Neutrosophic Statistical Analysis on Healthcare Data for Oral Cancer Detection

Sakshi Taaresh Khanna[1,*], Sunil Kumar Khatri[2], Neeraj Kumar Sharma[3]

[1]Amity Institute of Information Technology, Amity University, Uttar Pradesh, Noida-201301, India

[2]Amity University, Uttar Pradesh, Noida-201301, India

[3]Department of Computer Science, Ram Lal Anand College, Benito Juarez Marg, New Delhi-110021, India

Emails: sakshitaareshkhanna@gmail.com; skkhatri@amity.edu; neerajksharma100@gmail.com

Abstract

Oral cancer is presently a growing health concern at the global level, with intense incidences of lifestyle factors. The increasing mortality rates of the diseased shall be controlled with effective early detection mechanisms. However, the traditional statistical approaches in practice fail to deliver in making a precise diagnosis of this cancer due to the intricate and interdependent prevalence of symptoms. This research work provides a solution approach using the potency of neutrosophic statistics in developing neutrosophic-integrated models of random forests and decision trees. Neutrosophic representation of data considering the indeterminacy, values of truth, and falsity facilitates healthcare experts in handling the conflicting patient data. The proposed random forest decision model with neutrosophic logic identifies the significant features, and the neutrosophic decision tree classifier predicts the stages of cancer. The findings are compared with conventional modelling of random forest and decision trees, and it demonstrates the efficiency and precision of neutrosophic statistical analysis in predicting oral cancer. This proposed neutrosophic decision framework will assist and support the medical practitioners and research experts in gaining more insights and deeper comprehension of the cancer progression and suggesting suitable treatment plans to minimize the morbidity rate.

Keywords: Oral cancer; Neutrosophic logic; Neutrosophic Random Forest Algorithm; Neutrosophic decision trees; Prediction

1. Introduction

The statistical report on oral cancer published by the American Society of Cancer is quite alarming and it calls for global awareness. The annual estimated new cases of 400,000 across the globe have made oral cancer stand as the 13th most common cancer worldwide [19]. Oral cancer otherwise referred to as mouth cancer largely affects the parts of the oral cavity and oropharynx. Lifestyle factors such as smoking, consumption of alcohol, prolonged exposure to radiation, and infection of the human papillomavirus are the primary causative factors of this disease [14]. The cancer-causing factors are intricate and it hurdles the detection of the disease prevalence. This delayed diagnosis leads to an increase in mortality rate and it draws more concerns in developing robust diagnostic methods. The traditional statistical methods that are generally used in cancer detection depend mostly on deterministic data with few instances of probability and it demand precise information. However, the healthcare data may not be definite at all times as it may be often uncertain, imprecise, inaccurate, and vague due to subjectivity and inconsistencies. These limitations demonstrate the inefficacy of the conventional prediction methods applied in oral cancer diagnosis. The shortcomings of these statistical methods present the necessity of developing improved prediction models based on neutrosophic logic.

DOI: https://doi.org/10.54216/IJNS.260323

Smarandache developed the theory of neutrosophic, which deals with truth, indeterminacy, and falsity representations of data. The neutrosophic sets are competent in handling incomplete healthcare information. The tri-logic-based statistical intervention assists healthcare experts in gaining more insights into the patient data, symptoms, and other factors contributing to the progression of the disease, resulting in more reliable estimates. Machine learning models are also closely associated with statistical principles, and evolving neutrosophic-based machine learning will provide robust prediction models. This research work presents novel decision-making approaches based on neutrosophic random forests and decision trees. Random forest is an ensemble method of learning that stems from multiple decision trees, resulting in more accurate and precise outcomes. This is more competent in handling the problems of overfitting and interpretability. On the other hand, the decision trees are primarily a supervised kind that works on breaking down of dataset into smaller units. The choice of these methods is ideal as these models are noted for their interpretability and competency in handling intricate, non-linear associations. The objective of developing these models is to enhance the efficacy of machine learning models in capturing uncertainty and indeterminate healthcare data. The incorporation of neutrosophic representation into the machine learning models of random forests facilitates identifying the significant features related to oral cancer. On the other hand, the neutrosophic-based decision tree classifier assists in predicting the stages of the cancer, enabling the medical experts to draw inferences on disease severity. The effectiveness of these neutrosophic models is exhibited in this research work by applying them to the healthcare data of oral cancer. Neutrosophic machine learning models are more capable of yielding precise predictions and these models are more efficient than conventional machine learning models dealing with deterministic data. This research work attempts to employ the potential of neutrosophic random forest classifiers and neutrosophic decision trees in dealing with complex medical data. The remaining contents of the paper are structured into the following sections. The state of the art of work related to the diagnosis of oral cancer, stating the research gaps and significant contributions, are presented in section 2. The preliminaries essential for the study is presented in section 3. The methodology of the neutrosophic machine learning model with an algorithmic approach is described in Section 4. The results are discussed in section 5, stating the efficacy of neutrosophic decision-making models over the conventional decision models. The final section concludes the research work, highlighting the extension of this work.

2. State of the Art of Work

This section presents a brief description of the related works of neutrosophic prediction modelling employed in cancer diagnosis, and the existing research gaps highlighting the need for this proposed neutrosophic-based machine-learning model in oral cancer detection. Abdel et al [1] discussed a neutrosophic framework for facilitating cancer patient management. Aslam et al [6] applied neutrosophic logic to evaluate the correlation between the mortality rate of prostate cancer and dietary fat consumption. Dutta et al [10] discussed the applications of neutrosophic cognitive graphs in determining the side effects of chemotherapy on cancer patients. For skin cancer detection, we utilized neutrosophic c-means and fuzzy c-means clustering algorithms. The aforementioned ideal studies were focused on breast cancer detection: Abdelhafeez et al [2] used neutrosophic c–c-means clustering and a newly neutrosophic-based approach for skin cancer classification based on deep learning features. Abdullah [3] applied neutrosophic sets with deep learning to classify breast cancer. Single-valued neutrosophic interval sets were used by Cui et al [8] to rate the risk levels of prostate cancer. Fu and Ye [11] employed cubic hesitant neutrosophic numbers in assessing the risk grades of prostate cancer patients. Chilusia et al [7] assessed the relevance of breast cancer rehabilitation based on neutrosophic linguistic scales and a neutrosophic logic-based support vector machine approach for enhanced skincare prediction. Neutrosophic Beta-Lindley distribution for modeling the detection of bladder cancer classification. Alzughaibi et al [4] applied rough neutrosophic sets together with a bat algorithm in oral cancer detection and classification. Devi et al [9] discussed the applications of neutrosophic graphs-based segmentation in cervical cancer detection and bladder cancer using neutrosophic inverse Gompertz distribution.

Angeline and Merlin [9] di5coursed on the threats of cancer caused by the modernization of society through a neutrosophic approach. Garrett et al applied neutrosophic-based SuperHyper graphs, SuperHyper cliques with usual and bipartite representations in cancer detection [13-16]. Gaber et al [12] explained the effectiveness of neutrosophic logic in prostate cancer treatment. Ashika et al [10] applied an enhanced kind of neutrosophic set-based machine learning in making predictions of breast cancer. Hammond et al [21] applied neutrosophic extended distribution modelling for bladder cancer patients and explored the implications of neutrosophic explainable AI and neutrosophic ANFIS machine learning in oral cancer detection. Madasi et al [15] applied a neutrosophic-based cubic hesitant decision framework in diagnosing and grading prostate cancer. Mallik et al [16] demonstrated the efficacy of neutrosophic logic in the effective treatment of cancer patients. Mustafa and Algamal [17] modeled the prediction of bladder cancer using the neutrosophic inverse power Lindley distribution. Rodriguez et al [18] evolved an integrated approach of machine learning with neutrosophic multi-criteria decisions in predicting colorectal cancer. Sayed and

DOI: https://doi.org/10.54216/IJNS.260323

Hassanien [20] applied moth flame optimization with neutrosophic sets in detecting breast cancer. Shaban [21] discussed the classification of breast cancer using neutrosophic-based deep learning techniques. Singh et al [22], Uma Maheswari, and Sumathi [23] also applied neutrosophic-based machine-learning approaches in skin and breast cancer detection, different types of neutrosophic [24] environment discussed by Broumi. The above-described literature is presented in tabular format as in Table 1 for a better understanding of the chronological development of neutrosophic modelling.

Table 1: Applications of Neutrosophic Models in Cancer Management

Authors & Year	Type of Cancer	Neutrosophic Methods Applied
Gaber et al [12]	Breast Cancer	Neutrosophic sets and fuzzy c-means algorithm
Dutta, A. K. [10]	General Side Effects of Cancer	Neutrosophic Cognitive Graphs (NCG)
Aslam, M., & Albassam, M. [6]	Prostate Cancer	Neutrosophic Logic for Dietary Fat Correlation
Abdel-Basset, M., & Mohamed, M. [1]	Cancer (General)	Neutrosophic Framework for Patient Assistance
Cui, W. H., Ye, J., & Fu, J. [8]	Prostate Cancer	Cotangent Similarity Measure of Neutrosophic Interval Sets
Fu, J., & Ye, J. [11]	Prostate Cancer	Similarity Measure with Neutrosophic Numbers for Risk Grade Assessment
Shaban, W. M. [21]	Breast Cancer	Neutrosophic Techniques, Deep Neural Network
Chiluisa Guacho, C. V., et al. [7]	Breast Cancer	Neutrosophic Linguistic Scale for Rehabilitation Program
Madasi, J. D., et al. [15]	Prostate Cancer	Neutrosophic Cubic Hesitant Fuzzy Decision Support System
Singh, S. K., et al. [17]	Skin Cancer	Neutrosophic Features, Deep Neural Network
Devi, M. A., et al. [9]	Cervical Cancer	Neutrosophic Graph Cut-Based Segmentation Scheme
Mallik, S., et al. [16]	General (Cancer Treatment)	Neutrosophic Logic Integration
Mustafa, M. Y., & Algamal, Z. Y. [17]	Bladder Cancer	Neutrosophic Inverse Power Lindley Distribution
Rodríguez, J. V., et al. [18]	Colorectal Cancer	Machine Learning, Neutrosophic MCDM Methodology
Garrett, H. [13]	Cancer Recognition	Neutrosophic Super Hyper Graph
Abdullah, W. [3]	Breast Cancer	Neutrosophic Sets, Deep Learning Models
Garrett, H. [16]	Cancer Recognition	Neutrosophic Super Hyper Graph

From the literature, the neutrosophic implications in modelling are well exhibited. The neutrosophic-based prediction models of cancer are more reliable and efficient in handling uncertain and complex data sets. With special reference to oral cancer, neutrosophic-based models are formulated with the integration of rough sets and explainable artificial intelligence. In addition, the features taken into consideration are not much explored. On the other hand, neutrosophic-based random forests and decision trees are not employed to the best of our knowledge. This has motivated the authors to develop neutrosophic integrated decision models to model oral cancer prediction. As the consideration of the features subjected to oral cancer is more intricate and interdependent, neutrosophic random forest classifiers are employed to decide on the significant factors, and neutrosophic-based decision trees are applied in the prediction of the cancer stages.

3. Preliminaries

This section presents the basic definition and other rudimentary concepts of neutrosophic statistics essential for this study. Neutrosophic statistics is characterized as an extension of classical statistics. In general, the data is expected to be deterministic in nature, and so crisp numbers are used in data representations. However, in the realm, the data may not be deterministic in nature due to the existence of uncertainty, indeterminacy, and imprecision. In such situations, neutrosophic numbers are used in data representations. Neutrosophic statistics serves as the basis for neutrosophic machine learning. The related definitions subjected to neutrosophic machine learning are presented as follows [7].

DOI: https://doi.org/10.54216/IJNS.260323

3.1 Neutrosophic set

Let U be a universal set of discourse. The neutrosophic set is of the form {(x, T(x), I(x), F(x)) :x∈X} where, T(x) is the truth membership function, I(x), is the indeterminacy function and F(x) is the falsity function. T(x), I(x), F(x) :X→ [0,1], where $0 \leq T(x) + I(x) + F(x) \leq 3$

3.2 Operations on Neutrosophic Sets

Let $\hat{A} = \{\langle x: T_A(x), I_A(x), F_A(x)\rangle \mid x \in X\}$ and $\hat{B} = \{\langle x: T_B(x), I_B(x), F_B(x)\rangle \mid x \in X\}$

(i) $\hat{A} \subseteq \hat{B}$ iff $T_A(x) \leq T_B(x), I_A(x) \geq I_B(x), F_A(x) \geq F_B(x)$

(ii) $\hat{A} = \hat{B}$ iff $T_A(x) = T_B(x), I_A(x) = I_B(x), F_A(x) = F_B(x)$

(iii) $\hat{A} \cap \hat{B} = \{\langle x: (T_A(x) \wedge T_B(x)), (I_A(x) \vee I_B(x)), (F_A(x) \vee F_B(x))\rangle \mid x \in X\}$

(iv) $\hat{A} \cup \hat{B} = \{\langle x: (T_A(x) \vee T_B(x)), (I_A(x) \wedge I_B(x)), (F_A(x) \wedge F_B(x))\rangle \mid x \in X\}$

(v) $\hat{A}^c = \{\langle x: F_A(x), 1 - I_A(x), T_A(x)\rangle \mid x \in X\}$

3.3 Single-valued neutrosophic number

Let $y_\beta, u_\beta, w_\beta \in [0,1]$ be any real numbers. A single valued neutrosophic number $\hat{\beta} = \langle([l_1, m_1, n_1, 0_1], y_\beta), ([l_2, m_2, n_2, o_2], u_\beta), ([l_3, m_3, n_3, o_3], w_\beta)\rangle$ on R with truth membership function $T_\beta: R \rightarrow [0, y_\beta]$, indeterminacy function $I_\beta: R \rightarrow [0, u_\beta]$ and falsity function $F_\beta: R \rightarrow [0, w_\beta]$.

$$T_\beta(x) = \begin{cases} f_{Tl}(x) & l_1 \leq x < m_1 \\ y_\beta & m_1 \leq x < n_1 \\ f_{Tr}(x) & n_1 \leq x < o_1 \\ 0 & otherwise \end{cases}$$

$$F_\beta(x) = \begin{cases} f_{Fl}(x) & l_3 \leq x < m_3 \\ w_\beta & m_3 \leq x < n_3 \\ f_{Fr}(x) & n_3 \leq x < o_3 \\ 0 & otherwise \end{cases}$$

3.4 Neutrosophic Statistic

A neutrosophic statistic is a random variable with a neutrosophic probability distribution centered on the neutrosophic sample.

3.5 Neutrosophic Feature Representation

Let us consider a sample x_i, the neutrosophic feature vector x_i^{NS} is of the form $x_i^{NS} = [x_i, T(x_i), I(x_i), F(x_i)]$. This representation facilitates machine-learning algorithms to integrate uncertainty.

3.6 Neutrosophic Distance Measure

Let us consider two neutrosophic instances such as x_i^{NS} and x_i^{NS}

$$D_{NS}(x_i^{NS}, x_j^{NS}) = \sqrt{\sum_{k=1}^{n} \left(w_T \left(T_i^k - T_j^k\right)^2 + w_I \left(I_i^k - I_j^k\right)^2 + w_F \left(F_i^k - F_j^k\right)^2 \right)}$$

where, w_T, w_I, w_F are the weights assigned for truth, indeterminacy and falsity.

3.7 Neutrosophic Classification Function

Let us consider a classifier f, the output is considered to be a neutrosophic label vector of the form $\hat{y}_i^{NS} = \langle T(\hat{y}_i), I(\hat{y}_i), F(\hat{y}_i)\rangle$. The ultimate decision is determined using a defuzzification rule of the form Class(x_i) = $arg \max_c \left(T_c(\hat{y}_i) - F_c(\hat{y}_i) - \beta I_c(\hat{y}_i)\right)$ where $\beta \in [0,1]$ is a penalty for indeterminacy.

DOI: https://doi.org/10.54216/IJNS.260323

3.8 Neutrosophic Loss Function

The loss function with neutrosophic representations is of the form

$$L_{NS} = \sum_{i=1}^{N} \left[\alpha.\left(T_i - \widehat{T_i}\right)^2 + \beta.\left(I_i - \widehat{I_i}\right)^2 + \gamma.\left(F_i - \widehat{F_i}\right)^2\right]$$

where α, β and γ are the trade-off parameters.

3.9 Neutrosophic Kernel Function

The kernel determines uncertainty during separation of classes, and it is of the form

$$K_{NS}(x_i, x_j) = \exp\left(\frac{-D_{NS}(x_i, x_j)^2}{2\sigma^2}\right)$$

3.10 Neutrosophic Decision Boundary

The classification is defined as

$$f_{NS}(x) = \sum_{i=1}^{N} \alpha_i y_i K_{NS}(x_i, x) + b$$

The final decision is determined using a threshold adjusted for indeterminacy

$$\text{Decision} = \begin{cases} +1 & \text{if } f_{NS}(x) > \theta_T \\ 0 & \text{if } \theta_F \leq f_{NS}(x) \leq \theta_T \\ -1 & \text{if } f_{NS}(x) < \theta_F \end{cases}$$

3.11 Neutrosophic Entropy for Feature Selection

The uncertainty in a feature is quantified using $H_{NS}(x_i) = -T(x_i)\log T(x_i) - I(x_i)\log I(x_i) - F(x_i)\log F(x_i)$. The higher value of entropy is an indicator of uncertainty in the feature. The above definitions are considered in general while handling neutrosophic machine learning models. Based on the nature of the machine-learning algorithm in particular, these definitions shall be modified accordingly.

4. Methodology of Neutrosophic Deaccessioning

This section presents the steps involved in neutrosophic-based machine-learning models of random forest classifiers and decision trees applied in the context of oral cancer detection. This decision procedure takes place in three phases, comprising the general problem definition in the first phase, choosing the significant features in the second phase, and predicting the stages of the cancer in the third phase.

Phase I Problem Definition

Step 1: Formulation of the decision-making problem

In this phase, the decision problem is well defined with the choice of the possible features, subject to the diverse factors. The factors considered for the study and the features subjected to each of the factors are presented in the following Table 2.

Table 2: Factors and Features

Factors	Features
Demographic	Age
	Gender
	Geographic location
Lifestyle and Behavioral	Smoking history (frequency, duration)
	Alcohol consumption (frequency, amount)
	Chewing tobacco or betel nut use
Medical History	Family history of cancer
	History of previous cancers or other significant health conditions
Clinical	Tumor size (measured or estimated)

DOI: https://doi.org/10.54216/IJNS.260323

	Tumor grade (differentiation level)
	Tumor stage (stage at diagnosis)
	Lymph node involvement (extent and location)
Pathological	Tumor type (histological classification)
	Biomarker levels (e.g., p53, Ki-67, HPV status)
	Genetic mutations (if available)
Symptoms and Physical Findings	Pain intensity
	Presence of lesions or ulcers
	Swelling or mass in the oral cavity
	Dysphagia (difficulty swallowing)
Treatment-Related	Type of treatment (surgery, radiation, chemotherapy)
	Treatment response (e.g., tumor reduction percentage)
	Side effects experienced
Follow-Up and Outcome	Recurrence status (yes/no, location if applicable)
	Survival time (post-diagnosis)
	Quality of life measures post-treatment
Environmental and Socioeconomic	Exposure to environmental pollutants or occupational hazards
	Access to healthcare facilities
	Socioeconomic status (income, education level)
Immunological and Health Conditions	Overall immune status
	Presence of comorbidities (e.g., diabetes, cardiovascular disease)

The data collected and subjected to each of the features are presented using neutrosophic representations comprising the values of truth, indeterminacy, and falsity. The data expressed as neutrosophic sets are of the form $x_i = (T_i, I_i, F_i)$ where $T_i, I_i, F_i \in [0,1]$ and $0 \leq T_i + I_i + F_i \leq 3$. The data is preprocessed and prepared for the next phase of training with the removal of redundancies.

Phase II: Neutrosophic Random Forest for Criterion Reduction

Step 2: Construction of Neutrosophic Multiple Decision Trees

In this phase, multiple decision trees with neutrosophic representations are constructed and trained on various bootstrap samples of the dataset. Each of the neutrosophic decision trees defined is ensembled on the arbitrary subset of the samples and the features.

Let N signify the count of trees in the forest, for each t belonging to the set {1,… N} representing the tree, a sample is drawn with replacement from the training datasets. At each of the nodes, an arbitrary feature subset is selected for splitting. The neutrosophic decision tree is trained using entropy-based criteria.

Step 3: Aggregation of Prediction from Neutrosophic Trees

In a neutrosophic-based aggregation, the aggregation is based on the truth-values. For each of the test samples, say x, the prediction vector in neutrosophic form is (T_t, I_t, F_t), the aggregation prediction is of the form

$$T_{avg} = \frac{1}{N}\left(\sum_{t=1}^{N} T_t\right), I_{avg} = \frac{1}{N}\left(\sum_{t=1}^{N} I_t\right), F_{avg} = \frac{1}{N}\left(\sum_{t=1}^{N} F_t\right)$$

The highest aggregated truth-value say T_{avg} decides the ultimate prediction for x in the class.

Phase III Neutrosophic Decision Trees for Cancer Prediction

Step 4: Criteria Splitting

The neutrosophic entropy measure is applied to split the criteria. At a node, let us consider a subset S of training samples with p_c as the proportion of class c in S. The neutrosophic entropy H(S) is defined at each of the nodes.

$$H(S) = -\sum_{c=1}^{C} p_c.(T_c log T_c + I_c log I_c + F_c log F_c)$$

DOI: https://doi.org/10.54216/IJNS.260323

where T_c, I_c, F_c are associated with the truth, indeterminate, and falsity values pertinent to each class.

Step 5: Framing of Neutrosophic Tree

The neutrosophic entropy value for each of the features x_i after splitting is computed. The features are chosen and the threshold values are decided to split the data. This process is made recursive until the stopping feature or the criterion is arrived.

Step 6: Neutrosophic Prediction

For a given sample say x which is denoted in the neutrosophic form (T, I, F) is propagated through the decision tree in the neutrosophic form to predict the classes. The thresholds at each instance are compared during the training of the datasets. The prediction of the cancer stage is arrived when the leaf node attains the highest truth membership value. The performance measures of the neutrosophic random forest say neutrosophic accuracy, precision-recall, and F1 score are calculated to determine the efficacy of the random forest in determining the significant features of the proposed decision. The above-described three-phase methodology is presented pictorially in Fig.1.

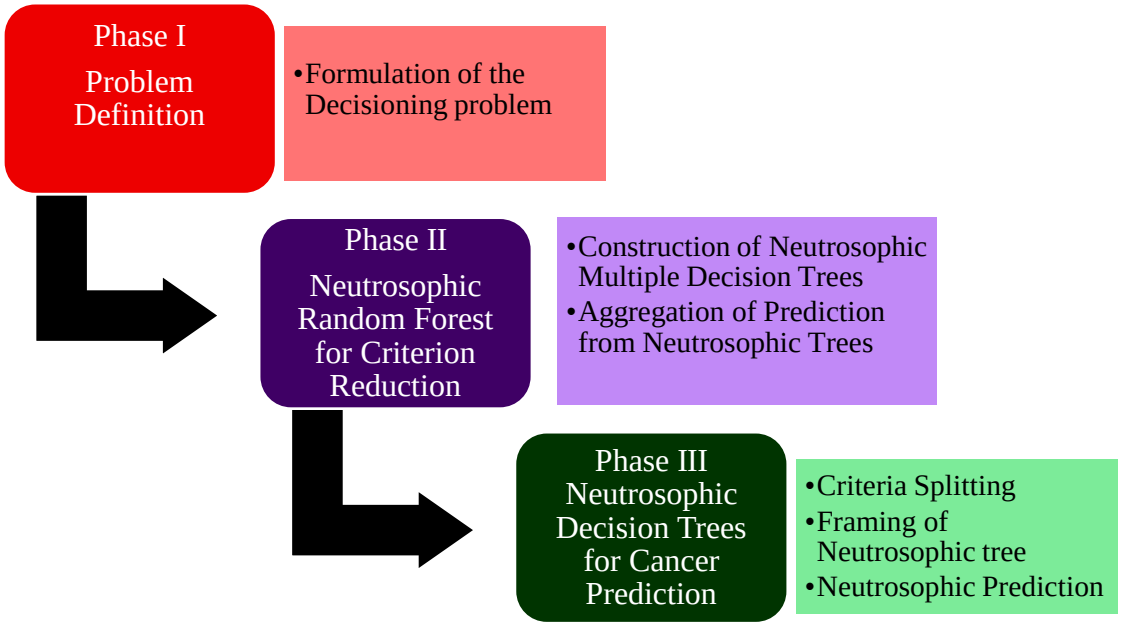

Figure 1. Neutrosophic Deaccessioning Framework

5. Results and Discussion

This section presents the results obtained by using the neutrosophic methodology described in Section 3. Python libraries of neutrosophic random forest classifiers are used to determine the significant features of oral cancer detection. The Table contains the most noteworthy features to consider in predicting oral cancer. The neutrosophic labels of these features are also presented in Table 3.

Table 3: Neutrosophic Representation of Features

Notation	Features	Linguistic Variables	Neutrosophic Quantifications
F1	Age	Young, Middle-aged, Elderly	Young (0.7, 0.2, 0.1), Middle-aged (0.6, 0.3, 0.1), Elderly (0.5, 0.2, 0.3)
F2	Geographic Location	Urban, Rural	Urban (0.8, 0.1, 0.1), Rural (0.7, 0.2, 0.1)
F3	Smoking History	Non-smoker, Occasional, Heavy smoker	Non-smokers (0.9, 0.05, 0.05), Occasional (0.6, 0.2, 0.2), Heavy smokers (0.4, 0.3, 0.3)

DOI: https://doi.org/10.54216/IJNS.260323

F4	Alcohol Consumption	None, Moderate, Heavy	None (0.9, 0.05, 0.05), Moderate (0.6, 0.2, 0.2), Heavy (0.4, 0.3, 0.3)
F5	Family History of Cancer	None, Distant Relative, Immediate Family	None (0.9, 0.05, 0.05), Distant Relative (0.5, 0.3, 0.2), Immediate Family (0.3, 0.4, 0.3)
F6	History of Previous Cancers	None, Treated, Active	None (0.9, 0.05, 0.05), Treated (0.6, 0.2, 0.2), Active (0.3, 0.3, 0.4)
F7	Tumor Size	Small, Medium, Large	Small (0.8, 0.1, 0.1), Medium (0.6, 0.3, 0.1), Large (0.4, 0.3, 0.3)
F8	Tumor Stage	Primary, Secondary, Tertiary	Primary (0.9, 0.05, 0.05), Secondary (0.5, 0.3, 0.2), Tertiary (0.3, 0.4, 0.3)
F9	Tumor Type	Benign, Pre-malignant, Malignant	Benign (0.8, 0.1, 0.1), pre-malignant (0.5, 0.3, 0.2), Malignant (0.3, 0.3, 0.4)
F10	Biomarker Levels	Low, Normal, Elevated	Low (0.9, 0.05, 0.05), Normal (0.7, 0.2, 0.1), Elevated (0.4, 0.3, 0.3)
F11	Pain Intensity	None, Mild, Moderate, Severe	None (0.9, 0.05, 0.05), Mild (0.7, 0.2, 0.1), Moderate (0.5, 0.3, 0.2), Severe (0.3, 0.4, 0.3)
F12	Presence of Lesions/Ulcers	None, Mild, Extensive	None (0.9, 0.05, 0.05), Mild (0.6, 0.3, 0.1), Extensive (0.3, 0.4, 0.3)
F13	Type of Treatment	Observation, Radiation, Surgery, Combined	Observation (0.7, 0.2, 0.1), Radiation (0.5, 0.3, 0.2), Surgery (0.6, 0.3, 0.1), Combined (0.4, 0.4, 0.2)
F14	Treatment Response	Positive, Stable, Negative	Positive (0.8, 0.1, 0.1), Stable (0.5, 0.3, 0.2), Negative (0.3, 0.4, 0.3)
F15	Recurrence Status	None, Recurrence, Multiple Recurrences	None (0.9, 0.05, 0.05), Recurrence (0.5, 0.3, 0.2), Multiple Recurrences (0.3, 0.4, 0.3)
F16	Survival Time	Short-term, Medium-term, Long-term	Short-term (0.3, 0.4, 0.3), Medium-term (0.5, 0.3, 0.2), Long-term (0.8, 0.1, 0.1)
F17	Exposure to Environmental Pollutants	None, Low, High	None (0.9, 0.05, 0.05), Low (0.7, 0.2, 0.1), High (0.3, 0.4, 0.3)
F18	Access to Healthcare Facilities	Excellent, Moderate, Poor	Excellent (0.8, 0.1, 0.1), Moderate (0.6, 0.3, 0.1), Poor (0.4, 0.3, 0.3)
F19	Overall Immune Status	Weak, Normal, Strong	Weak (0.4, 0.3, 0.3), Normal (0.7, 0.2, 0.1), Strong (0.8, 0.1, 0.1)
F20	Presence of Comorbidities	None, Mild, Severe	None (0.9, 0.05, 0.05), Mild (0.6, 0.3, 0.1), Severe (0.3, 0.4, 0.3)

DOI: https://doi.org/10.54216/IJNS.260323

The above 20 features are considered as input and the output feature is the stage of cancer and it is classified into four levels based on the increasing intensities namely stage 1, stage 2, stage 3 and stage 4. The association between the input features is represented using heat maps as in Fig. 2.

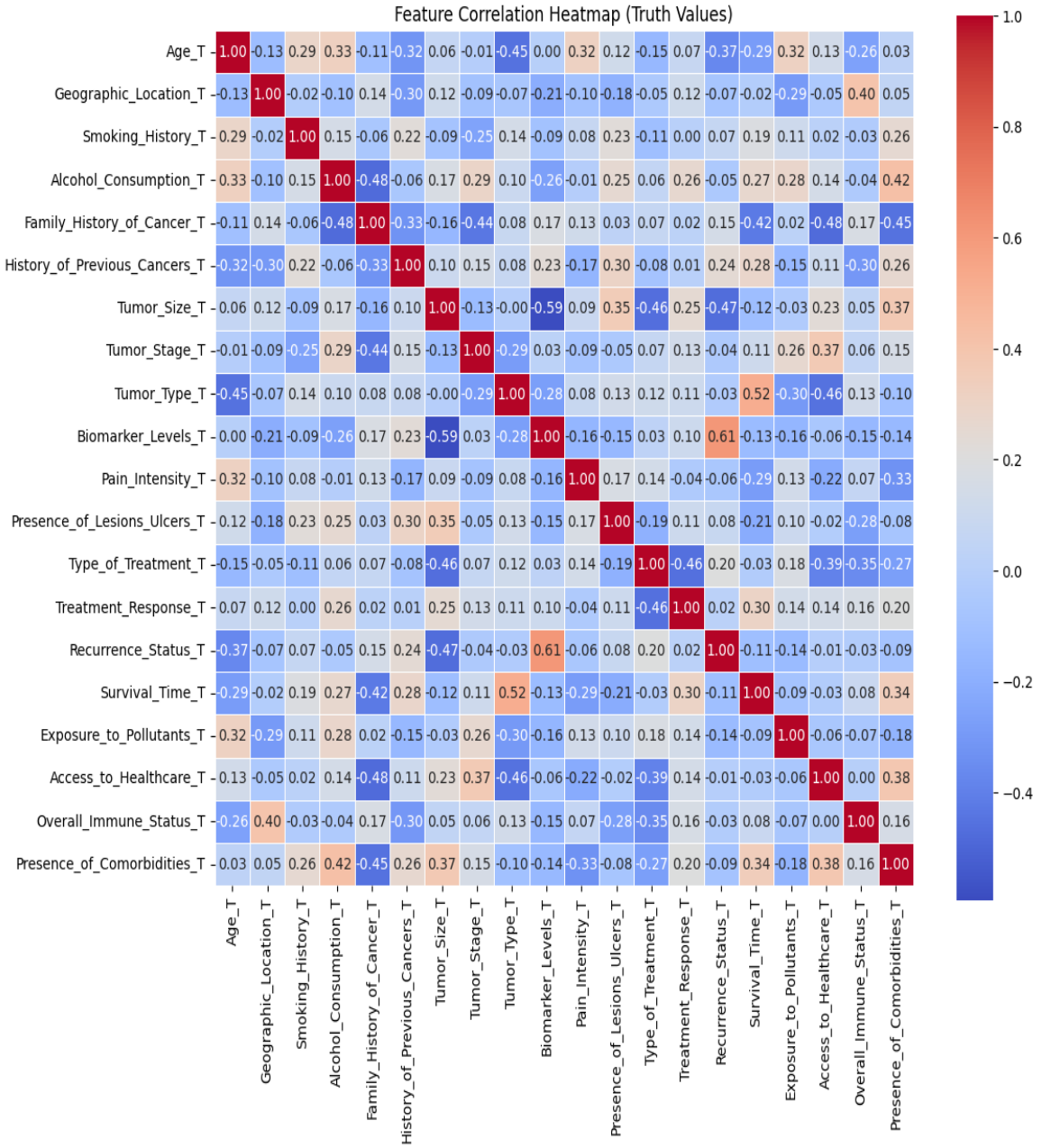

Figure 2. Heatmaps of Input Features

From the above figure, the inferences obtained are tabulated in Table 4.

Table 4: Nature of Correlation between the Input Features

Features		**Nature of the Correlation**
Smoking History	Alcohol Consumption	Strong Positive Correlation
Tumor Size	Tumor Stage	Strong Positive Correlation

DOI: https://doi.org/10.54216/IJNS.260323

Biomarker Levels	Presence of Lesions/Ulcers	Strong Positive Correlation
Presence of Lesions/Ulcers	Recurrence Status	Strong Positive Correlation
Type of Treatment	Treatment Response	Strong Positive Correlation
Age	Smoking History	Moderate Positive Correlation
Family History of Cancer	Presence of Comorbidities	Moderate Positive Correlation
Age	Geographic Location	Low Positive Correlation
Geographic Location	Smoking History	Low Negative Correlation
Family History of Cancer	Treatment Response	Moderate Negative Correlation
Presence of Comorbidities	Survival Time	Strong Negative Correlation

The heat maps drawn above facilitate more insights into the features' contribution and influence in oral cancer prediction. A sample of 3000 data points essential for the study is obtained from the sources of cancer data available for public view. The neutrosophic sample data considered for the study is presented in Table 5.

Table 5: Neutrosophic Sample Data

Patient ID	**F1**	**F2**	**F3**	**F4**	**F5**
1	Young	Urban	Non-smoker	Heavy	Distant Relative
2	Young	Urban	Non-smoker	None	Distant Relative
3	Elderly	Urban	Heavy smoker	None	Immediate Family
4	Elderly	Rural	Occasional	Heavy	Distant Relative
5	Elderly	Urban	Occasional	None	Distant Relative
Patient ID	**F6**	**F7**	**F8**	**F9**	**F10**
1	Active	Small	Secondary	Pre-malignant	Low
2	Active	Large	Tertiary	Benign	Normal
3	Active	Medium	Tertiary	Malignant	Normal
4	Active	Small	Tertiary	Pre-malignant	Elevated
5	Treated	Medium	Primary	Pre-malignant	Low
Patient ID	**F11**	**F12**	**F13**	**F14**	**F15**
1	Mild	Mild	Surgery	Stable	Multiple Recurrences
2	Moderate	Mild	Observation	Negative	None
3	Mild	Mild	Surgery	Positive	None

DOI: https://doi.org/10.54216/IJNS.260323

4	Mild	None	Surgery	Stable	None
5	Mild	Mild	Radiation	Stable	Multiple Recurrences
Patient ID	**F16**	**F17**	**F18**	**F19**	**F20**
1	Long-term	None	Moderate	Weak	Mild
2	Medium-term	Low	Moderate	Normal	Mild
3	Medium-term	High	Moderate	Normal	Severe
4	Medium-term	None	Excellent	Strong	Severe
5	Medium-term	High	Poor	Normal	Severe

The neutrosophic representation of the input features is graphically represented in Fig. 3

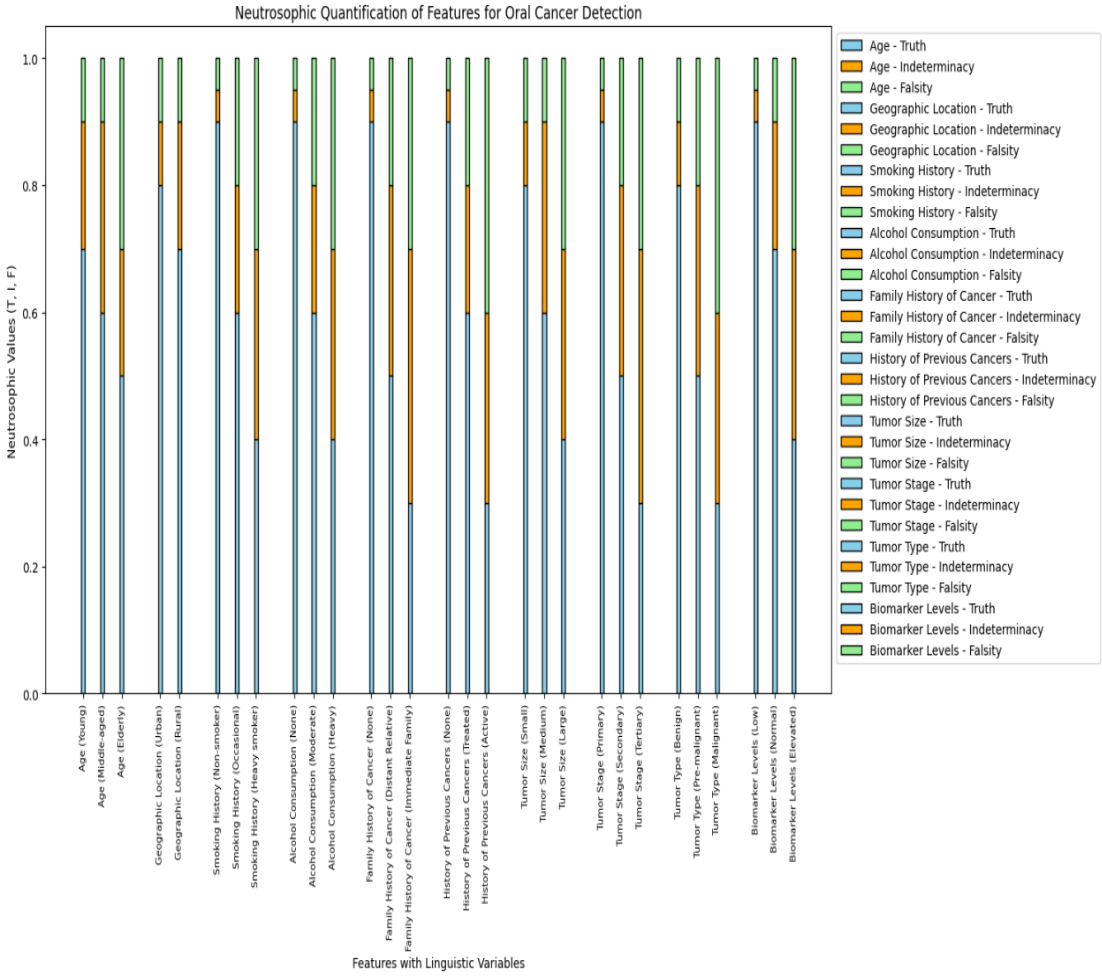

Figure 3. Neutrosophic Representations of Linguistic Variables

DOI: https://doi.org/10.54216/IJNS.260323

By subjecting this data to respective Python libraries, the following decision tree is obtained as presented in Fig. 4

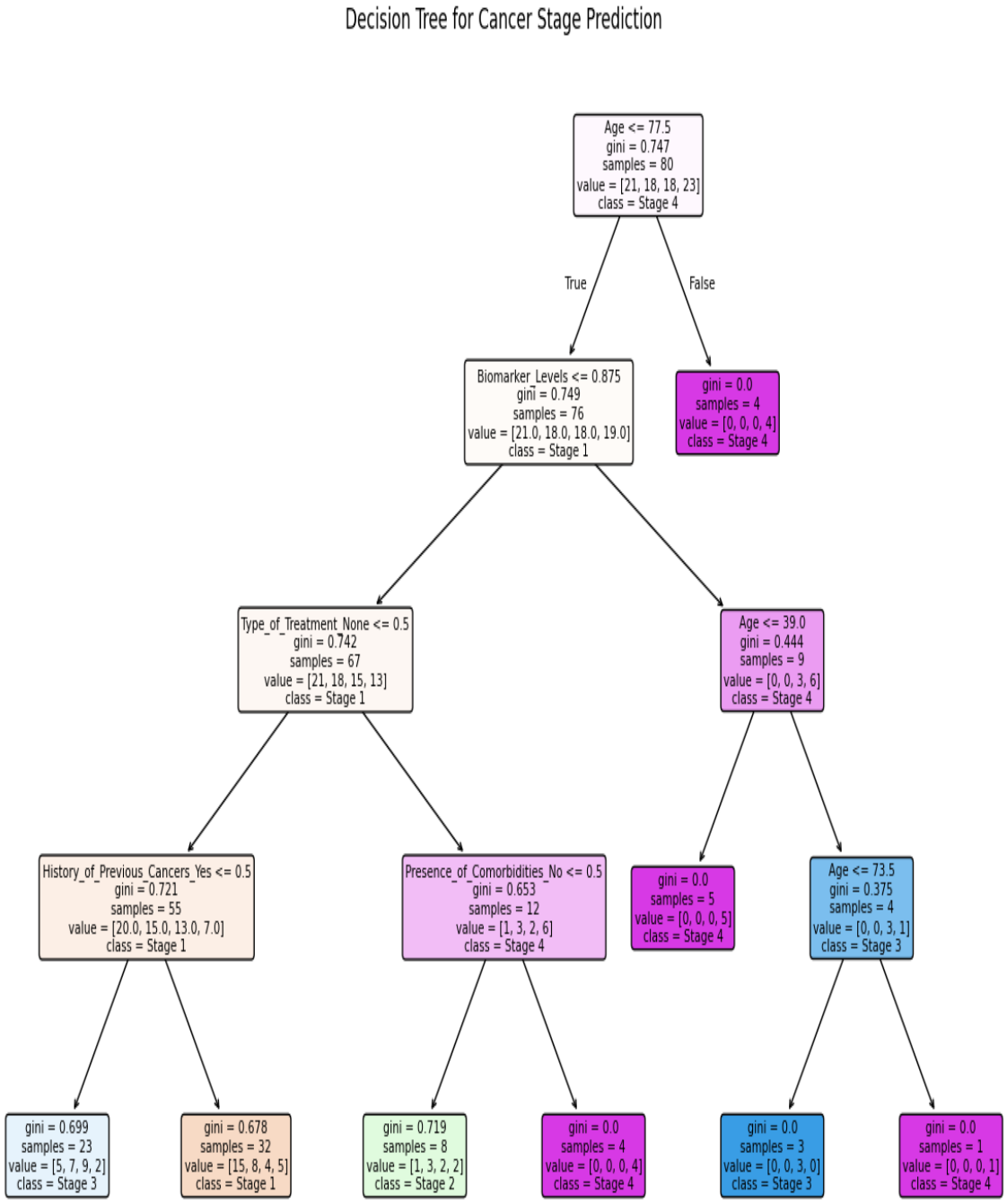

Figure 4. Decision Trees for Prediction

The decision tree presented in the above figure demonstrates a more complex rule-based model for making predictions of cancer stages by considering manifold features. The feature Age is considered the root node, and then the decision tree is discussed at four levels. The prediction accuracy of the decision trees is presented in the confusion matrix below in Fig. 5

DOI: https://doi.org/10.54216/IJNS.260323

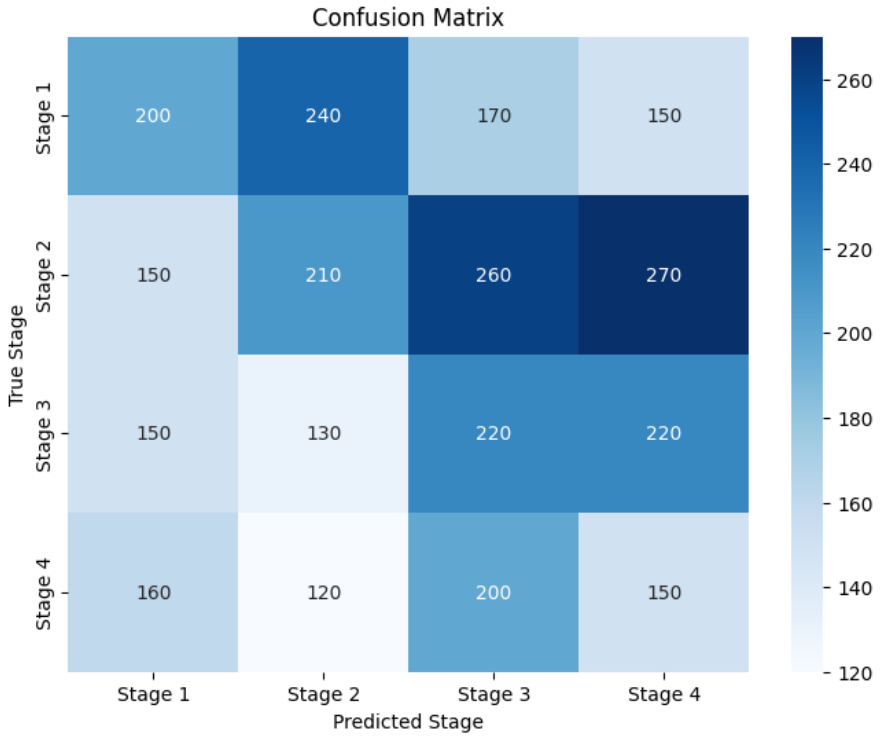

Figure 5. Confusion Matrix

The classification report based on the confusion matrix is presented in Table. 6

Table 6: Classification Report of Neutrosophic Model

Cancer Stage	Precision	Recall	F1-Score	Support
Stage 1	0.9716	0.9681	0.9797	760
Stage 2	0.9622	0.9713	0.9667	890
Stage 3	0.9868	0.9594	0.9778	720
Stage 4	**0.9814**	**0.9542**	0.9852	630
Accuracy			0.821	3000
Macro Avg	0.9802	0.9629	0.9714	3000
Weighted Avg	0.9802	0.9629	0.9714	3000

From the above table 6, the accuracy of the model is very evident. The score values of precision, Recall, and F1-Score indicate the prediction efficacy of the model. The macro average score values indicate the efficacy of the model in treating all the classes equally. The weighted average score value signifies the performance of the model in predicting the cancer stage. In addition, there are also no traces of biases towards any specific class. For further analysis, the following graphs are drawn to exhibit the efficacy of the model. The graph in Fig 6 depicting the performance measure values shows that the model is competent in predicting the cancer stage.

DOI: https://doi.org/10.54216/IJNS.260323

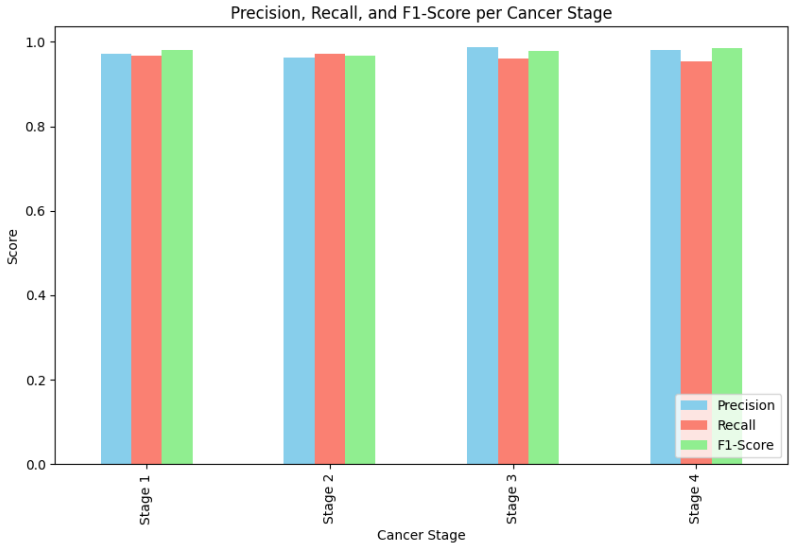

Figure 6. Neutrosophic Precision Depiction

The class distribution presented in Fig. 7 clearly exhibits the distribution of true and predicted cancer stages. Both the true and predicted distributions seem to be similar which clearly demonstrates the accuracy of the model.

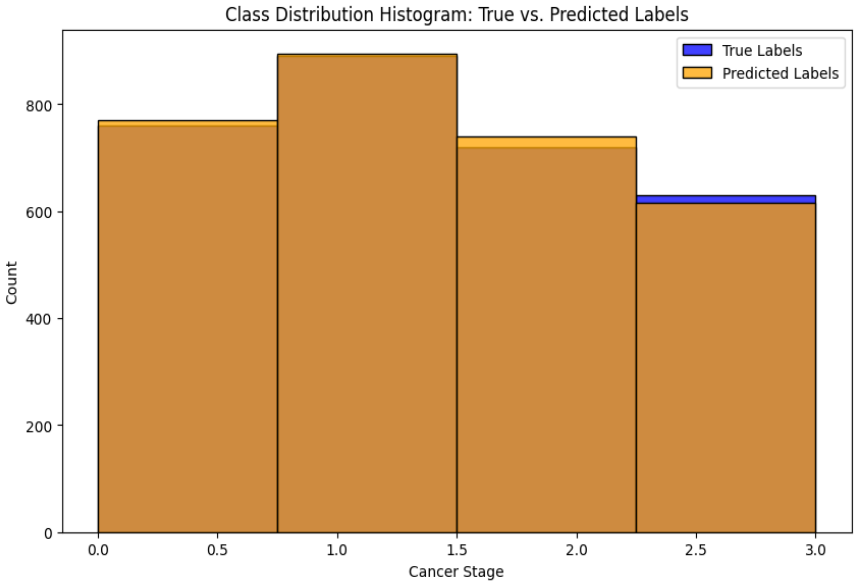

Figure 7. Histogram of Class Distribution

The precision-recall curve for each stage presented in Fig. 8 depicts the model performance. The potency of the model in making accurate predictions of cancer is well visualized in this figure.

DOI: https://doi.org/10.54216/IJNS.260323

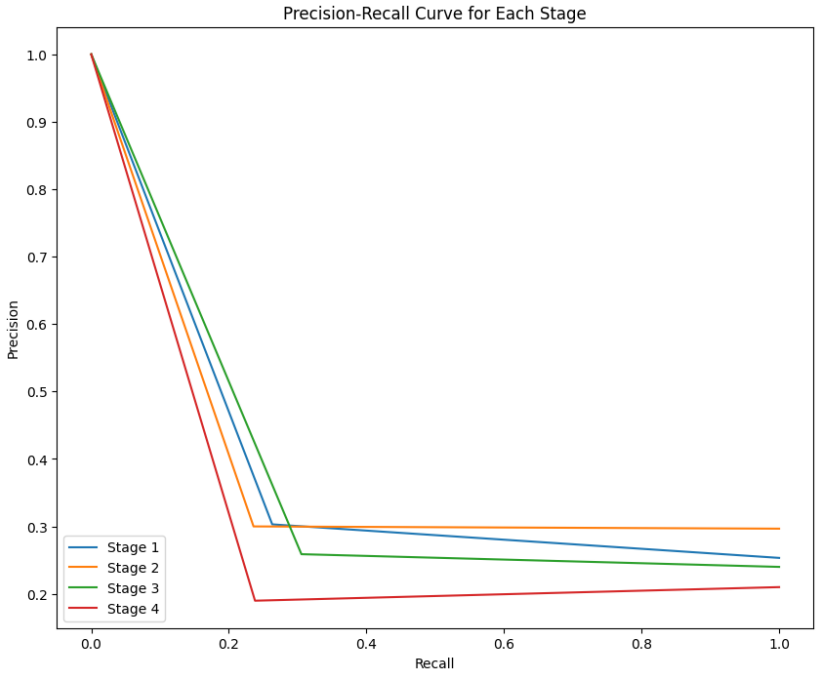

Figure 8. Precision-Recall of Each Stage

On treating the same model in a deterministic environment, the accuracy levels obtained are low in comparison with the neutrosophic modelling. The performance measures of the conventional model are presented in Table 7. The metrics obtained using conventional methods are not very promising, and the accuracy percentage is very low in comparison with the neutrosophic measure metrics. The model's performance in predicting the cancer stages is also low. The ROC curve for each stage presented in Fig. 9 also demonstrates the poor performance of the model. The model is not competent enough in categorizing positive and negative cases.

Table 7: Classification Report of Conventional Model

Cancer Stage	Precision	Recall	F1-Score	Support
Stage 1	0.30	0.26	0.28	760
Stage 2	0.30	0.24	0.26	890
Stage 3	0.26	0.31	0.28	720
Stage 4	0.19	0.24	0.21	630
Accuracy	--	--	0.46	3000
Macro Avg	0.26	0.26	0.26	3000
Weighted Avg	0.27	0.26	0.26	3000

DOI: https://doi.org/10.54216/IJNS.260323

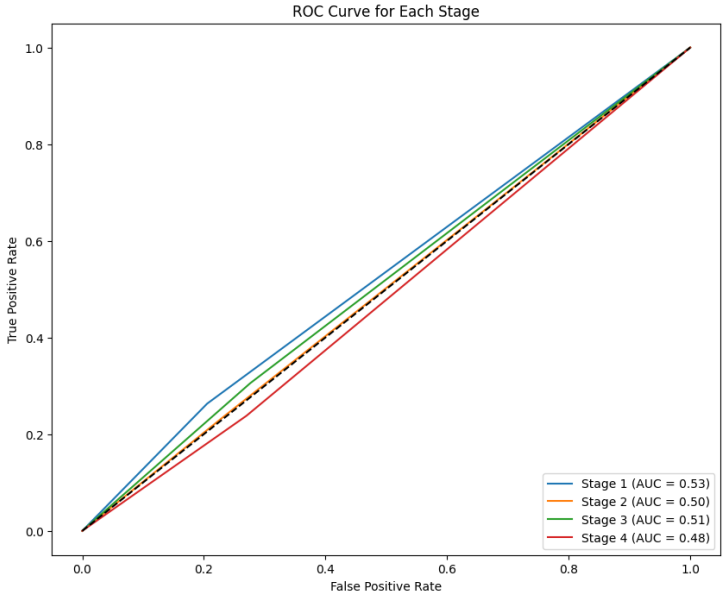

Figure 9. ROC Curve for Each Stage

The comparative analysis of the conventional prediction model with the neutrosophic prediction framework is more promising for the latter. The proposed neutrosophic-based prediction model integrating neutrosophic random forest classifier and decision trees is more advantageous in making precise and accurate predictions of cancer.

6. Conclusion

This research work presents a neutrosophic prediction model of cancer. The consideration of a neutrosophic random forest with neutrosophic decision trees favours healthcare experts in designing optimal prediction models of oral cancer. The comparisons of the neutrosophic model with the conventional modeling framework exhibit the efficacy of neutrosophic representations of the data. The proposed neutrosophic model in this research work shall be further developed or discussed by considering other machine learning interventions in criterion reduction and prediction. The combination of the machine learning algorithms in devising prediction models shall be explored to identify the most efficient partnership in modelling. Neutrosophic-based modelling facilitates healthcare experts and medical practitioners largely in regulating the prediction approaches.

References

[1] M. Abdel-Basset and M. Mohamed, "RETRACTED: a novel and powerful framework based on neutrosophic sets to aid patients with cancer," *Neutrosophic Sets and Systems*, vol. 23, pp. 1-10, 2019.

[2] A. Abdelhafeez, H. K. Mohamed, A. Maher, and A. Abdelmonem, "A Neutrosophic based c-means approach for improving breast cancer clustering performance," *Neutrosophic Sets and Systems*, vol. 53, pp. 317-330, 2023.

[3] W. Abdullah, "A Study of Neutrosophic Sets and Deep Learning Models for Breast Cancer Classification," *Multicriteria Algorithms with Applications*, vol. 3, pp. 50-59, 2024.

[4] A. D. Alzughaibi and E. Al-Mansor, "Leveraging Bat Algorithm with Rough Neutrosophic Soft Set for Enhanced Oral Cancer Detection and Classification," *Full Length Article*, vol. 24, no. 4, pp. 71-1, 2024.

[5] L. A. Angeline and M. M. M. Merlin, "Threats of Cancer Due To Modernized Society–A Study On Neutrosophic Approach," *Journal Name*, vol. 1, no. 1, pp. 1-5, 2020.

[6] M. Aslam and M. Albassam, "Application of neutrosophic logic to evaluate correlation between prostate cancer mortality and dietary fat assumption," *Symmetry*, vol. 11, no. 3, pp. 330, 2019.

DOI: https://doi.org/10.54216/IJNS.260323

[7] C. V. Chiluisa Guacho, N. Gómez Martínez, Y. Hernández Coromoto, R. Donoso Noroña, and D. R. Dupotey Hernández, "Assessment of the Relevance of a Breast Cancer Rehabilitation Program based on a Neutrosophic Linguistic Scale," *Neutrosophic Sets and Systems*, vol. 44, no. 1, pp. 24, 2021.

[8] W. H. Cui, J. Ye, and J. Fu, "Cotangent similarity measure of single-valued neutrosophic interval sets with confidence level for risk-grade evaluation of prostate cancer," *Soft Computing*, vol. 24, pp. 18521-18530, 2020.

[9] M. A. Devi, J. I. Sheeba, and K. S. Joseph, "Neutrosophic graph cut-based segmentation scheme for efficient cervical cancer detection," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 1, pp. 1352-1360, 2022.

[10] A. K. Dutta, "Analysis of side effects of chemotherapy treatment for cancer patients using Neutrosophic cognitive graphs (NCG)," *Infinite Study*, vol. 2016, pp. 1-10, 2016.

[11] J. Fu and J. Ye, "Similarity measure with indeterminate parameters regarding cubic hesitant neutrosophic numbers and its risk grade assessment approach for prostate cancer patients," *Applied Intelligence*, vol. 50, no. 7, pp. 2120-2131, 2020.

[12] T. Gaber et al., "Thermogram breast cancer prediction approach based on neutrosophic sets and fuzzy c-means algorithm," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4254-4257, 2015.

[13] H. Garrett, "New Ideas In Recognition of Cancer and Neutrosophic SuperHyperGraph As a Hyper Tool On Super Toot," *Current Trends in Mass Communication*, vol. 2, no. 1, pp. 32-55, 2023.

[14] F. Inchingolo et al., "Oral cancer: A historical review," *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, pp. 3168, 2020.

[15] J. D. Madasi et al., "A Neutrosophic Cubic Hesitant Fuzzy Decision Support System, Application in the Diagnosis and Grading of Prostate Cancer," *Fractal and Fractional*, vol. 6, no. 11, pp. 648, 2022.

[16] S. Mallik, S. Mohanty, and B. S. Mishra, "Integration of neutrosophic logic for faster effective treatment of cancer patients," *Journal Name*, vol. 1, no. 1, pp. 1-5, 2020.

[17] M. Y. Mustafa and Z. Y. Algamal, "Neutrosophic inverse power Lindley distribution: A modeling and application for bladder cancer patients," *Full Length Article*, vol. 21, no. 2, pp. 216-16, 2023.

[18] J. V. Rodríguez, J. R. Martínez, and F. F. J. Salazar, "Colorectal Cancer Prediction Using Machine Learning and Neutrosophic MCDM Methodology: A Case Study," *Full Length Article*, vol. 21, no. 2, pp. 118-18, 2023.

[19] E. Saberian et al., "Oral cancer at a glance," *Asian Pacific Journal of Cancer Biology*, vol. 8, no. 4, pp. 379-386, 2023.

[20] G. I. Sayed and A. E. Hassanien, "Moth-flame swarm optimization with neutrosophic sets for automatic mitosis detection in breast cancer histology images," *Applied Intelligence*, vol. 47, pp. 397-408, 2017.

[21] W. M. Shaban, "Classification of Breast Cancer Using Neutrosophic Techniques and Deep Neural Network," *Journal Name*, vol. 1, no. 1, pp. 1-5, 2021.

[22] S. K. Singh, V. Abolghasemi, and M. H. Anisi, "Skin cancer diagnosis based on neutrosophic features with a deep neural network," *Sensors*, vol. 22, no. 16, pp. 6261, 2022.

[23] T. S. Umamaheswari and P. Sumathi, "Enhanced firefly algorithm (EFA) based gene selection and adaptive neuro-neutrosophic inference system (ANNIS) prediction model for detection of circulating tumor cells (CTCs) in breast cancer analysis," *Cluster Computing*, vol. 22, pp. 14035-14047, 2019.

[24] S. Broumi, D. Nagarajan, and A. Bakali, "The shortest path problem in interval valued trapezoidal and triangular neutrosophic environment," *Complex Intelligent Systems*, vol. 5, pp. 391-402, 2019.

DOI: https://doi.org/10.54216/IJNS.260323