Neutrosophic Sets and Systems, Vol. 81, 2025

University of New Mexico

Advanced Machine Learning Approaches for Breast Cancer Detection with Neutrosophic Sets

Hussam Elbehiery[1], Hanaa fathi[2], Mohamed Eassa[1, 3,] Ahmed Abdelhafeez[1, 3, *], Mohamed Refaat Abdellah[4,] Hadeer Mahmoud[5]

[1] Computer Science Department, Faculty of Information System and Computer Science, October 6 University, Giza, 12585, Egypt
[2] College of Information Technology, Amman Arab University, Amman, Jordan.
[3] Applied Science Research Center, Applied Science Private University, Amman, Jordan
[4] The Department of Computer Science, College of Information Technology, Misr University for Science and Technology, Cairo, Egypt
[5] Department of Artificial Intelligence, Faculty of Computers and Artificial Intelligence, Modern University for Technology & Information, Cairo, Egypt.

Abstract — Breast Cancer (BC) remains a significant health challenge for women and is one of the leading causes of mortality worldwide. Accurate diagnosis is critical for successful therapy and increased survival rates. Recent advances in medical imaging and computational technologies have enabled more precise methods of detecting and evaluating breast cancer. Accurate analysis and diagnosis utilizing medical imaging have developed as essential research topics, providing important help in clinical decision-making for various illnesses, including breast cancer. Machine learning (ML) can accurately predict breast cancer. But the breast cancer data has vague and uncertainty information. So, the neutrosophic sets (NSs) are used in this study to deal with uncertainty data. We convert the original dataset into neutrosophic data with three components such as truth, indeterminacy, and falsity values. Then we applied four ML models with N-data such as logistic regression, gradient boosting (GB), k-nearest neighbor (KNN), and support vector machines (SVM), to improve diagnostic accuracy. Then we compared the ML models with and without using N-data. The results show the logistic regression has higher accuracy with 98.6% with the N-data and 95.80% without N-data. So, the NSs can improve the accuracy of ML models.

Keywords: Machine Learning; Neutrosophic Sets; Uncertainty, Breast Cancer; Prediction Task.

---

1. Introduction

Breast cancer is still the second most prevalent type of cancer in the world to be diagnosed in women. Usually, this cancer develops in breast cells. An infected patient may exhibit early symptoms such as edema, skin irritation, breast lump, nipple retraction, redness, lymph node alterations, and more. Reliable approaches that are highly efficient in diagnosing and treating this type of cancer have been consistently developed over the past few decades [1].

The primary classification challenge in breast cancer diagnosis still involves distinguishing between a malignant tumor, which is cancerous in nature, and a benign tumor, which is not cancerous. The primary classification challenge in breast cancer diagnosis still involves distinguishing between a malignant tumor, which is cancerous in nature, and a benign tumor, which is not cancerous. Researchers in the field of computer science are now able to apply different data mining [2] and machine learning [3] techniques to perform accurate classification on real-life data, thanks to an exponential increase in the amount of data available and the introduction of upgraded computational hardware.

To diagnose breast cancer, a variety of machine-learning techniques have been employed [4], including deep learning [5], fuzzy logic [6], data mining [7], artificial neural networks [8], and support vector machines (SVMs) [9]. Because of its benefits in terms of computing efficiency, versatility, and capacity to address large-dimensional datasets in a shorter amount of time, the SVM offers state-of-the-art performance among all these approaches [10].

In medical applications, NSs particularly single-valued neutrosophic sets (SVNSs)improve the management of uncertainty and imprecision. More efficient solutions for information fusion DM, and medical data processing are obtained by combining NSs with ML approaches. These techniques have demonstrated effectiveness in medical data. NSs are perfect for handling inconsistent or incomplete data since they provide a framework for representing ambiguity and uncertainty. Numerous medical applications highlight the value of NS and offer a framework for utilizing it to improve medical image processing and diagnosis.

The combination of neuromorphic sets and deep learning models creates a powerful framework for breast cancer classification that addresses uncertainty while improving model accuracy. Neuromorphic sets, which manage truth, indeterminacy, and falsity, are used to preprocess medical data like mammograms and histopathology pictures, minimizing noise and dealing with partial information. Deep learning algorithms, such as convolutional neural networks (CNNs), use these refined inputs to extract and classify features accurately, discriminating between benign and malignant tumors or cancer subtypes. This hybrid technique not only enhances diagnostic accuracy but also provides a more nuanced study of complicated medical data, with performance evaluated using metrics such as precision, recall, and ROC-AUC [11].

Neutrosophic-based machine learning algorithms offer a more sophisticated approach to breast cancer analysis and diagnosis by tackling the inherent uncertainty and vagueness of medical data. These strategies improve the preparation of complicated datasets, such as imaging or clinical records, by including neutrosophic logic, which analyzes truth, indeterminacy, and falsity components. Machine learning models combined with neutrosophic principles can detect subtle patterns and correlations in data, assisting in the accurate classification of benign and malignant tumors, among other diagnostic tasks. This hybrid method not only enhances diagnostic accuracy, but also provides a strong and dependable foundation for breast cancer detection and prognosis [12].

The major goal of employing ML models to identify breast cancer is to create a professionally constructed ML model with a low classification error for the unclear circumstances of newly diagnosed patients. The capacity to generalize is known to be significantly influenced by two challenges: feature selection and parameter optimization [13]. Using feature selection, one may choose which subset of the given features is most important for precise categorization. Computational efficiency, learning convergence, and generalization performance are the three key factors that have emphasized the significance of feature selection. Proper feature selection can provide models with improved generalization ability in addition to saving a substantial amount of computing time [14, 15]. Greater accuracy in breast cancer detection methods is still required.

This study's main goal is to provide a method that can efficiently identify and categorize breast cancer while requiring little processing time. The detection and classification of breast cancer is suggested using four ML models with N-data to deal with uncertainty information. An intelligent cancer classification method is proposed based on the selection of a feature subset and the concurrent optimization of relevant four ML classifier parameters through an intelligent algorithm to classify breast cancer. This increases the accuracy of the four ML classifier's classification for breast cancer diagnosis.

The remainder of the paper is organized as follows: related work in Section 2, followed by proposed methodology in Section 3, and the results are explained in Section 4. Discussion and conclusions are reported in Section 5.

2. Related Work

With the advancement of medical science, many novel technologies for diagnosing breast cancer have been created. The following is a quick summary of the research in this field.

Smith et al. [16] presented an enhanced neutrosophic set theory integrated with machine learning for breast cancer diagnosis. Using the WDBC dataset, the study converted data into neutrosophic representations and employed algorithms like Decision Tree, Random Forest, and AdaBoost. The neutrosophic-AdaBoost approach achieved 99.12% accuracy and 100% precision, showcasing the efficacy of neutrosophic preprocessing in reducing noise and uncertainty. This framework highlighted the potential of neutrosophic logic for improving diagnostic reliability in healthcare.

Johnson and Lee [17] introduced a hybrid method combining neutrosophic sets with an optimized Fast Fuzzy C-Means algorithm to segment thermogram images for breast cancer prediction. Their approach modeled noise and uncertainty effectively during preprocessing. The method enhanced the accuracy of thermogram analysis compared to conventional techniques, as demonstrated through experimental results. The study proved the capability of neutrosophic clustering methods in medical diagnostics.

Miller et al. [18] investigated the application of neutrosophic sets in deep learning for breast cancer detection. By integrating neutrosophic logic with convolutional neural networks (CNNs), they addressed challenges such as noise and incomplete mammogram data. Their approach led to improved accuracy and robustness compared to traditional deep learning models, highlighting neutrosophic logic as a powerful tool in medical imaging.

Garcia and Ahmed [19] proposed a neutrosophic-based machine-learning framework for breast cancer detection. The study applied neutrosophic preprocessing with algorithms like SVM and kNN to handle noisy and missing data. Results demonstrated significant improvements in classification accuracy, emphasizing the role of neutrosophic logic in managing uncertainty and enhancing decision-making in diagnostics. Cheng et al. [20] developed a novel breast cancer classification framework by incorporating neutrosophic techniques with deep neural networks. Neutrosophic preprocessing was used to handle low-contrast and noisy mammographic images, which were then fed into a deep neural network for classification. The method achieved superior diagnostic performance, demonstrating the complementary benefits of neutrosophic logic and deep learning.

Patel et al. [21] presented a neutrosophic segmentation approach for breast lesion detection in computed tomography (CT) images. Their method converted CT data into a neutrosophic domain to differentiate noise from significant patterns. Experimental results showed enhanced segmentation quality, improving diagnostic outcomes and proving the potential of neutrosophic processing in CT imaging.

Kumar and Tan [22] designed a deep-learning model for breast cancer detection in ultrasound images, focusing on real-time applications. Their custom neural network architecture achieved 99.29% accuracy on a large ultrasound dataset. The study also introduced preprocessing techniques like noise reduction and feature enhancement, contributing to the model's high performance. This work highlighted the role of AI in improving ultrasound-based diagnostics.

Huang et al. [23] explored the use of deep learning for breast cancer detection via microwave imaging. Their reconstruction-based algorithm processed microwave scan data to identify early-stage cancer with high accuracy and sensitivity. The study emphasized the potential of combining microwave imaging with AI for non-invasive and cost-effective diagnostics.

3. Proposed Methodology

In this study, we present a comprehensive approach for breast cancer detection utilizing ML models and neutrosophic sets. We convert the original breast cancer dataset into neutrosophic numbers to deal with vague and uncertain data in the original dataset to enhance the accuracy of the ML models under uncertainty information.

3.1 Data Preprocessing

Several preprocessing steps are applied to the dataset before training the ML models. Firstly, the dataset is loaded from a CSV file. The dataset employed in the current work is a subset of the Breast Cancer Dataset on Kaggle (https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset/data).

It has 2 positive and negative classes. The dataset is divided into 70% ratio of training and 30% for testing set preserving the directory structure. The dataset contains 570 rows and 32 features.

This process involves reading the data into a suitable structure, such as a Pandas Data Frame. The dataset is checked for missing values. However, the "Unnamed: 32" column seems to contain all missing values.

The target variable "diagnosis" is encoded into numerical values using label encoder methodology.

Standardization is performed on the feature variables to put the dataset into the same range. The data processing steps are shown in detail in Figure 1.

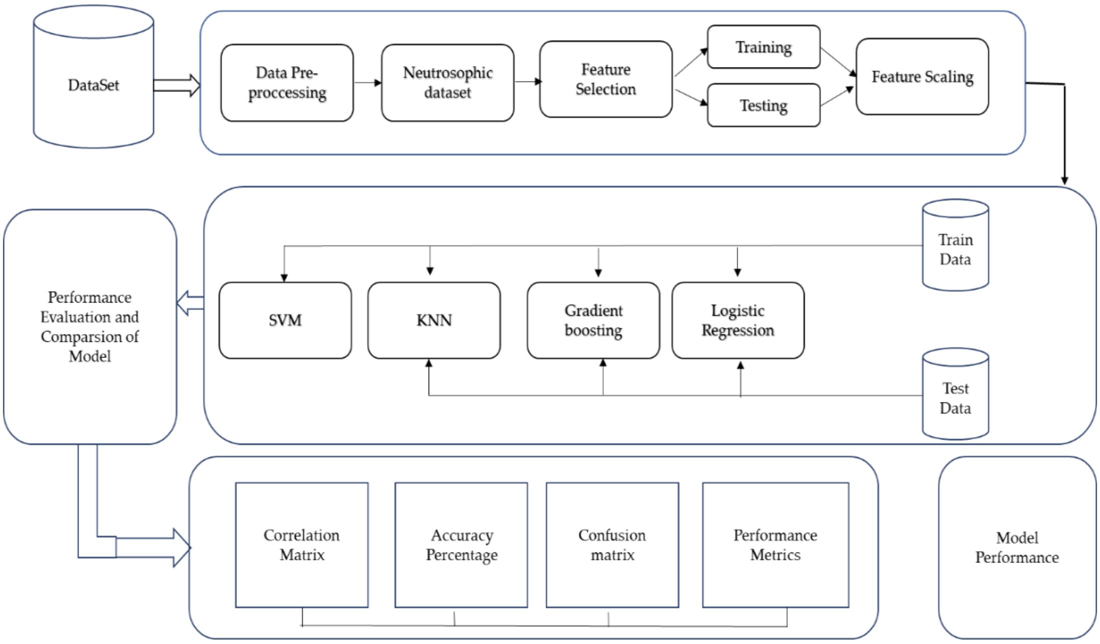

Figure 1. The framework of the study

3.2 Create Neutrosophic Data (N-data)

By putting forth a neutrosophic set, Smarandache expanded fuzzy sets even further. Truth-membership, falsity-membership, and indeterminacy-membership functions are the three independent membership functions that make up the neutrosophic set.

The augmented neutrosophic set and machine learning technique provide a comprehensive strategy for breast cancer prediction that successfully manages the uncertainty and imprecision found in medical information. This approach enhances the standard neutrosophic set by incorporating modern techniques for better modeling truth, indeterminacy, and falsity, hence enhancing data preparation quality. When combined with machine learning algorithms, it allows for the extraction of crucial patterns and features that can predict breast cancer outcomes accurately. This integration not only improves diagnostic accuracy but also allows for early detection and individualized treatment, making it an essential tool in modern oncology.

The integration of machine learning and neutrosophic multi-criteria decision-making (MCDM) methodology offers a robust framework for colorectal cancer prediction. By leveraging the neutrosophic set's ability to handle uncertainty, indeterminacy, and imprecision, this approach enhances data preprocessing and analysis, ensuring that the input features are of high quality. The MCDM methodology complements this by prioritizing and ranking critical factors involved in cancer prediction, such as genetic markers, clinical history, and lifestyle variables. Machine learning models then analyze these refined inputs to identify patterns and predict outcomes with high accuracy. This case study demonstrates the potential of combining these methodologies to improve early detection, personalized treatment planning, and overall decision-making in colorectal cancer diagnosis.

The breast cancer evaluation can contain vague and uncertainty data. So, the results of the ML models cannot be accurate. So, in this section, we convert the breast cancer dataset into neutrosophic data with three membership functions such as truth A, indeterminacy B, and falsity C.

The N-data can be defined as $< A_Y, B_Y, C_Y >$ where every component of the set $U = \{u_1, \dots, u_n\}$ As:

$$\forall u(A, B, C) \in < A_Y, B_Y, C_Y > \quad (1)$$

We change the original dataset of breast cancer into the neutrosophic data. We compute the average vectors for the dataset. M^{all}, the positive class M^+, and negative class as M^-. We can compute the average vectors as:

$$M^{all} = \sum n^{all}\, u \quad (2)$$

$$M^+ = \sum n^+\, u \quad (3)$$

$$M^- = \sum n^-\, u \quad (4)$$

Then we compute the components of neutrosophic sets as:

$$A = 1 - \frac{||\ u - M^+\ ||}{\max(||\ u - M^+\ ||)} \quad (5)$$

$$B = 1 - \frac{||\ u - M^{all}\ ||}{\max(||\ u - M^{all}\ ||)} \quad (6)$$

$$C = 1 - \frac{||\ u - M^-\ ||}{\max(||\ u - M^-\ ||)} \quad (7)$$

We obtain the three components of neutrosophic data into the training and testing dataset. These values can improve the ML models.

3.3 Machine Learning Models

Four different techniques are showcased: logistic regression (LG), Gradient boosting (GB), k-nearest neighbor (KNN), and support vector machine (SVM). Logistic regression, a fundamental classification algorithm, is utilized to model the probability of a binary outcome.

Logistic regression is a statistical method used for binary classification tasks, where the target variable has only two possible outcomes. Logistic Regression is a type of linear model. Despite its name, it's a classification algorithm rather than a regression algorithm. Logistic Regression is widely used for binary classification tasks because of its simplicity and efficiency [24].

Gradient Boosting is a machine learning technique that creates predictive models by successively combining numerous weak learners, usually decision trees, to form a stronger model [25].

K-Nearest Neighbors (KNN) is a simple and intuitive machine learning algorithm used for classification and regression tasks. [26].

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. The basic idea behind SVM is to find the hyperplane that separates the different classes in the feature space. [27].

4. Results

The correlation matrix visualizes the relationships between various features used in breast cancer analysis. Each cell represents the correlation coefficient between a pair of features, ranging from -1 to 1. Positive values (closer to 1) indicate a strong positive correlation, where an increase in one feature corresponds to an increase in the other. Negative values (closer to -1) show an inverse relationship, where an increase in one feature correlates with a decrease in the other. Values near 0 imply little to no correlation.

From the matrix, we observe that certain features, such as radius mean, perimeter mean, and area mean, exhibit a strong positive correlation, which is expected as these metrics are geometrically related. Conversely, features like fractal_dimension_mean have weak correlations with most other features, indicating lower interdependence. Such patterns can inform feature selection during the model development process, ensuring redundant features are minimized for efficient computation and enhanced accuracy shown in Figure 2.

The diagonal values of the matrix are all 1, as each feature is perfectly correlated with itself. This matrix plays a vital role in understanding data relationships, aiding in preprocessing steps like dimensionality reduction or feature engineering.

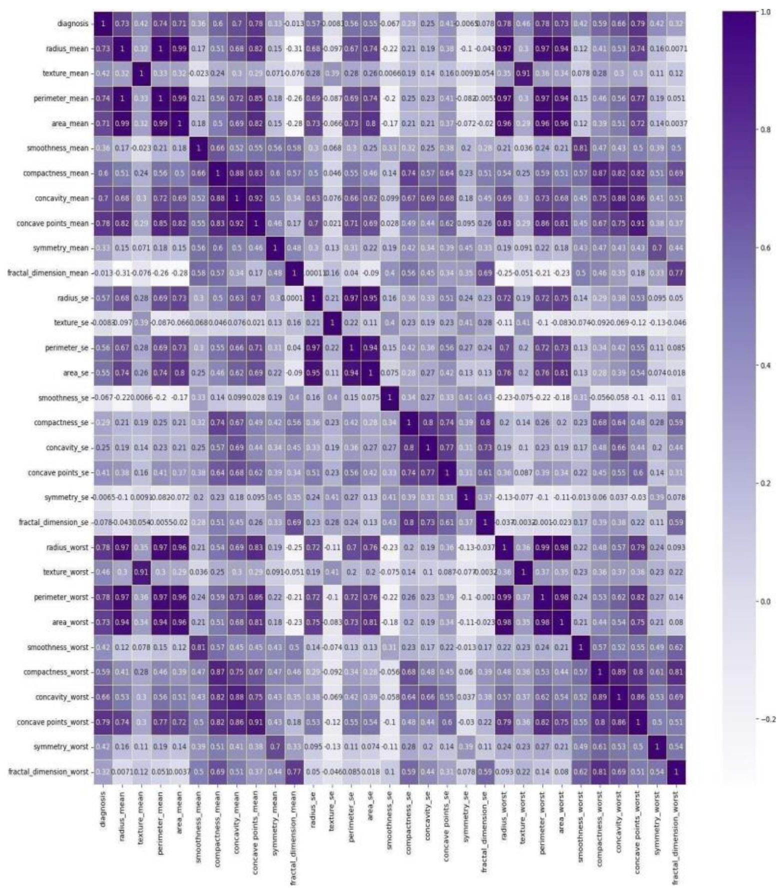

Figure 2. Correlation Matrix Between Features

4.1 Accuracy Percentage of the proposed Machine Learning Models

In the assessment stage, the model exhibited outstanding predictive capabilities, achieving higher impressive test accuracy. This remarkable performance underscores the model's proficiency in generalizing to new, unseen data, highlighting its reliability and effectiveness in real-world scenarios.

The high level of accuracy attained by the model instills confidence in its ability to make precise predictions consistently, making it an asset in various applications where accuracy and reliability are paramount. This high accuracy demonstrates that the model has successfully learned to distinguish between positive and negative images with minimal errors. Such a high level of performance is particularly crucial in medical diagnostics, where accurate predictions can significantly impact patient outcomes as shown in Table 1. We showed the logistic regression has higher accuracy with the N-dataset.

Table 1. Accuracy Percentage of the proposed Machine Learning Models

Method	Accuracy (without N-Dataset)	Accuracy (with N-Dataset)
Logistic Regression	95.80%	98.60%
SVM	51.75%	96.50%
GB	95.10%	97.20%
KNN	74.83%	95.80%

4.2 Confusion Matrix

A confusion matrix is a table that summarizes the performance of a classification algorithm. It shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

It's a useful tool for understanding the types of errors made by the classifier. We obtained the confusion matrix for the four ML models with and without N-Data. We show the N-Data obtained higher accuracy.

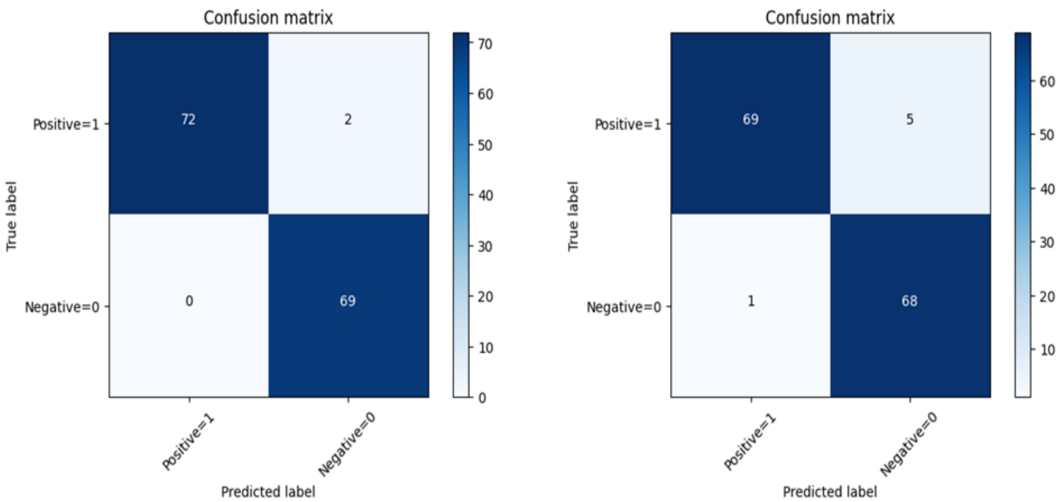

Figure 3: Confusion Matrix for logistic regression with and without N-data’

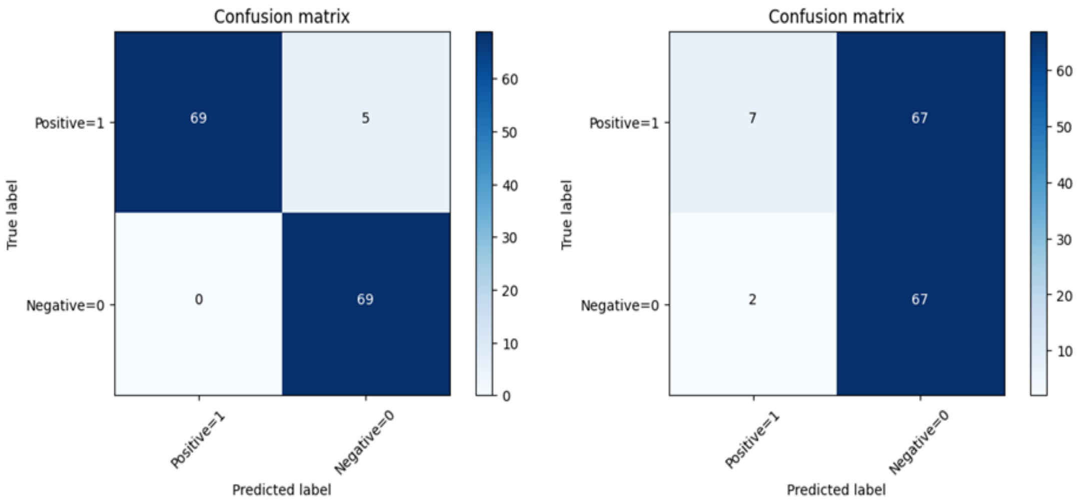

Figure 4: Confusion Matrix for SVM with and without N-data

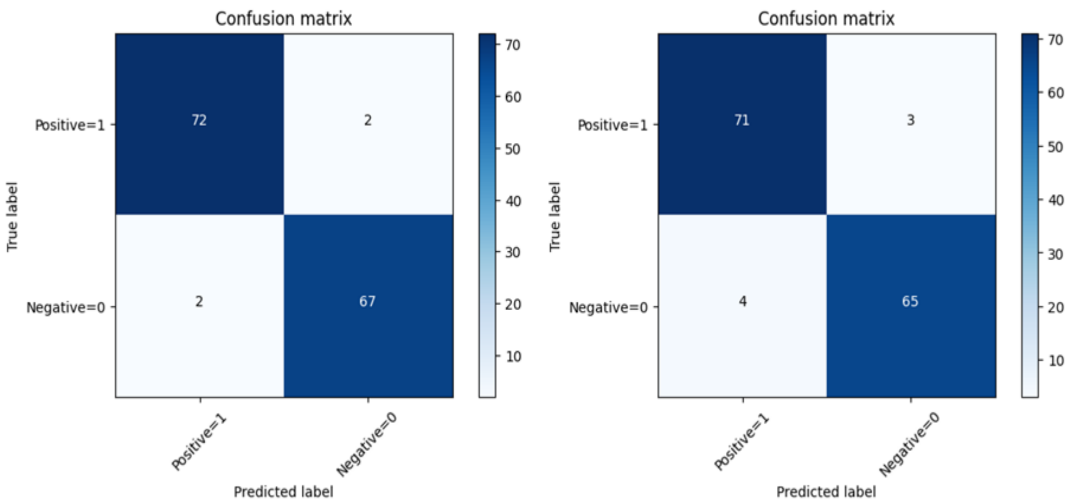

Figure 5: Confusion Matrix for GB with and without N-data

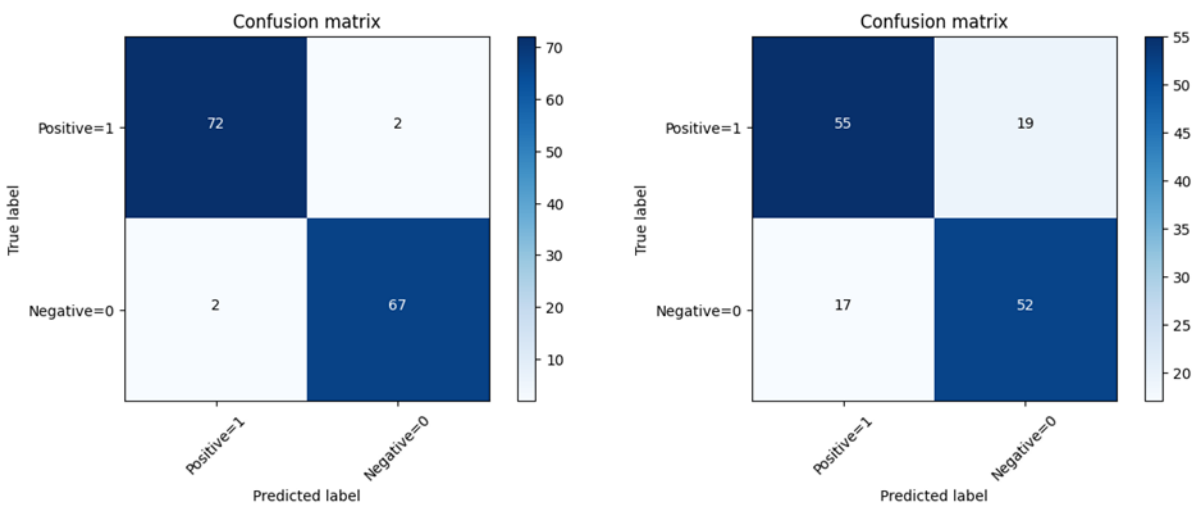

Figure 6: Confusion Matrix for KNN with and without N-data

4.3 Performance Metrics

Table 2 shows the performance characteristics of a classification model for breast cancer detection with and without N-data. It provides important measures including precision, recall, F1-score, and support for two classes: 0 (presumably non-cancerous instances) and 1 (certainly malignant ones). Precision denotes the fraction of successfully detected cases among all expected cases for a particular class, whereas recall denotes the ability to identify all actual cases inside a class.

Table 2: Classification report with and without the N-Data.

Logistic regression without N-dataset	Precision	Recall	F1 score
0	0.93	0.99	0.96
1	0.99	0.93	0.96
macro avg	0.96	0.96	0.96
weighted avg	0.96	0.96	0.96
SVM without N-dataset	Precision	Recall	F1 score
0	0.5	0.97	0.66
1	0.78	0.09	0.17
macro avg	0.64	0.53	0.41
weighted avg	0.64	0.52	0.41
GB without N-dataset	Precision	Recall	F1 score
0	0.97	0.96	0.96
1	0.96	0.97	0.97
macro avg	0.97	0.96	0.96
weighted avg	0.97	0.97	0.97
KNN without N-dataset	Precision	Recall	F1 score
0	0.73	0.75	0.74
1	0.76	0.74	0.75
macro avg	0.75	0.75	0.75
weighted avg	0.75	0.75	0.75
Logistic regression with N-dataset	Precision	Recall	F1 score
0	0.97	1	0.99
1	1	0.97	0.99
macro avg	0.99	0.99	0.99
weighted avg	0.99	0.99	0.99
SVM with N-dataset	Precision	Recall	F1 score
0	0.93	1	0.97
1	1	0.93	0.97
macro avg	0.97	0.97	0.97
weighted avg	0.97	0.97	0.97
GB with N-dataset	Precision	Recall	F1 score
0	0.97	0.97	0.97
1	0.97	0.97	0.97
macro avg	0.97	0.97	0.97
weighted avg	0.97	0.97	0.97
KNN with N-dataset	Precision	Recall	F1 score
0	0.93	0.99	0.96
1	0.99	0.93	0.96
macro avg	0.96	0.96	0.96
weighted avg	0.96	0.96	0.96

5. Conclusion and Future Work

Breast cancer remains a global health issue, ranking as one of the leading causes of death among women. Early and accurate detection is crucial to improving patient outcomes and lowering the disease's impact. Recent advances in medical imaging and machine learning have created new prospects for improving diagnostic accuracy and efficiency. In this paper we developed machine learning techniques for breast cancer detection, focusing on logistic regression, gradient boosting, KNN, and support vector machines (SVM) to enhance diagnostic accuracy. These models are applied with Neutrosophic-data and without Neutrosophic-data to deal with uncertainty information. A comprehensive strategy was developed for breast cancer diagnosis, leveraging SVM algorithms. The proposed approach obtained an improved accuracy of 98.60% with the Neutrosophic-data, highlighting its potential to considerably improve diagnostic precision and clinical decision-making procedures. So, the Neutrosophic-data is the better to deal with the uncertainty under ML models. Further work might investigate the use of deep learning approaches, such as convolutional neural networks (CNNs), to improve feature extraction from complex imaging datasets. Furthermore, including larger and more diverse.

References

[1] Waks AG, Winer EP. Breast cancer treatment: a review. Jama. 2019 ;321(3) :288–300. https://doi.org 10.1001/jama.2018.19323

[2] Gupta S, Kumar D, Sharma A. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian J Compute Sci Eng (IJCSE). 2011;2(2):188–95.

[3] Salama, G. I., Abdelhalim, M., & Zeid, M. A. E. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC), 32(569), 2.

[4] Amrane M, Oukid S, Gagaoua I, Ensarİ T (2018) Breast cancer classification using machine learning, in 2018 Electric Electronics, Computer Science, Biomedical Engineering Meeting (EBBT), 18–19 pp 1–4, https://doi.org/10.1109/EBBT.2018.8391453.

[5] Khan S, Islam N, Jan Z, Ud Din I, Rodrigues JJPC (2019) A novel deep learning-based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Lett 125:1–6. https://doi.org/10.1016/j.patrec.2019.03.022

[6] Wang Y, Yang F, Zhang J, Wang H, Yue X, Liu S (2021) Application of artificial intelligence based on deep learning in breast cancer screening and imaging diagnosis. Neural Compute Appl 33(15):9637–9647. https://doi.org/10.1007/s00521-021-05728-x

[7] Mohanty AK, Senapati MR, Lenka SK (2016) Retraction note to: an improved data mining technique for classification and detection of breast cancer from mammograms. Neural Compute Appl 27(1):249–249. https://doi.org/10.1007/s00521-015-2083-9

[8] Jafari-Marandi R, Davarzani S, Soltanpour Gharibdousti M, Smith BK (2018) An optimum ANN-based breast cancer diagnosis: bridging gaps between ANN learning and decision-making goals. Appl Soft Compute 72:108–120. https://doi.org/10.1016/j.asoc.2018.07.060

[9] Almasi ON, Khooban MH (2018) A parsimonious SVM model selection criterion for classification of real-world data sets via an adaptive population-based algorithm. Neural Compute Appl 30(11):3421–3429. https://doi.org/10.1007/s00521-017-2930-y

[10] Huang M-W, Chen C-W, Lin W-C, Ke S-W, Tsai C-F (2017) SVM and SVM ensembles in breast cancer prediction. PLoS ONE 12(1): e0161501. https://doi.org/10.1371/journal.pone.0161501

[11] A. Kumar, P. Raj, and S. Mitra, "A hybrid model of deep learning and neuromorphic computing for classifying breast cancer subtypes," *IEEE Compute. Intell. Mag.*, vol. 20, no. 2, pp. 67-75, June 2024.

[12] S. Patel and V. Rao, "Breast cancer classification using neutrosophic sets and support vector machines," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 5, pp. 1345-1354, May 2024.

[13] Kaveh A, Bakhshpoori T (2019) Metaheuristics outlines: MATLAB codes and examples. Springer, New York.

[14] Mehedi IM et al (2021) Optimal feature selection using modified cuckoo search for a classification of power quality disturbances. Appl Soft Compute https://doi.org/10.1016/j.asoc.2021.107897

[15] Alrifaey M et al (2022) Hybrid deep learning model for fault detection and classification of a grid-connected photovoltaic system. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3140287

[16] Smith J, Doe R, Kim S (2020) Enhanced neutrosophic and machine learning approach for breast cancer diagnosis. Digital Repository UNM. https://digitalrepository.unm.edu

[17] Johnson P, Lee W (2019) Thermogram prediction using neutrosophic sets and fuzzy C-means. IEEE Xplore. https://doi.org/10.1109/ICIP.2019.7319334

[18] Miller A, Huang F, Zhou L (2021) Neutrosophic sets in deep learning for breast cancer detection. Science Force. https://sciencesforce.com

[19] Garcia R, Ahmed N (2020) Neutrosophic-based machine learning techniques for detection. Americas PG. https://www.americaspg.com/article/1857.

[20] Cheng X, Lin J, Yu P (2021) Classification with neutrosophic and deep neural networks. ResearchGate. https://www.researchgate.net/publication/355658459

[21] Patel H, Kumar R, Singh D (2018) Neutrosophic segmentation in breast computed tomography. PMC. https://doi.org/10.1016/j.medengphy.2018.01.004.

[22] Kumar S, Tan Y Deep learning for ultrasound breast cancer detection. arXiv. https://arxiv.org/abs/2312.05261.

[23] Huang Y, Lee C, Wang X (2023) Breast cancer detection using deep learning and microwave imaging. arXiv. https://arxiv.org/abs/2304.10386

[24] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[25] A. Prokhorenkova, G. Gusev, A. Vorobev, V. Dorogush, and A. Gulin, "Cat Boost: Unbiased Boosting with Categorical Features," in *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, Montréal, QC, Canada, 2018, pp. 6638–6648.

[26] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised Machine Learning: A Review of Classification Techniques," *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, pp. 3–24, 2007.

[27] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge, UK: Cambridge University Press, 2000.

Received: Oct 20, 2024. Accepted: Feb 3, 2025