - \rho^{\text{all}}\|}{\max\left(\|X_{\text{train}} - \rho^{\text{all}}\|\right)} \tag{3}$$

$$F = 1 - \frac{\|x - \rho^{-}\|}{\max\left(\|X_{\text{train}} - \rho^{-}\|\right)} \tag{4}$$

These computations are applied to each sample in both the training and testing datasets, thereby transforming them into neutrosophic representations. The resulting N-dataset incorporates semantic nuances by reflecting the degree to which a sample belongs to the positive class, is ambiguous, or belongs to the negative class. This enriched representation improves the robustness of machine learning classifiers in handling data that exhibits inherent ambiguity, inconsistency, or incompleteness. The transformation process ultimately leads to the construction of an N-representation dataset O_N, which is better suited for uncertain and complex biomedical applications.

2.6. *Classification Algorithms*

2.6.1. *Logistic Regression*

Logistic Regression is a widely used supervised learning technique designed for solving classification problems [26]. It estimates the probability that a given input instance falls into a particular class, typically in binary settings. Unlike linear regression, which predicts continuous outputs, logistic regression models the relationship between the dependent variable and one or more independent variables using a logistic function.

This method is particularly suited for binary classification tasks, where the response variable can assume only two possible outcomes. The core mathematical component of logistic regression is the sigmoid (or logistic) function, which transforms linear combinations of input features into values constrained between 0 and 1, representing probabilities. The sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{5}$$

In this equation, z represents the linear combination of input variables and their corresponding coefficients. The output of the sigmoid function is interpreted as the probability of the instance belonging to the positive class. Logistic regression does not produce class labels directly but assigns probabilities, which are then thresholded (commonly at 0.5) to assign class labels.

Instead of fitting a straight line, logistic regression fits an S-shaped curve, allowing it to effectively model classification boundaries between two classes, typically labeled as 0 and 1.

2.6.2. *Random Forest*

Random Forest (RF) [27] is an ensemble-based supervised learning algorithm that enhances prediction performance by constructing a collection of decision trees and aggregating their results. Each tree in the forest is trained on a bootstrap sample drawn randomly with replacement from the original training set. Furthermore, during the construction of each node, a random subset of features is considered for splitting, which introduces decorrelation among trees and enhances model robustness.

In classification tasks, the final prediction is determined through majority voting among the individual trees, whereas in regression problems, the average of all tree outputs is computed. This ensemble strategy reduces the model's variance and mitigates the risk of overfitting that is often encountered in single decision trees.

Random Forest excels in handling high-dimensional datasets and can effectively model complex, non-linear relationships. It is particularly suitable for domains involving noisy, incomplete, or heterogeneous data. One of its key strengths lies in its ability to estimate feature importance by evaluating the impact of each feature on the overall prediction accuracy, offering valuable insights into the underlying data structure.

Additionally, Random Forest is known for its scalability, robustness to outliers, and minimal requirement for parameter tuning. Due to these advantages, it has become a widely adopted approach in a variety of real-world applications, including bioinformatics, medical diagnosis, environmental modeling, and financial forecasting.

2.6.3. *K-Nearest Neighbors*

K-Nearest Neighbors (KNN) [14] is a non-parametric, instance-based learning algorithm commonly employed for both classification and regression tasks. The core principle of KNN involves identifying the k nearest data points in the training set to a given input sample, based

on a chosen distance metric (typically Euclidean distance), and inferring the output based on these neighbors. In classification, the algorithm assigns the most frequent class among the k neighbors, while in regression, it predicts the average of their target values.

KNN is categorized as a lazy learning algorithm because it does not involve an explicit training phase or model construction. Instead, it stores the training data and defers computation until a prediction is requested, at which point it performs distance-based retrieval and aggregation. This characteristic allows KNN to adapt flexibly to complex decision boundaries, as it makes no assumptions about the underlying data distribution.

Despite its conceptual simplicity, KNN is effective in a variety of practical applications, especially where the local structure of data plays a significant role. However, its performance is sensitive to the choice of k, the distance metric, and the presence of irrelevant or noisy features. To address such limitations, dimensionality reduction techniques and feature scaling are often employed in preprocessing.

2.7. *Performance Evaluation*

To comprehensively evaluate the effectiveness of the proposed classification approaches, four key performance metrics were employed: accuracy, precision, recall, and F1-score. These metrics offer a balanced perspective on the model's predictive ability, especially in binary classification tasks.

Accuracy measures the overall proportion of correctly predicted instances among the total number of samples, providing a general indication of the classifier's success.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Instances}}$$

Precision quantifies the ratio of correctly predicted positive instances to all instances predicted as positive, reflecting the classifier's exactness in labeling positive cases.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall, also referred to as sensitivity, captures the proportion of actual positive cases that were correctly identified, emphasizing the model's ability to detect relevant instances.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F1-score is the harmonic mean of precision and recall, offering a single metric that balances both aspects, particularly useful when the class distribution is imbalanced.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics together provide a robust framework for assessing the strengths and limitations of each classification model, ensuring a thorough and reliable performance analysis.

3. Results

3.1. *Experimental Setup*

In this study, three distinct tree-based machine learning classifiers [17] were utilized to predict lung cancer outcomes using the LC dataset. Prior to model training, the original dataset was transformed into an N-dataset format to better capture and manage inherent uncertainty within the data. Each classifier was trained separately on this N-dataset, and their predictive performances were evaluated using commonly accepted metrics. Furthermore, the outcomes achieved using the N-dataset were systematically compared against those obtained from the unmodified original dataset. All computational experiments were performed on the Google Colab platform, leveraging GPU acceleration to efficiently process the data and expedite model training.

3.2. *Experimental Results*

Table [1,2,3] summarizes the comparative performance of the employed machine learning algorithms on both the original and N-transformed datasets. Here, N-LR, N-RF, and N-KNN correspond to logistic regression, random forest, and K-nearest neighbors models trained on the N-dataset, respectively. This comparison highlights the impact of incorporating neutrosophic components on classification accuracy and robustness.

Table 1. Comparative evaluation of N-Logistic Regression Model and Logistic Regression Model

Metrics	**N-LR**	**LR**
Accuracy	**90.03**	87.5
Precision	**92.89**	91
Recall	**83.67**	79.5
F1 score	**82.73**	80.5

Table 2. Comparative evaluation of N-Random Forest Classifier and Random Forest Classifier

Metrics	**N-RF**	**RF**
Accuracy	**87.79**	85.71
Precision	**93.05**	90
Recall	**79.74**	76.5
F1 score	**82.04**	77

TABLE 3. Comparative evaluation of N-K Nearest Neighbors Model and K Nearest Neighbors Model

Metrics	N-KNN	KNN
Accuracy	**94.35**	92.86
Precision	**93.87**	92.5
Recall	**92.45**	89.5
F1 score	**93.67**	89.5

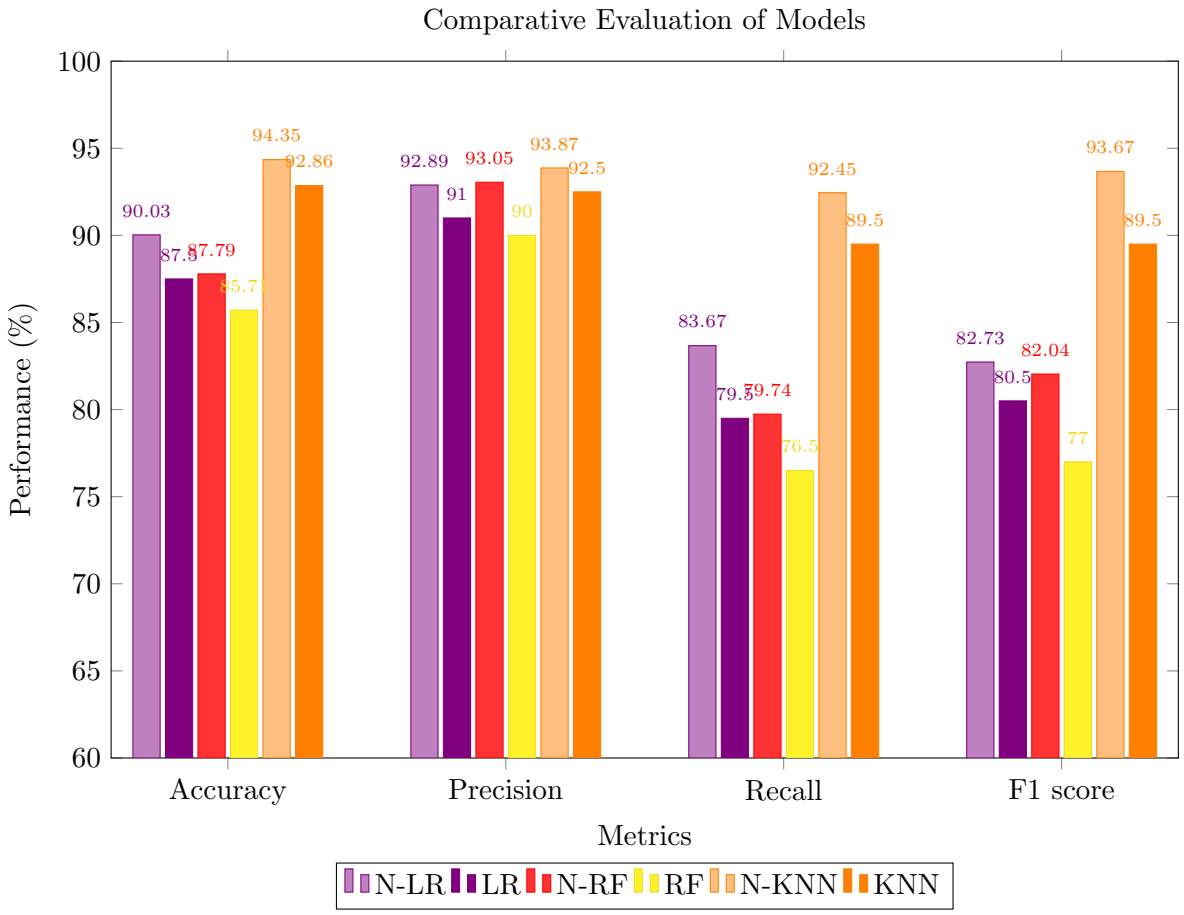

FIGURE 2. Classification of the following models on Accuracy, Precision, Recall, and F1 Score

4. Discussion

The proposed hybrid framework, which integrates Enhanced Neutrosophic Sets (ENS) with machine learning (ML) classifiers, demonstrates significant promise in advancing the early diagnosis of lung cancer. By modeling uncertainty, indeterminacy, and fuzziness through neutrosophic logic, the ENS framework addresses one of the key challenges in medical data analysis—handling imprecise and ambiguous information.

Advantages: One of the major strengths of this approach is its ability to enhance predictive accuracy and robustness across different ML classifiers. Among the tested models, the Neutrosophic K-Nearest Neighbors (N-KNN) consistently outperformed others, including Logistic Regression (LR) and Random Forest (RF), particularly in terms of accuracy and precision. This superior performance can be attributed to the neutrosophic representation, which effectively captures subtle variations and uncertainty present in clinical and imaging datasets. Additionally, the ENS framework refines the feature selection process by converting raw input into a structured format, thereby facilitating improved decision-making. The model's adaptability and extensibility make it suitable for other diagnostic applications where uncertainty is prevalent.

Limitations: Despite these advantages, the proposed approach has certain limitations. The transformation of data into neutrosophic space introduces computational complexity, which may hinder real-time deployment in resource-constrained clinical settings. Furthermore, the model's success is highly dependent on the completeness and quality of the dataset; missing or biased data may adversely impact its performance. Another challenge is the necessity of tuning the membership functions for truth, indeterminacy, and falsity, which often requires expert domain knowledge. Finally, while the experimental results are promising, the framework has not yet been validated on large-scale clinical datasets or through real-world hospital trials, which is essential before practical implementation.

In conclusion, the fusion of neutrosophic theory with machine learning techniques presents a promising direction for improving diagnostic accuracy in complex medical applications. However, future work should focus on optimizing computational efficiency, automating parameter selection, and validating the model in clinical environments to fully realize its potential.

5. Conclusion

This study introduces an advanced framework based on Enhanced Neutrosophic Sets (ENS) integrated with machine learning techniques to improve the accuracy and reliability of lung cancer prediction. By transforming the original lung cancer dataset into its neutrosophic representation [25, 29], the performance of multiple classifiers was significantly enhanced.

The experimental results, as presented in Tables 1, 2, and 3, clearly demonstrate that models trained on the neutrosophic dataset consistently outperform those trained on the conventional dataset. For instance, the N-KNN classifier achieved an impressive accuracy of 94.35% and an F1-score of 93.67, surpassing the traditional KNN model which recorded 92.86% accuracy and an F1-score of 89.5 (see Table 3). Similarly, the N-Logistic Regression and N-Random Forest models exhibited notable improvements in precision and recall, as shown in Tables 1 and 2.

These results highlight the strength of the ENS framework in handling uncertainty [16, 28, 30] and improving classification performance in medical data analysis. By incorporating the three membership components—truth, indeterminacy, and falsity—the ENS-based approach provides a more nuanced and comprehensive interpretation of patient information [32, 34], thereby enhancing the reliability of diagnostic decisions in lung cancer detection [12, 14].

Future research can explore the extension of this methodology to other areas of biomedical prediction, including early cancer screening and patient risk stratification, with a focus on real-time clinical applicability [11, 17].

5.1. *Acknowledgment:*

We thank the editor and referees for valuable comments and suggestions for improving the paper. The second authors would like to thank the Council of Scientific and Industrial Research ($CSIR$), Government of India for JRF(File N0. 09/0112(18967)/2024-EMR-I).

References

[1] Jett, J.R.; Cortese, D.A.; Fontana, R.S. Lung cancer: Current concepts and prospects. *CA* **1983**, *33*, 74–86.

[2] Davies, D.F. A review of detection methods for the early diagnosis of lung cancer. *J. Chronic Dis.* **1966**, *19*, 319–845.

[3] Russell, W.O.; Neihardt, H.W.; Mountain, C.F.; et al. Cytodiagnosis of lung cancer: A report of a four-year laboratory, clinical, and statistical study with a review of the literature on lung cancer and pulmonary cytology. *Acta Cytol.* **1963**, *7*, 1–44.

[4] Brett, G.Z. Earlier diagnosis and survival in lung cancer. *Br Med J* **1969**, *4*, 260–262.

[5] Lilienfeld, A.; Archer, P.G.; Burnett, L.H. *et al.* An evaluation of radiologic and cytologic screening for early detection of lung cancer: A cooperative pilot study of the American Cancer Society and Veterans Administration. *Cancer Res* **1966**, *26*, 2083–2121.

[6] Hulka, B.S. Screening for cancer: Lessons learned. *J Occup Med* **1986**, *28*, 687–691.

[7] Berlin, N.I.; Buncher, C.R.; Fontana, R.S. *et al.* The National Cancer Institute Cooperative Early Lung Cancer Detection Program: Results of the initial screen (prevalence). Introduction. *Am Rev Respir Dis* **1984**, *130*, 545–549.

[8] Fontana, R.S.; Sanderson, D.R.; Taylor, W.F. *et al.* Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic Study. *Am Rev Respir Dis* **1984**, *130*, 561–565.

[9] Frost, J.K.; Ball, W.C. Jr.; Levin, M.L. *et al.* Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study. *Am Rev Respir Dis* **1984**, *130*, 549–554.

[10] Flehinger, B.J.; McLamed, M.R.; Zaman, M.B. *et al.* Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Memorial Sloan-Kettering study in New York. *Am Rev Respir Dis* **1984**, *130*, 555–560.

[11] Amin, B.; Salama, A.A.; El-Henawy, I.M.; Mahfouz, K.; Gafar, M.G. Intelligent neutrosophic diagnostic system for cardiotocography data. *Comput. Intell. Neurosci.* **2021**, *2021*, 6656770.

[12] Jennifer, J.S.; Sharmila, T.S. A neutrosophic set approach on chest X-rays for automatic lung infection detection. *Inf. Technol. Control* **2023**, *52*, 37–52.

[13] Saqlain, M.; Garg, H.; Kumam, P.; Kumam, W. Uncertainty and decision-making with multi-polar interval-valued neutrosophic hypersoft set: A distance, similarity measure and machine learning approach. *Alex. Eng. J.* **2023**, *84*, 323–332.

[14] Akbulut, Y.; Sengur, A.; Guo, Y.; Smarandache, F. NS-k-NN: Neutrosophic set-based k-nearest neighbors classifier. *Symmetry* **2017**, *9*, 179.

[15] https://www.kaggle.com/code/hasibalmuzdadid/lung-cancer-analysis

[16] Nguyen, G.N.; Son, L.H.; Ashour, A.S.; Dey, N. A survey of the state-of-the-arts on neutrosophic sets in biomedical diagnoses. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 1–13.

[17] Ahmed, R.; Nasiri, F.; Zayed, T. A novel Neutrosophic-based machine learning approach for maintenance prioritization in healthcare facilities. *J. Build. Eng.* **2021**, *42*, 102480.

[18] Khan, V.A.; Arshad, M.; Khan, M.D. Some Results on Neutrosophic Normed Spaces via Tribonacci Convergent Sequence Spaces. *J. Inequal. Appl.* **2022**, *42*(1), 1–27.

[19] Khan, V.A.; Arshad, M. On Some Properties of Nörlund Ideal Convergence of Sequence in Neutrosophic Normed Spaces. *Ital. J. Pure Appl. Math.* **2023**, *50*, 352–373.

[20] Khan, V.A.; Arshad, M. On Neutrosophic Normed Spaces of I-Convergence Difference Sequences Defined by Modulus Function. *Neutrosophic Sets Syst.* **2024**, *73*, 178–190.

[21] Khan, V.A.; Arshad, M. Application of Neutrosophic Normed Spaces to Analyze the Convergence of Sequences Involving Neutrosophic Operators. *Math. Found. Comput.* **2025**.

[22] Fetanat, A.; Tayebi, M. Sustainability and reliability-based hydrogen technologies prioritization for decarbonization in the oil refining industry: A decision support system under single-valued neutrosophic set. *Int. J. Hydrog. Energy* **2024**, *52*, 765–786.

[23] Rogulj, K.; Kilić Pamuković, J.; Ivić, M. Hybrid MCDM based on VIKOR and cross entropy under rough neutrosophic set theory. *Mathematics* **2021**, *9*, 1334.

[24] Mondal, K.; Pramanik, S. Rough neutrosophic multi-attribute decision-making based on rough accuracy score function. *Neutrosophic Sets Syst.* **2015**, *8*, 14–21.

[25] Wang, H.; Smarandache, F.; Zhang, Y.; Sunderraman, R. Single valued neutrosophic sets. *Infinite Study* **2010**.

[26] Alshikho, M.; Jdid, M.; Broumi, S. Artificial Intelligence and Neutrosophic Machine Learning in the Diagnosis and Detection of COVID-19. *J. Prospect. Appl. Math. Data Anal.* **2023**, *1*, 2.

[27] Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

[28] Smarandache, F. (2019). Neutrosophic Set is a Generalization of Intuitionistic Fuzzy Set, Inconsistent Intuitionistic Fuzzy Set (Picture Fuzzy Set, Ternary Fuzzy Set), Pythagorean Fuzzy Set, Spherical Fuzzy Set, and q-Rung Orthopair Fuzzy Set, while Neutrosophication is a Generalization of Regret Theory, Grey System Theory, and Three-Ways Decision (revisited). *Journal of New Theory* **(2019)** *29*, 1-31.

[29] Smarandache, F., "A Geometric Interpretation of the Neutrosophic Set – A Generalization of the Intuitionistic Fuzzy Set," *2011 IEEE International Conference on Granular Computing*, edited by Tzung-Pei Hong et al., IEEE Computer Society, National University of Kaohsiung, Taiwan, pp. 602–606, November 8–10, 2011. `http://fs.unm.edu/IFS-generalized.pdf`

[30] Krassimir Atanassov and Peter Vassilev, "Intuitionistic fuzzy sets and other fuzzy sets extensions representable by them," *Journal of Intelligent & Fuzzy Systems*, IOS Press, 2019. DOI: `10.3233/JIFS-179426` (under press).

[31] Smarandache, F., "Neutrosophy, A New Branch of Philosophy," *Multiple-Valued Logic / An International Journal*, vol. 8, no. 3, pp. 297–384, 2002. `http://fs.unm.edu/Neutrosophy-A-New-Branch-of-Philosophy.pdf`

[32] Smarandache, F., *Neutrosophic Overset, Neutrosophic Underset, and Neutrosophic Offset. Similarly for Neutrosophic Over-/Under-/Off- Logic, Probability, and Statistics*, Pons Editions, Bruxelles, Belgique, 168 p., 2016. https://arxiv.org/ftp/arxiv/papers/1607/1607.00249.pdf

[33] Smarandache, F., "n-Valued Refined Neutrosophic Logic and Its Applications in Physics," *Progress in Physics*, vol. 4, pp. 143–146, 2013. https://arxiv.org/ftp/arxiv/papers/1407/1407.1041.pdf

[34] Smarandache, F., *Law of Included Multiple-Middle & Principle of Dynamic Neutrosophic Opposition*, EuropaNova & Educational, Brussels-Columbus (Belgium-USA), 136 p., 2014. http://fs.unm.edu/LawIncludedMultiple-Middle.pdf

[35] Gavrilova, Y., "Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials," 2020.

[36] Deepa, N., et al., "An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier," *The Journal of Supercomputing*, vol. 77, pp. 1998–2017, 2021.

[37] Salehi, A.W., et al., "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023.

[38] Chowdhury, S.R., Y. Khare, and S. Mazumdar, "Classification of diseases from CT images using LSTM-based CNN," in *Diagnostic Biomedical Signal and Image Processing Applications with Deep Learning Methods*, Elsevier, pp. 235–249, 2023.

[39] Özkaraca, O., et al., "Multiple brain tumor classification with dense CNN architecture using brain MRI images," *Life*, vol. 13, no. 2, p. 349, 2023.

[40] Şahin, R.; Çelik, Y. A Fermatean fuzzy multicriteria group decision making approach for medical diagnosis. *Soft Comput.* **2021**, *25*, 12267–12282.

[41] Wang, X.; Zhang, X.; Liu, S. A Pythagorean fuzzy decision support system for lung cancer prediction. *J. Intell. Fuzzy Syst.* **2018**, *35*, 345–352.

[42] N. Akhtar, N. Agarwal, and A. Burjwal, "K-mean algorithm for Image Segmentation using Neutrosophy," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Delhi, India, 2014, pp. 2417–2421.

[43] N. Akhtar and M. V. Ahmad, "A Modified Fuzzy C Means Clustering Using Neutrosophic Logic," in *Proceedings of the 2015 Fifth International Conference on Communication Systems and Network Technologies (CSNT)*, Gwalior, India, 2015, pp. 1124–1128.

Received: Dec. 15, 2024. Accepted: July 9, 2025