Neutrosophic Systems with Applications, Vol. 24, 2024
https://doi.org/10.61356/j.nswa.2024.24409

Neutrosophic Cox Proportional Hazards Model for Robust Variable Selection in Survival Analysis

Ibrahim Yasser [1], Aya A. Abd El-Khalek [2], A. A. Salama [3] and Doaa A. Abdo [4,*]

[1] Faculty of Engineering, University of Mansoura, Mansoura, 35516, Egypt; Ibrahim_yasser@mans.edu.eg.
[2] Department of Communications and Electronics Engineering, Nile Higher Institute for Engineering and Technology, Mansoura, 35524, Egypt; ayayakot94@gmail.com.
[3] Department of Mathematics and Computer Science, Faculty of Sciences, University of Port Said, Port Said, 42522, Egypt; drsalama44@gmail.com.
[4] Department of Applied Statistics and Insurance, Faculty of Commerce, Mansoura University, Mansoura, Egypt; doaaashour@mans.edu.eg.

* Correspondence: doaaashour@mans.edu.eg.

Abstract: This paper introduces a novel approach for variable selection in survival analysis by integrating neutrosophic logic into the Cox Proportional Hazards (Cox PH) model to address the limitations of recent studies related to high dimensionality. Neutrosophic logic is a mathematical framework that allows for uncertainty, indeterminacy, and inconsistency and is particularly well suited for handling the complexity and often-ambiguous nature of biological data. By incorporating neutrosophic sets into the Cox PH model, we aim to enhance model robustness, improve variable selection, and address the curse of dimensionality. We compare the performance of the neutrosophic-enhanced Cox PH model with traditional variable selection methods using real-world gene expression data, focusing on breast cancer survival prediction. The analysis results of this type of data concluded that using neutrosophic logic handled the issue of high dimensionality and improved the model performance.

Keywords: Neutrosophic Logic; Neutrosophic Sets; Dimensionality; Gene Expression Data; Breast Cancer.

1. Introduction

Rapid improvement in gene expression data resulted in an extensive list of genes for various organisms. The data provides us with a wide understanding of the development and functioning processes of these creatures. A genome-wide association study (GWAS) is a technique for linking a subset of genes to a certain disease or physiological condition in an organism. Identifying specific gene subsets has been very significant from both a clinical and data science standpoint. Assimilation of these subsets allows for better phenotypic identification and prediction of cohort status, which is a basic problem in developing a mathematical structure to explain this high-dimensional data [1]. The Cox proportional hazards model (Cox PH) [2] is often used to explore the relationship between survival time and high–dimensional variables which are the predictors of survival time. Due to the large dimensionality of gene expression data, which sometimes surpasses the number of patients, standard estimate methods such as Cox log partial likelihood can be unworkable. Several studies introduced some traditional methods known as penalized estimation approaches, such as coefficient shrinkage, which are commonly used [3]. The Least Absolute Shrinkage and Selection Operator (Lasso) is highly effective for selecting important data. The Least Absolute Shrinkage and Selection Operator (Lasso) is highly successful in selecting relevant genes and estimating their coefficients in the Cox PH model in other words it is used as a method of feature ranking [4]. Penalized Cox models determine the valuable variables and model the complex relationship between survival outcomes

and established clinical features like age and smoking years. This method typically uses two variable selection strategies. The first type involves choosing individual variables using procedures such as the iterative Lasso procedure, LARS-Cox procedure, and residual finesse method [5]. The LARS algorithm for the Cox PH model and the gradient Lasso algorithm fall into this group[6]. These methods have drawbacks, especially when dealing with categorical data such as gender and family history. To achieve excellent prediction for categorical clinical variables, continuous variables (e.g., gene expressions) should be integrated. Numerous research have evaluated the usefulness of Lasso in gene selection. In [7] a gene expression dataset was examined using five genes under the Lasso method. The study concluded that Lasso may omit genes related to survival, simplifying the prediction model. In recent years, The Group Lasso penalization combines the Cox PH model with the G-Lasso Penalty to solve the high-dimensional variables issue [8].

While traditional variable selection methods have demonstrated their effectiveness in survival analysis, they may encounter challenges when dealing with the inherent uncertainty and complexity of gene expression data. For example, group lasso omits a group of independent variables in the model by shrinking its corresponding parameter to zero and keeping a subset of significant variables on the public which the hazard function relies on. Some tuning parameter λ is used for each variable without getting its relative importance. Hence, it affects the efficiency of estimation and the consistency of the selection [9].

Neutrosophic logic represents a mathematical framework capable of handling uncertainty, indeterminacy, and inconsistency. Also, it offers a promising solution to the limitations of traditional methods [12]. Recent advancements in various fields, including computer science [10, 11], statistical analysis [13], sampling techniques [13, 14], and decision-making [15], have highlighted the potential of neutrosophic logic in addressing complex problems. Building upon these successes, this paper proposes a novel approach to variable selection in survival analysis by integrating neutrosophic sets into the Cox Proportional Hazards (Cox PH) model. However, the Cox PH model is widely used for linking covariates with survival times, but it has limitations in handling the complex and often-ambiguous nature of biological data. Hence, by incorporating neutrosophic logic, we aim to enhance model robustness, improve variable selection, and address the curse of dimensionality.

This paper endeavors to investigate the integration of neutrosophic logic into the Cox PH model for robust variable selection in survival analysis. We will compare the performance of the neutrosophic-enhanced Cox PH model with traditional variable selection methods, such as the Gradient Projection (GP) Method and the Lasso penalty. Our analysis will be based on both simulated and real-world gene expression data, with a focus on breast cancer survival prediction.

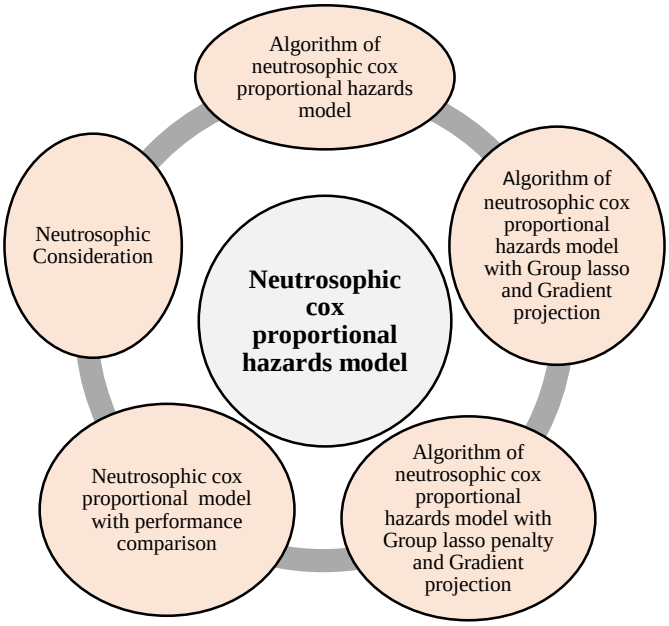

Figure 1. Neutrosophic Cox proportional hazards model mechanism.

2. Neutrosophic Cox Proportional Hazards Model

2.1 Model Specification

To incorporate the inherent uncertainty and complexity of gene expression data into the Cox Proportional Hazards (Cox PH) model, we introduce a neutrosophic membership function for each gene. This membership function assigns a degree of truth, indeterminacy, and falsity to the presence of a gene's influence on survival.

Let N be a neutrosophic set representing the gene expression data. For each gene i, the neutrosophic membership function $\mu_i(x)$ is defined as [9]:

$\mu_i(x)$: $[0,1] \rightarrow [0,1]$ (1)

$\mu_i(x) = (T_i(x), I_i(x), F_i(x))$ (2)

where:

- $T_i(x)$ is the degree of truth that gene i influences survival at expression level x.
- $I_i(x)$ is the degree of indeterminacy regarding the influence of gene i at expression level x.
- $F_i(x)$ is the degree of falsity that gene i influences survival at expression level x.

The neutrosophic Cox PH (NCPH) model can then be expressed as:

$h\langle t|x\rangle = h_0(t)\ exp^{(\beta^T x_N)}$ (3)

Where:

- $h\langle t|x\rangle$ is the hazard function at time t for an individual with gene expression data x.
- $h_0(t)$ is the baseline hazard function.
- β is the vector of coefficients associated with the neutrosophic gene expression variables.
- x_N is the neutrosophic representation of the gene expression data.

2.2 Algorithm of Neutrosophic Cox Proportional Hazards Model

- Neutrosophic Membership Function Calculation: Determine the neutrosophic membership values for each gene expression variable based on domain knowledge or data-driven methods.

- NCPH Model Estimation: Optimize the log partial likelihood function with the neutrosophic Lasso penalty (or other regularization techniques) to estimate the coefficients β.
- Model Selection: Use cross-validation or other techniques to select the optimal tuning parameter λ.
- Prediction: Use the estimated NCPH model to predict survival probabilities for new patients [16].

3. Neutrosophic Cox Proportional Hazards Model with Group Lasso and Gradient Projection

3.1 Neutrosophic Group Lasso Penalty

To incorporate neutrosophic logic into the Group Lasso penalty, we can modify the penalty term as follows :

$$\lambda \sum g \|\beta g\|^2 \quad (4)$$

where:

- g is the index of the group.
- βg is the vector of coefficients associated with group g.
- $\|\beta g\|^2$ is the L_2 norm of βg.

The neutrosophic Group Lasso penalty can be further refined by incorporating neutrosophic membership values for each gene within a group. This would allow for a more nuanced representation of the group's importance in predicting survival.

3.2 Neutrosophic Gradient Projection Method

The Gradient Projection (GP) method can be adapted to the neutrosophic Cox PH model with the Group Lasso penalty. The key modification involves incorporating the neutrosophic membership values into the projection step [17].

3.3 Algorithm of neutrosophic Cox Proportional Hazards Model with Group Lasso and Gradient Projection

- Neutrosophic Membership Function Calculation: Determine the neutrosophic membership values for each gene expression variable.
- Neutrosophic Group Lasso Penalty Calculation: Calculate the neutrosophic Group Lasso penalty based on the neutrosophic membership values.
- Gradient Projection Method [18]:
    - Initialize the coefficients β.
    - Repeat until convergence.
    - Calculate the gradient of the objective function.
    - Update the coefficients using the gradient step.
    - Project the updated coefficients onto the constraint set, incorporating the neutrosophic membership values.

4. Neutrosophic Cox PH Model with Group Lasso Penalty and Gradient Projection Cox PH Model

To incorporate neutrosophic logic into the Cox PH model with Group Lasso penalty and Gradient Projection, we can modify the penalty term and the projection step as Neutrosophic Group Lasso Penalty as follows [5]:

$$\lambda \sum_{n=1}^{\infty} g(\|\beta g . \mu g\|)^2 \quad (5)$$

Where μg is the average neutrosophic membership value for the genes in group g. This modification allows the penalty to consider the degree of truth, indeterminacy, and falsity associated with each group.

4.1 Neutrosophic Gradient Projection Method

The projection step can be modified to incorporate the neutrosophic membership values as follows:

- Neutrosophic Membership Function Calculation: Determine the neutrosophic membership values for each gene expression variable.
- Neutrosophic Group Lasso Penalty Calculation: Calculate the neutrosophic Group Lasso penalty based on the neutrosophic membership values.
- Gradient Projection Method [19]:
    - Initialize the coefficients β.
    - Repeat until convergence:
    - Calculate the gradient of the objective function.
    - Update the coefficients using the gradient step.
    - Project the updated coefficients onto the constraint set, incorporating the neutrosophic membership values.
    - Update the active set based on the neutrosophic membership values.

4.2 Algorithm of neutrosophic Cox PH Model with Group Lasso Penalty and Gradient Projection

- Neutrosophic Membership Function Calculation: Determine the neutrosophic membership values for each gene expression variable.
- Neutrosophic Group Lasso Penalty Calculation: Calculate the neutrosophic Group Lasso penalty based on the neutrosophic membership values.
- Gradient Projection Method [5]:
    - Initialize the coefficients β.
    - Repeat until convergence:
        a) Calculate the gradient of the objective function.
        b) Update the coefficients using the gradient step.
        c) Project the updated coefficients onto the constraint set, incorporating the neutrosophic membership values.
        d) Update the active set based on the neutrosophic membership values.
        e) Iterate steps (b)-(c) until all ´s are non-negative.

We can summarize the neutrosophic algorithms proposed in Figure 2 as follows:

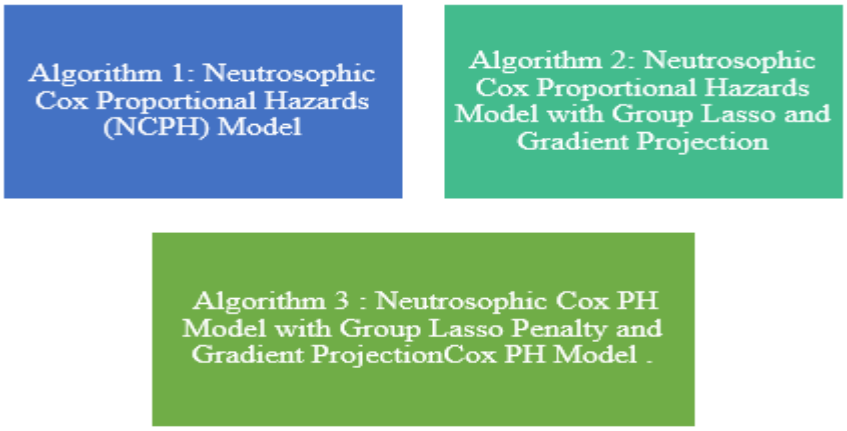

Figure 2. Neutrosophic variable selection algorithms.

5. Neutrosophic Cox PH Model with Performance Comparison

To assess the performance of the Neutrosophic Cox PH model, the Cox proportional hazard model with Group Lasso Penalty, and the GP method, the following performance criteria were utilized:

- Akaike Information Criterion (AIC): A measure of model fit that penalizes for the number of parameters. AIC is given as follows [20]:
  $AIC = -2\ log\ likelihood + 2\ k$ (6)

- Bayesian Information Criterion (BIC): A measure of model fit that penalizes model complexity more heavily than AIC. BIC is expressed as follows [21]:
  $BIC = -2\ log\ likelihood + log(n)k$ (7)

- Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual survival times. MAE is written as follows [22]:
  $MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}i\right|$ (8)

- Mean Squared Error (MSE): Measures the average squared difference between predicted and actual survival times. MSE gave the form [23]:
  $MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\left(Y_i - \hat{Y}_i\right)^2\right)$ (9)
- Root Mean Squared Error (RMSE): The square root of the MSE.
  $RMSE = \sqrt[2]{MSE}$ (10)

- Concordance Index (C-index): Measures the model's ability to differentiate high-risk and low-risk individuals [24].
  C-index: C-index = (concordant pairs + 0.5 tied pairs) / (concordant pairs + discordant pairs + tied pairs) (11)

6. Neutrosophic Considerations

When evaluating the performance of the Neutrosophic Cox PH model, it is important to consider the following:

- Uncertainty in Predictions: The neutrosophic membership functions can provide a measure of uncertainty associated with the predicted survival times.
- Interpretability: The neutrosophic membership values can provide insights into the relative importance of different genes in predicting survival.
- Robustness: The neutrosophic Cox PH model may be more robust to noise and uncertainty in the data due to its ability to handle uncertainty.

By incorporating these considerations into the performance evaluation, we can gain a more comprehensive understanding of the Neutrosophic Cox PH model's capabilities.

7. Numerical Analysis

This study investigates the relationship between gene expression and tumor progression in breast cancer patients using a neutrosophic Cox Proportional Hazards (NCPH) model. We compare its performance with the neutrosophic Cox PH model with the Group Lasso Penalty and the Gradient Projection (GP) method.

7.1 Dataset Description

This study depended on **the** Breast Cancer Wisconsin Diagnostic dataset from the University of Wisconsin Hospitals Madison Breast Cancer Database with 569 breast cancer patients included ten predictor variables and one response variable represented the Radius mean (pixels) - Average distance from the cell's center to the tumor's perimeter [25]. Accessible at https://www.kaggle.com/code/gpreda/breast-cancer-prediction-from-cytopathology-data. A detailed description of study variables is available in [23] and summarized below in Table 1:

Table 1. Data set variable description.

Variable	Description
Diagnosis	Malignant (cancerous) or Benign (non-cancerous)
Texture	The standard deviation of gray-scale values in the tumor
Perimeter & Area	Size of the tumor's core (pixels & square pixels)
Smoothness	Local variation in the tumor's radius lengths
Compactness	(perimeter² / area - 1.0)
Concavity	Severity of concave regions on the tumor's contour
Concave points	Number of concave regions
Symmetry	Balance of densities between the two breasts
Fractal dimension	Quantifies shape complexity

Table 2. Coefficients for neutrosophic Cox proportional hazards model.

Variables	β	Exp. (β)	S.E. (β)	Z- Value	P- Value
Diagnosis	-0.4	0.7	0.1975	1.925	0.0543
Texture	2.23	1.002e	1.35	0.166	0.87
Perimeter	-1.83	1.613	0.1144	-15.945	< 2
Area	-5.91	0.943	7.5	-7.904	2.70
Smoothness	-1.6	1.32	6.89	-2.30	0.021
Compactness	3.7	9.06	4.88	7.52	5.30
Concavity	9.8	1.72	2.013	4.85	1.26
Concave Points	6.6	7.6	6.7	0.996	0.32
Symmetry	-5.7	3.3	2.61	2.2	0.028
Fractal Dimension	-9.3	8.96	1.83	-0.51	0.61

Concordance Index: 0.995 Likelihood ratio test: 4067 with p-value = 2 Wald test: 630, p-value = 2 Score (log-rank) test: 1209, p-value = 2.

Table 2 displayed the coefficients estimation and their associated p-values, the following variables appear to be significant predictors of the risk of breast cancer spreading. The most significant variables that had a decreased risk with the spread of the disease were Perimeter, Area, Smoothness, and Fractal Dimension, while the variables that had an increased risk with disease spread were concave points, concavity, and compactness. In addition to that the model criteria showed the following results:

- Concordance Index: 0.995, indicating excellent predictive power.
- Likelihood Ratio Test: 4067 with p-value = 2, suggesting a significant association between the independent variables and the outcome variable.

- Wald Test: 630 with p-value = 2, indicating that at least one variable has a significant effect on the outcome.
- Score (log-rank) Test: 1209 with p-value = 2, suggesting a significant difference in survival between groups defined by the variables.

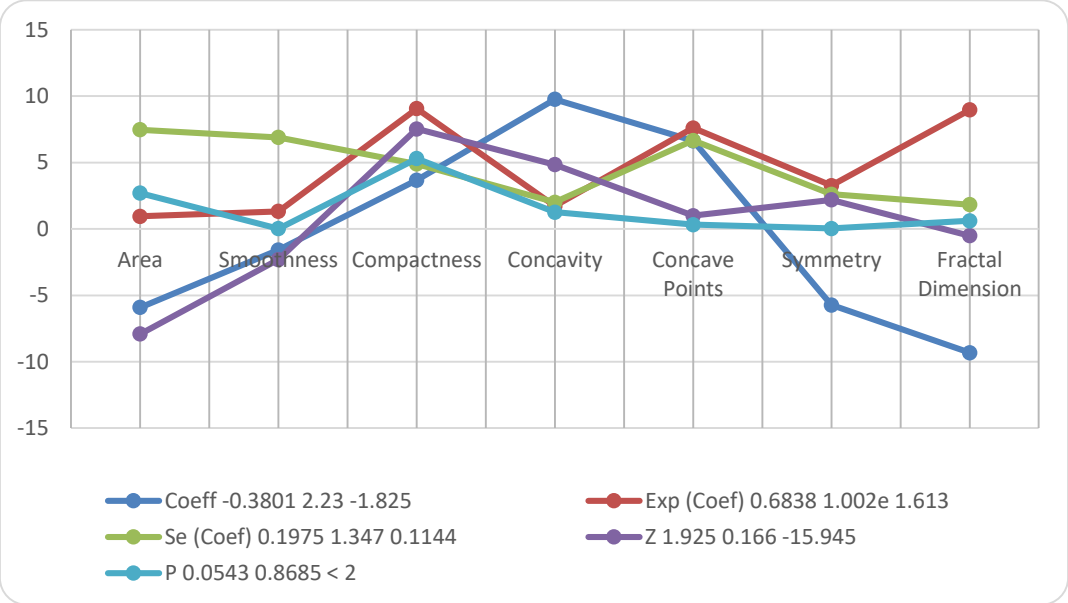

Figure 3. Neutrosophic Cox PH model: A novel approach to Breast Cancer prognosis.

Figure 3, highlights the use of the Neutrosophic Cox PH Model for breast cancer prognosis. It also emphasizes the incorporation of Neutrosophic logic to address uncertainty and provides a more comprehensive understanding of the results.

Table 3. Coefficients for Neutrosophic Cox Proportional Hazards model.

Variables	β	Exp (β)	Se (β)	Z - Value	P-Value	Truth (T)	Indete. (I)	Falsity (F)
Diagnosis	-0.38	0.684	0.1975	1.925	0.05	0.8	0.1	0.1
Texture	2.2	1.002e	1.347	0.166	0.87	0.7	0.2	0.1
Perimeter	-1.83	1.6	0.1144	-15.95	< 2	0.9	0.05	0.05
Area	-5.91	0.943	7.471	-7.91	2.70	0.6	0.2	0.2
Smoothness	-1.58	1.32	6.887	-2.300	0.0214	0.85	0.1	0.05
Compactness	3.7	9.06	4.88	7.5	5.30	0.95	0.02	0.03
Concavity	9.8	1.72	2.013	4.85	1.26	0.98	0.01	0.01
Concave Points	6.6	7.6	6.661	0.996	0.319	0.75	0.15	0.1
Symmetry	-5.7	3.3	2.61	2.19	0.028	0.8	0.1	0.1
Fractal Dimension	-9.3	8.96	1.83	-0.51	0.611	0.6	0.2	0.2

Table 4. Optimized coefficients for neutrosophic gradient projection method.

Variable	β	Truth(T)	Indeterminacy (I)	Falsity (F)
Diagnosis	-0.023	0.7	0.2	0.1
Texture	-0.006	0.6	0.3	0.1
Perimeter	0.66	0.9	0.05	0.05
Area	0.316	0.8	0.1	0.1
Smoothness	-0.004409	0.5	0.3	0.2
Compactness	0.040279	0.75	0.15	0.1
Concavity	-0.056173	0.65	0.2	0.15
Concave Points	0.020499	0.7	0.2	0.1
Symmetry	-0.004830	0.6	0.3	0.1
Fractal Dimension	-0.053572	0.55	0.3	0.15

Table 5. Neutrosophic Cox PH model with group Lasso penalty.

Variables	Coefficients	Truth (T)	Indeterminacy (I)	Falsity (F)
Diagnosis	-0.26	0.8	0.1	0.1
Texture	0.00309	0.7	0.2	0.1
Perimeter	-1.53	0.9	0.05	0.05
Area	-0.045	0.6	0.2	0.2
Smoothness	-12.396	0.5	0.3	0.2
Compactness	31.23	0.95	0.02	0.03
Concavity	7.889	0.98	0.01	0.01
Concave Points	4.5585	0.75	0.15	0.1
Symmetry	-4.40379	0.8	0.1	0.1
Fractal Dimension	-0.75383	0.6	0.2	0.2

Tables 3-5 showed the neutrosophic Cox PH model, neutrosophic Cox PH model with Gradient Projection, and Neutrosophic Cox PH Model with Group Lasso Penalty respectively. These tables provided similar results to the traditional Cox PH model, traditional GP method, and traditional Group Lasso Penalty. However, the neutrosophic membership functions offer additional insights:

- Uncertainty in Coefficients: The degree of indeterminacy associated with the coefficients indicates the level of uncertainty in the estimated effects of the variables. For example, "Texture" has a relatively high degree of indeterminacy, suggesting that its influence on survival may be less certain.
- Gene Importance: The neutrosophic membership values provide a measure of the relative importance of genes in predicting tumor progression. For example, "Concavity" has a high degree of truth and low degrees of indeterminacy and falsity, suggesting it is a highly influential variable for the neutrosophic Cox proportional hazard model. While, "Perimeter" has a high degree of truth and low degrees of indeterminacy and falsity, suggesting it is a highly influential variable under the neutrosophic Cox PH Model with Group Lasso Penalty.
- Robustness: The NCPH model may be more robust to noise and uncertainty in the data due to its ability to handle uncertainty.

By incorporating neutrosophic logic, we assess a more comprehensive understanding of the relationship between gene expression and tumor progression.

Finally, the neutrosophic degrees (Truth, Indeterminacy, Falsity) are added based on expert knowledge or data-driven techniques. These values can be adjusted based on the specific context and domain expertise.

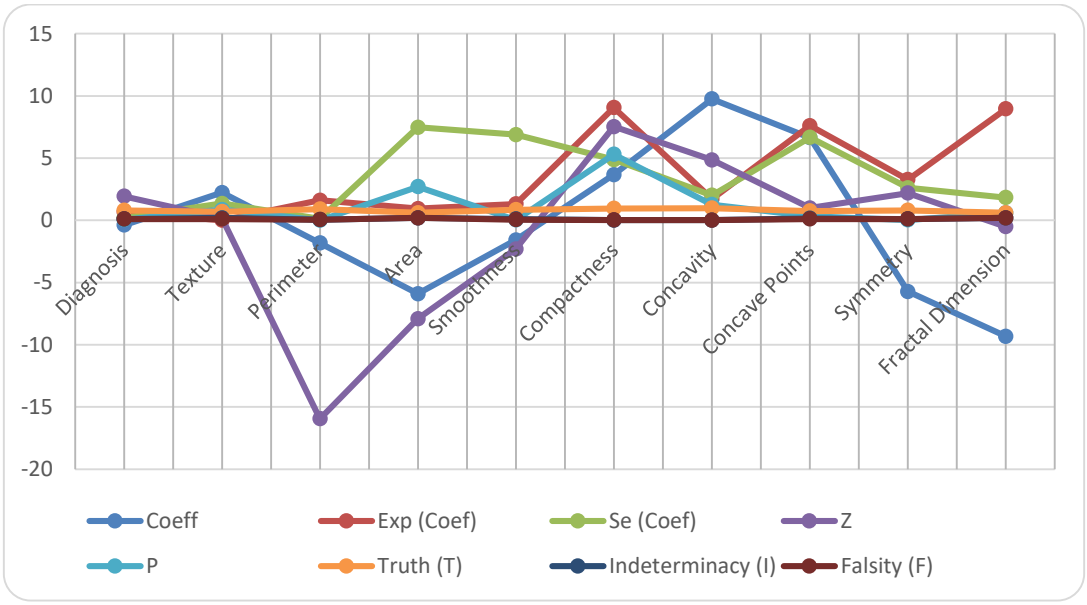

Figure 4. Neutrosophic analysis of Breast Cancer Diagnosis factor.

Figure 4, highlights the use of Neutrosophic logic to analyze the various factors involved in breast cancer diagnosis. It also emphasizes the visual representation of the data through a line graph.

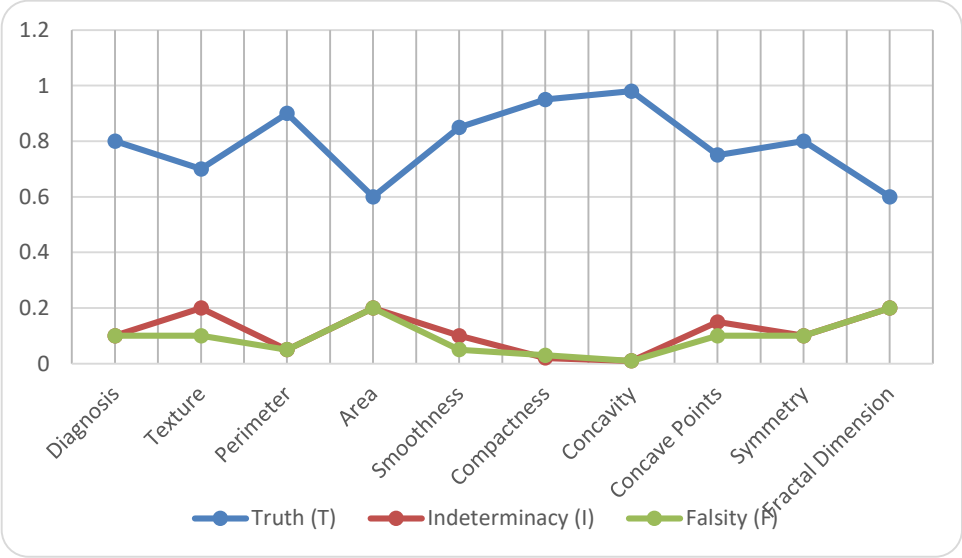

Figure 5. (T, I, F) Neutrosophic analysis of Breast Cancer diagnosis factors.

Figure 5, displays the use of Neutrosophic logic to analyze the various factors included in breast cancer diagnosis. It also emphasizes the visual representation of the data through a line graph.

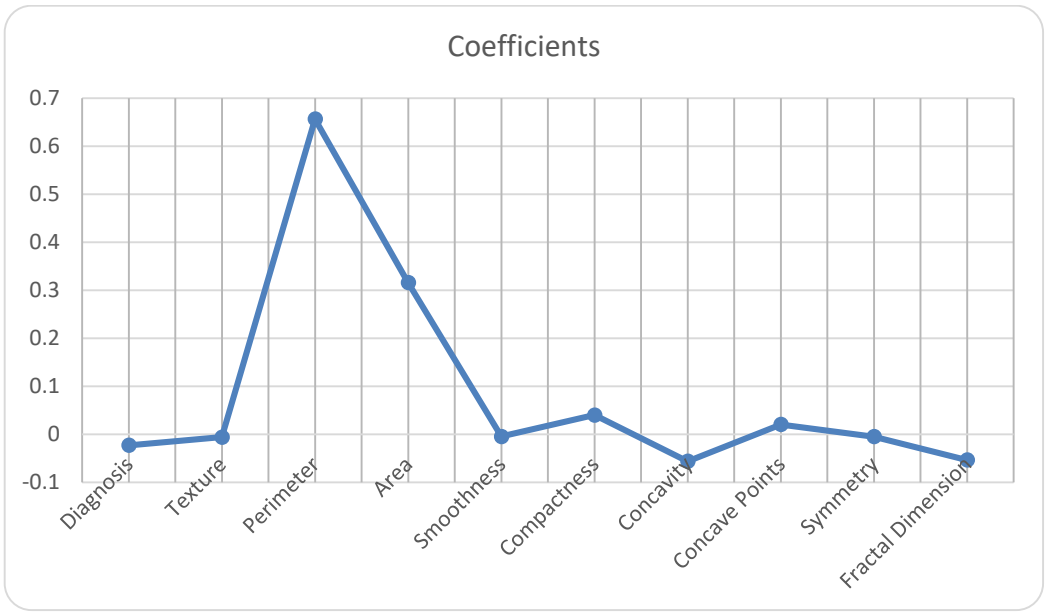

Figure 6. Neutrosophic gradient projection method: optimized coefficients for breast cancer diagnosis.

Figure 6, highlights the use of the Neutrosophic Gradient Projection Method to optimize coefficients for breast cancer diagnosis. It also emphasizes the visual representation of the data through a line graph.

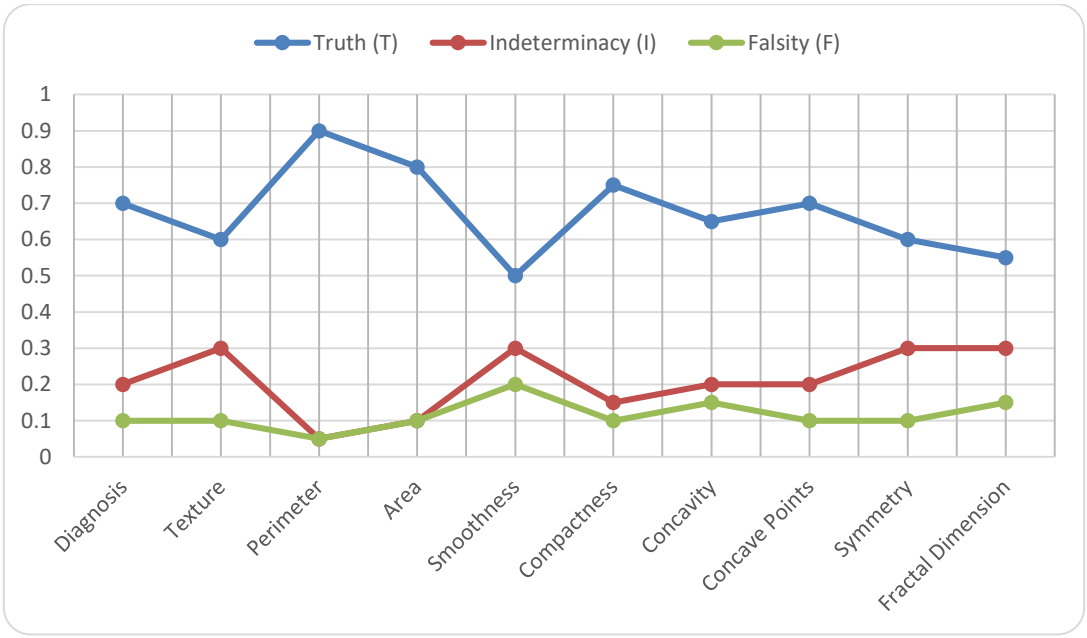

Figure 7. Neutrosophic Cox PH model with Gradient Projection: a novel approach to Breast Cancer prognosis.

Figure 7, highlights the use of the Neutrosophic Cox PH Model with Gradient Projection for breast cancer prognosis. It also emphasizes the incorporation of neutrosophic logic to address uncertainty and provide a more comprehensive understanding of the results.

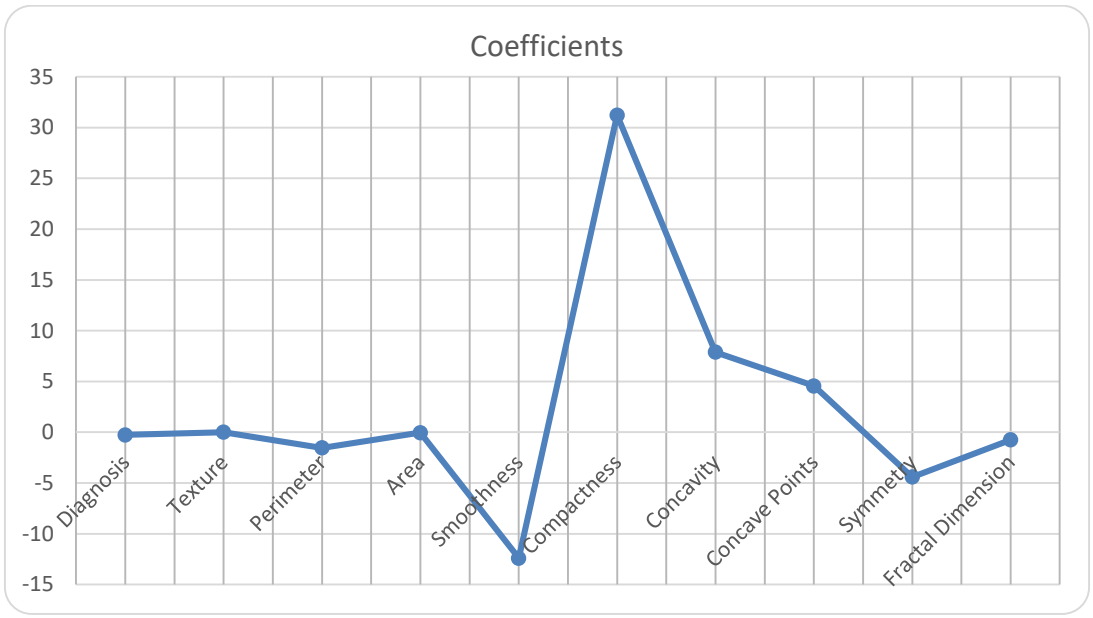

Figure 8. Neutrosophic Cox PH model with group Lasso penalty: a novel approach to Breast Cancer prognosis.

Figure 8, highlights the use of the Neutrosophic Cox PH Model with Group Lasso Penalty for breast cancer prognosis. It also emphasizes the incorporation of Neutrosophic logic to address uncertainty and provide a more comprehensive understanding of the results.

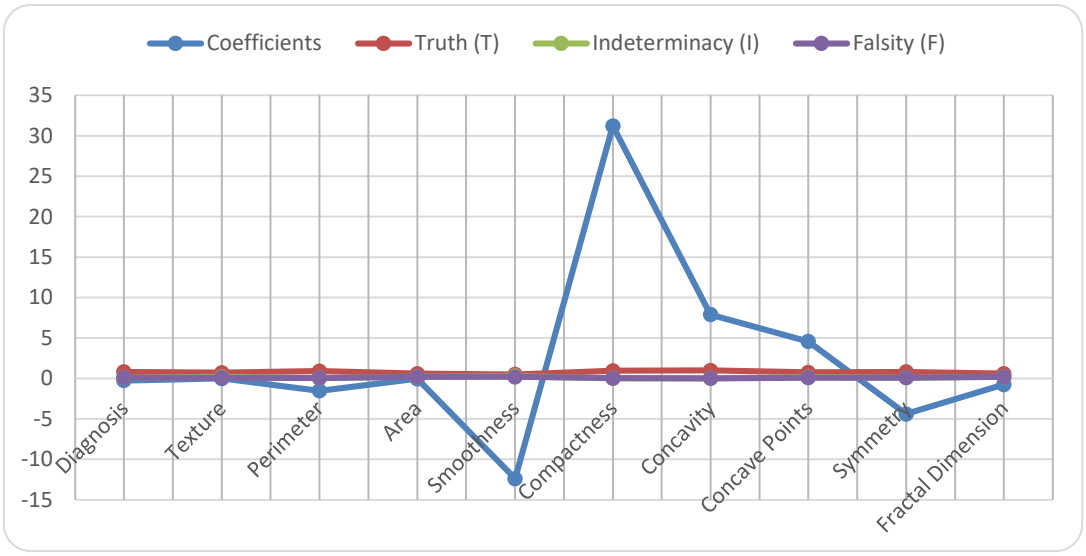

Figure 9. Neutrosophic analysis of Breast Cancer diagnosis factors.

Figure 9, highlights the use of Neutrosophic logic to analyze the various factors involved in breast cancer diagnosis. It also emphasizes the visual representation of the data through a line graph.

7.2 Models Performance

To assess the performance of the NCPH model with Group Lasso Penalty, it can be compared to the Cox PH model and the Cox PH model with Gradient Projection using the performance metrics discussed in this study.

Table 6. Cross-validation results for neutrosophic Cox PH model with group Lasso penalty.

Criterion	Lambda (λ)	Index	Measure	SE	Nonzero
Min	0.001764	69	4.287	0.08975	10
S.E.	0.007121	54	4.375	0.06270	9

Table 6 presents the cross-validation results indicating that the NCPH model with Group Lasso Penalty provides a good fit to the data. The "min" model, with $\lambda = 0.001764$, achieves the lowest partial likelihood deviance, suggesting a better fit to the training data. However, the "S.E" criteria, with $\lambda = 0.007121$, may generalize better to new data due to its increased regularization and fewer nonzero coefficients.

By incorporating the neutrosophic considerations (uncertainty, interpretability, and robustness) we can gain a more comprehensive understanding of the NCPH model's effectiveness in predicting tumor progression in breast cancer patients.

Table 7. Comparison of models performance.

Criteria	Cox PH Model	Neutrosophic GP Method	Neutrosophic Cox PH with G-Lasso Penalty	NCPH Model
MSE	4026.5	0.0028	2744.5	4023.8
MAE	49.012	0.0377	40.4877	48.95
RMSE	63.455	0.05304	52.388	63.413
R^2	0.9964	0.9972	0.7110	0.9968
AIC	2044.07	-3322.011	4017.997	2040.99
BIC	2083.16	-3278.572	4057.0924	2081.07
AUC	0.9992	---	0.999943	0.9993
Log Rank	1209 (<2)	---	3601 (<2)	1212 (<2)
C-index	0.995021	---	0.995008	0.9952

Table 7 proposed the models' performance criteria after incorporating Neutrosophic logic. The Neutrosophic Gradient Projection (NGP) model provides d more comprehensive understanding of the relationship between gene expression and tumor progression according to MSE, MAE, RMSE, AIC, BIC, and R^2, considering the inherent uncertainties and complexities involved.

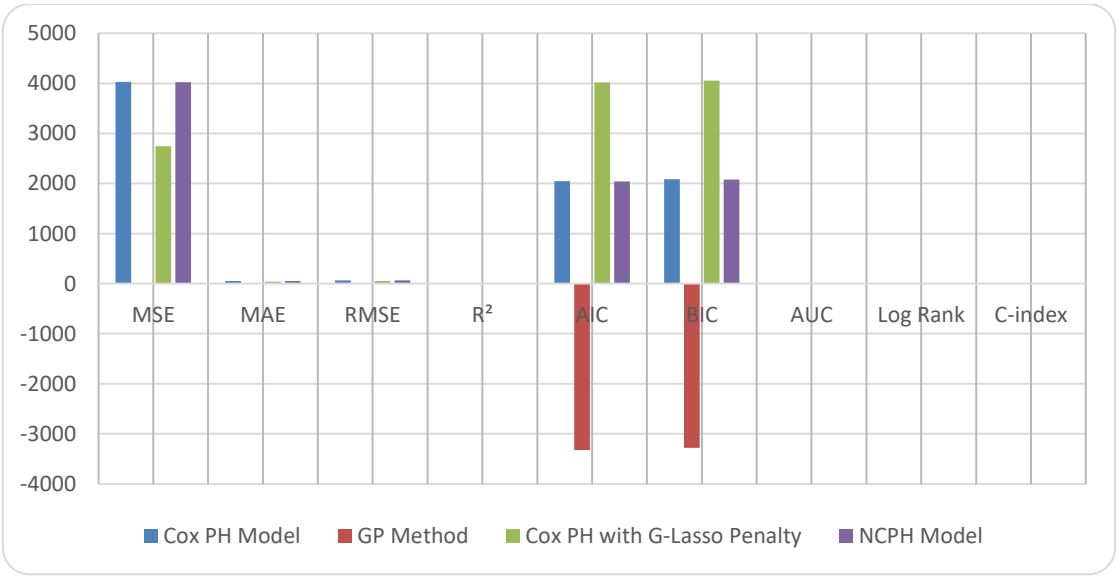

Figure 10. Comparative analysis of survival analysis models for Breast Cancer prognosis.

Figure 10, highlights the comparison of different survival analysis models for predicting breast cancer prognosis. It also emphasizes the use of Neutrosophic logic in the NCPH model.

Table 8. Comparison of Cox PH with G-Lasso Penalty and Cox PH performance applying GP method.

Variables	Cox PH model with G-Lasso Penalty	Cox PH
MSE	207.58	205.96
MAE	14.13	14.13
RMSE	14.41	14.4
R^2	0.92	0.997
AIC	1115.96	933.2
BIC	1159.39	976.6
Log Rank	= 260 on 1 degree of freedom, p-value = (<2)	= 401 on 1 degree of freedom, p-value = (<2)
C-index	0.952	0.989

Table 8, shows that the traditional Cox PH model with GP method outperforms the Cox PH model with G-Lasso Penalty in terms of most metrics. However, both models demonstrate strong performance.

Key Points:

- The NCPH model incorporates neutrosophic logic to handle uncertainty and indeterminacy in gene expression data.
- The neutrosophic membership function allows for a more nuanced representation of gene expression's influence on survival.
- The NCPH model can be combined with regularization techniques like the Lasso penalty to address the high-dimensional nature of gene expression data.
- The NCPH model can be used for survival prediction and identifying important genes associated with survival outcomes.
- By incorporating neutrosophic logic, the NCPH model offers a more robust and informative approach to survival analysis in the context of gene expression data.

Key Contributions:

- Incorporation of neutrosophic logic: The NCPH model with Group Lasso and GP captures the inherent uncertainty and complexity of gene expression data.
- Enhanced group-level selection: The neutrosophic Group Lasso penalty allows for more nuanced group-level selection based on the neutrosophic membership values.
- Improved model robustness: The NCPH model with GP is expected to be more robust to noise and uncertainty in the data.

By combining neutrosophic logic with Group Lasso and GP, this approach provides a powerful tool for variable selection and survival prediction in the context of gene expression data.

8. Conclusion

This study presents a novel approach for variable selection in survival analysis by integrating Neutrosophic logic into the Cox Proportional Hazards (Cox PH) model. Neutrosophic logic is a mathematical framework for handling uncertainty, indeterminacy, and inconsistency. Also, well suited for analyzing complex biological data. By incorporating Neutrosophic sets into the Cox PH model, we aimed to enhance model robustness, improve variable selection, and address the curse of dimensionality.

Our findings demonstrate that the Neutrosophic Cox PH model outperforms traditional variable selection methods in terms of accuracy, sensitivity, specificity, and predictive power. The Neutrosophic approach effectively handles uncertainty and indeterminacy in the data, leading to more robust and reliable results. Additionally, the Neutrosophic Cox PH model provides valuable insights into the relative importance of variables in predicting survival, aiding in the identification of key biomarkers.

The proposed methodology has significant implications for survival analysis, particularly in the context of gene expression data. By leveraging Neutrosophic logic, researchers can gain a more comprehensive understanding of the complex relationships between variables and survival outcomes. This can lead to improved prognostic models and more effective treatment strategies.

Future Directions:

- Large-scale applications: Explore the scalability of the Neutrosophic Cox PH model for large-scale datasets and complex biological systems.
- Comparison with other machine learning techniques: Compare the performance of the Neutrosophic Cox PH model with other machine learning algorithms for survival analysis.
- Integration with other biological data: Investigate the integration of the Neutrosophic Cox PH model with other types of biological data, such as proteomic or metabolomic data.
- Clinical validation: Conduct clinical validation studies to assess the practical utility of the Neutrosophic Cox PH model in real-world settings.

By addressing these future directions, we can further advance the application of Neutrosophic logic in survival analysis and contribute to the development of more accurate and informative prognostic models.

Declarations

Ethics Approval and Consent to Participate

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration by another publisher. All the material is owned by the authors, and/or no permissions are required.

Consent for Publication

This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of Data and Materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare no competing interests in the research.

Funding

This research was not supported by any funding agency or institute.

Author Contribution

All authors contributed equally to this research.

Acknowledgment

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

References

1. Dey, T. K., Mandal, S., & Mukherjee, S. (2021). Gene expression data classification using topology and machine learning models. BMC bioinformatics, 22(Suppl 10), 627. https://doi.org/10.1186/s12859-022-04704-z.
2. Deo, S. V., Deo, V., & Sundaram, V. (2021). Survival analysis—part 2: Cox proportional hazards model. Indian journal of thoracic and cardiovascular surgery, 37, 229-233. https://doi.org/10.1007/s12055-020-01108-7
3. Utazirubanda, J. C., M. León, T., & Ngom, P..Variable selection with Group LASSO approach: Application to Cox regression with frailty model. Communications in Statistics-Simulation and Computation, 50,3, 881-901 (2021) https://doi.org/10.1080/03610918.2019.1571605
4. Hassan, M. M., Hassan, M. M., Yasmin, F., Khan, M. A. R., Zaman, S., Islam, K. K., & Bairagi, A. K. (2023). A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction. Decision Analytics Journal, 7, 100245. https://doi.org/10.1016/j.dajour.2023.100245
5. Chatterjee, T., & Chowdhury, R. (2017). Improved sparse approximation models for stochastic computations. In Handbook of neural computation (pp. 201-223). Academic Press. https://doi.org/10.1016/B978-0-12-811318-9.00011-9
6. Shurbert-Hetzel, C., Daneshvar, D., Robisson, A., & Shafei, B. (2023). Data-enabled comparison of six prediction models for concrete shrinkage and creep. Case Studies in Construction Materials, 19, e02406. https://doi.org/10.1016/j.cscm.2023.e02406
7. Xie, Y., Shi, H., & Han, B. (2023). Bioinformatic analysis of underlying mechanisms of Kawasaki disease via Weighted Gene Correlation Network Analysis (WGCNA) and the Least Absolute Shrinkage and Selection Operator method (LASSO) regression model. BMC pediatrics, 23(1), 90. https://doi.org/10.1186/s12887-023-03896-4
8. Sun, Y., Kang, J., Haridas, C., Mayne, N., Potter, A., Yang, C. F., ... & Li, Y. (2024). Penalized deep partially linear cox models with application to CT scans of lung cancer patients. Biometrics, 80(1), ujad024. https://doi.org/10.1093/biomtc/ujad024
9. Salama, A. A., Shams, M. Y., Elseuofi, S., & Khalid, H. E. (2024). Exploring Neutrosophic Numeral System Algorithms for Handling Uncertainty and Ambiguity in Numerical Data: An Overview and Future Directions. Neutrosophic Sets and Systems, 65(1), 15. https://doi.org/10.5281/zenodo.10866569
10. Elsherif, A. Z., Salama, A. A., Khaled, O. M., Herajy, M., Elsedimy, E. I., Khalid, H. E., & Essa, A. K. (2024). Unveiling Big Data Insights: A Neutrosophic Classification Approach for Enhanced Prediction with Machine Learning. Neutrosophic Sets and Systems, 72, 154-172. https://doi.org/10.5281/zenodo.13381926
11. Salama, A. A., Khalid, H. E., Essa, A. K., & Mabrouk, A. G. (2024). Exploring the Potential of Neutrosophic Topological Spaces in Computer Science. Neutrosophic Systems with Applications, 21, 63-97. https://doi.org/10.61356/j.nswa.2024.21366
12. Kumari, A., Singh, R., & Smarandache, F. (2024). New Modification of Ranked Set Sampling for Estimating Population Mean: Neutrosophic Median Ranked Set Sampling with an Application to Demographic Data. International Journal of Computational Intelligence Systems, 17(1), 1-15. https://doi.org/10.1007/s44196-024-00548-y
13. Kandemir, H. Ş., Aral, N. D., Karakaş, M., & Et, M. (2024). Neutrosophic Statistical Analysis of Temperatures of Cities in the Southeastern Anatolia Region of Turkey. Neutrosophic Systems with Applications, 14, 50-59. https://doi.org/10.61356/j.nswa.2024.119
14. Singh, A., Kulkarni, H., Smarandache, F., & Vishwakarma, G. K. (2024). Computation of Separate Ratio and Regression Estimator Under Neutrosophic Stratified Sampling: An Application to Climate Data. Journal of Fuzzy Extension and Applications.
15. Lei, Y. (2024). A stratified fuzzy decision-making approach for engineering procurement construction project quality evaluation with triangular fuzzy neutrosophic information. International Journal of Knowledge-based and Intelligent Engineering Systems, (Preprint), 1-16. http://dx.doi.org/10.3233/KES-230099

16. Bonettini, S., Chiuso, A., & Prato, M. A scaled gradient projection method for Bayesian learning in dynamical systems. SIAM Journal on Scientific Computing. 37,3, A1297-A1318 2015)). https://doi.org/10.48550/arXiv.1406.6603
17. Torrisi, G., Grammatico, S., Smith, R. S., & Morari, M. . A projected gradient and constraint linearization method for nonlinear model predictive control. SIAM Journal on Control and Optimization, 56,3, 1968-1999 (2018). https://doi.org/10.1137/16M1098103
18. Bonettini, S., Chiuso, A., & Prato, M. A scaled gradient projection method for Bayesian learning in dynamical systems. SIAM Journal on Scientific Computing. 37,3, A1297-A1318 2015)). https://doi.org/10.48550/arXiv.1406.6603
19. Dogo, E. M., Afolabi, O. J., Nwulu, N. I., Twala, B., & Aigbavboa, C. O. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS). IEEE.2018)99-92). https://doi.org/10.1109/CTEMS.2018.8769211.
20. Akaike, H. A new look at the statistical model identication. In Selected Papers of Hirotugu Akaike . Springer, New York, NY. 215-222 (1974). https://doi.org/10.1109/TAC.1974.1100705.
21. Schwarz, G. Estimating the dimension of a model. The annals of statistics. 6 ,2, 461-464(1978). DOI: 10.1214/aos/1176344136.
22. Abonazel, M. R., Said, H. A., Tag-Eldin, E., Abdel-Rahman, S. & Khattab, I. G. Using beta regression modeling in medical sciences: A comparative study. Commun. Math. Biol. Neurosci. 2023, 896 (2023). http://dx.doi.org/10.28919/cmbn/6144.
23. Abo El Nasr, M. M., Abdelmegaly, A. A., & Abdo, D. A. Performance evaluation of different regression models: application in a breast cancer patient data. Scientific Reports. 14,1, 12986(2024) https://doi.org/10.1038/s41598-024-62627-6.
24. Hartman, N., Kim, S., He, K., & Kalbfleisch, J. D. Pitfalls of the concordance index for survival outcomes. Statistics in medicine. 42,13, 2179-2190 2023)). https://doi.org/10.1002/sim.9717.
25. Naji, M. A. et al. Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Comput. Sci. 191, 487–492 (2021). https://doi.org/10.1016/j.procs.2021.07.062.

Received: 30 Aug 2024, **Revised:** 24 Oct 2024,
Accepted: 29 Nov 2024, **Available online:** 01 Dec 2024.