International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

AI-Driven Decision Support System for Early Breast Cancer Detection

Dr. G. Jothi[1], Dr. S. Nivetha[2], Dr. H Hannah Inbarani[3]

[1]Assistant Professor, Department of Computer Applications, Sona College of Arts and Science, Affiliate to Periyar University Salem. Tamil Nadu.

[2]Independent Researcher, Department of Computer Science, Salem. Tamil Nadu.

[3] Professor, Department of Computer Science, Periyar University, Salem. Tamil Nadu.

Abstract

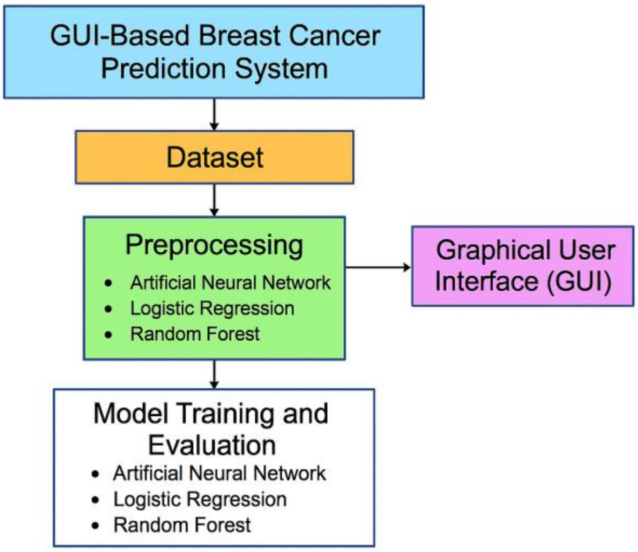
Breast cancer is one of the leading cancer types among women worldwide. The early and accurate diagnosis of breast cancer is crucial to important to increase the patient's life time. In this research a GUI-based breast cancer prediction system using Artificial Intelligence techniques to classify breast cancer as benign or malignant. The proposed model utilizes three different models, namely, Artificial Neural Network, Logistic Regression, and Random Forest, to evaluate the breast cancer prediction system. The model is trained with the Breast Cancer Wisconsin (Diagnostic) dataset, which consists of 30 numerical features derived from cell nucleus characteristics. The model's performance is assessed using standard evaluation metrics, including accuracy, precision, recall, F1 score, ROC curves, and confusion matrices. In addition to that, the trained models are integrated into a graphical user interface, enabling user interaction, real-time prediction, model comparison, and visualization of evaluation results. The empirical results show that a deep learning-based ANN model achieves superior performance in the accuracy of 97% when compared to other models. The proposed GUI-based breast cancer prediction system provides an efficient, interpretable, and user-friendly decision support tool for early breast cancer detection.

Keywords: Breast Cancer, Machine Learning, Deep Learning, ANN, Graphical User Interface

1. Introduction

Breast cancer is an illness defined by the excessive growth of abnormal cells in the breast tissue. It has been recorded as the most common form of cancer among women across the world (Fatima, et al., 2020). This illness constitutes a significant public health problem because of its high prevalence, mortality rate, and morbidity. Even though breast cancer mostly targets females, some male patients can be found. Early detection increases the survival chances, but failure to detect the condition early can lead to a late diagnosis and adverse consequences. For this reason, there is a need to develop accurate tools for diagnosing breast cancer to lower mortality rates. As per WHO, there have been about 2.3 million new diagnoses for breast cancer, leading to about 670,000 fatalities across the world. Regarding India, there have been roughly 192,020 incidences of new breast cancer cases and 98,337 mortalities. These figures account for the 26.6% percentage of female cancer cases in the nation. Moreover, the disease presents itself at a relatively young age in India, mainly in females aged between 40 and 50 years (Giaquinto, et al., 2022). These problems highlight the need for effective and supportive technology for early detection of high-risk cases. Over the last few decades, the use of artificial intelligence (AI) has gained significant popularity in designing intelligent systems for cancer detection (Abdikenov, et al., 2025). This technology provides better precision, decision support, efficiency, and speed in medical applications. Machine learning (ML) and deep learning (DL) are two major subfields of AI that help computer systems learn and predict results based on medical information without any kind of coding. This project designs a graphical user interface (GUI) based cancer prediction system with the help of machine learning and deep learning methods. In this research paper, the performance of popular AI systems such as Artificial Neural Network (ANN), Logistic Regression, and Random Forest will be analyzed with respect to breast cancer classification task. The objective of this project is to identify the effectiveness of traditional machine learning techniques as well as deep learning method by comparing them against each other. Furthermore, the integration of all the selected models within graphical user interface makes the system user-friendly and helps users to get real-time predictions about their tumor status along with some other useful statistics.

Figure 1: Framework of the GUI-Based Breast Cancer Prediction System



The architecture of the suggested GUI-based breast cancer prediction system is made up of various consecutive parts, aimed at ensuring high prediction quality and usability. First of all, the initial data set is subjected to preprocessing, which comprises data cleansing, scaling, and normalization. The next step involves training three AI models (ANN, Logistic Regression, and Random Forest) with the aid of the processed data set. These models will find discriminative features in the dataset, which will help identify whether breast tumors are benign or malignant. Afterward, the trained models will be assessed for their effectiveness in terms of accuracy, precision, recall, F1 score, ROC curves, and confusion matrices. Finally, the trained AI models along with the evaluation results will be incorporated into a GUI, created using Python. This way, end-users will be able to input data, load test samples, obtain predictions, compare model performance, and visualize ROC curves and confusion matrices.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive review of related work on breast cancer detection. Section 3 describes the proposed methodology and model architectures for the GUI-based prediction system. Section 4 discusses the experimental setup and results for breast cancer prediction. Finally, Section 5 concludes the paper by summarizing the key findings and future work.

2. Related Work

In recent times, various ML and DL algorithms have been used for breast cancer detection owing to their capacity to capture useful patterns within clinical and diagnostic data. Various research studies have been successful in implementing classification and deep neural approaches, yielding favorable results, especially on benchmarks like the WDBC dataset.

A research study by Rabiei et al., (2022) described an ML-based breast cancer prediction approach that utilized demographic, laboratory, and mammography data. In their experiments, they observed that the algorithm produced a high value of sensitivity at 95%, highlighting the efficiency of the method in detecting malignancy; however, its accuracy and AUC measures were found to be relatively low. Similarly, in a research study by Wei et al., (2023) the authors compared various approaches like Logistic Regression, Decision Trees, and Random Forest classifiers on the WDBC dataset. They found that Random Forests produced high levels of accuracy up to 95%, but there was no analysis of their real time applicability and visualizations of their model.

Other studies have moved towards utilizing deep learning techniques. Agrawal et al., (2025) studied the impact of machine learning and deep learning algorithms on breast cancer classification, considering various CNN-based architectures among others. It was noted that the use of deep learning resulted in an increase in the model's accuracy; however, there were no interactions with users or other components required to implement the system. Arravalli et al., (2025) have developed explainable machine learning models for breast cancer detection, which gave about 84% F1 Score using Random Forest algorithm. Despite the focus on interpretability via SHAP and LIME, this performance was worse compared to the one achieved by modern optimized ANN models. Hybrid and ensemble approaches are actively studied as well. Khan et al., (2026) offered a hybrid approach with clustering followed by a supervised classifier from among Gradient Boosting and Naïve Bayes.

The performance obtained was extremely accurate (99.3%) in some cases; however, the high complexity and a lack of means for deployment decreased the system's applicability in practice. Review articles highlight recent tendencies in the field. For example, Jothi et al., (2025) have conducted a literature review focusing on the application of ML, DL, and transfer learning to breast cancer detection. According to the article, CNN-based models can reach an accuracy of more than 99%. In addition, further research shows that classical machine learning models such as logistic regression can achieve high accuracy when used in highly structured datasets and perform better than DL models in terms of interpretability needed for clinical decision-making. Moreover, the use of Random Forest decreases the risk of overfitting and increases robustness, although it might be somewhat inferior in recall to ANNs in malignant tumor detection. In contrast, the present research distinguishes itself by integrating Artificial Neural Network, Logistic Regression, and Random Forest models into a GUI-based decision-support framework. The summarization of related work has been presented in Table 1.

Table 1. Summary of Related Work

Article	Dataset / Approach	Models Used	Key Performance Results	Limitations / Observations
Rabiei et al., (2022)	Demographic, laboratory, and mammography data	ML-based prediction model	Sensitivity: 95%	Low accuracy and AUC, limiting overall diagnostic reliability
Wei et al., (2023)	WDBC dataset	Logistic Regression, Decision Tree, Random Forest	Random Forest accuracy up to 95%	No real-time applicability study; lack of visualization
Agrawal et al., (2025)	Breast cancer image data	CNN-based deep learning models	Improved accuracy compared to ML	No user interaction or system-level implementation
Arravalli et al., 2025)	Breast cancer dataset	Explainable Random Forest (SHAP, LIME)	F1 Score ≈ 84%	Lower performance compared to optimized ANN models
Khan et al., (2026)	Hybrid learning framework	Clustering + Gradient Boosting / Naïve Bayes	Accuracy up to 99.3%	High complexity; no deployment mechanism
Jothi et al., (2025)	Literature survey	ML, DL, Transfer Learning (CNNs, LR, RF)	CNN accuracy > 99%	Notes trade-off between accuracy and interpretability
Present Research Work	WDBC dataset	ANN, Logistic Regression, Random Forest	Balanced performance across models	Novel GUI-based, clinically applicable framework

3. Methodology

In this research, a GUI-based breast cancer prediction system is designed through machine learning and deep learning in this study. In this section, the overall process used for the design of the GUI-based breast cancer prediction system is highlighted. There are five main processes involved in the development of the proposed GUI-based breast cancer prediction system: dataset description, data pre-processing, model development, and graphical user interface design.

3.1 Dataset

The breast cancer prediction system designed in this study makes use of the Breast Cancer Wisconsin (Diagnostic) Dataset, which is available on the UCI Machine Learning Repository (Asuncion et al., 2007). The dataset includes 569 instances that have been described with 30 numeric attributes that were extracted from the digitized image of fine needle aspirate (FNA) of breast mass. These numeric attributes include characteristics like radius, texture, perimeter, area, concavity, smoothness, and symmetry of the cell nuclei. Each instance has two possible classes – benign and malignant.

3.2 Data Preprocessing

Preprocessing is an essential stage that can significantly increase data quality and optimize model results. Within this experiment, the target class labels were numerically transformed, where 0 is the label for malignancy, and 1 is the class for benign tumors. The next stage involves splitting the sample into training (80%) and testing (20%) datasets. Because the feature values have different ranges and units, it is recommended to apply normalization with the help of z-score normalization to preprocess the information. Such a transformation makes sure that the mean value and variance of each attribute will be equal to 0 and 1 correspondingly. Normalization is particularly critical for gradient-based models and artificial neural networks. The scaled dataset was then used model training.

3.3 Model Development

In this research, three classification algorithms were tested to verify the efficiency of artificial intelligence methods for predicting breast cancer. These models are Artificial Neural Network, Logistic Regression, and Random Forest.

3.3.1 Artificial Neural Network (ANN)

The deep learning model employed in this research is a feed-forward Artificial Neural Network (ANN) designed to capture complex non-linear relationships among breast cancer diagnostic features (Agatonovic et al., 2000). The ANN architecture consists of an input layer, multiple fully connected hidden layers, and an output layer for binary classification. Each hidden layer utilizes the Rectified Linear Unit (ReLU) activation function, defined as:

$$ReLU(z) = max(0, z) \quad (1)$$

This activation function introduces non-linearity and helps mitigate the vanishing gradient problem. The output layer applies the sigmoid activation function, which maps the network output to a probability value between 0 and 1.

3.3.2 Logistic Regression

Logistic Regression (LaValley et al., 2008) was employed as a baseline machine learning model due to its simplicity, computational efficiency, and widespread use in medical research. It models the probability of a class label using a sigmoid function applied to a linear combination of input features. The logistic regression model is mathematically expressed as,

$$P(y = 1 \mid x) = \sigma(w^T x + b) \quad (2)$$

where w is the weight vector, x represents the input feature vector, b is the bias term, and σ denotes the sigmoid function.

3.3.3 Random Forest

Random Forest (Rigatti et al., 2017) is an ensemble learning algorithm that combines multiple decision trees to enhance classification performance and robustness. Each decision tree is trained on a randomly selected subset of training samples using bootstrap aggregation (bagging), while feature selection at each split further introduces randomness.

For a given input xxx, the Random Forest prediction is obtained by majority voting across all individual trees:

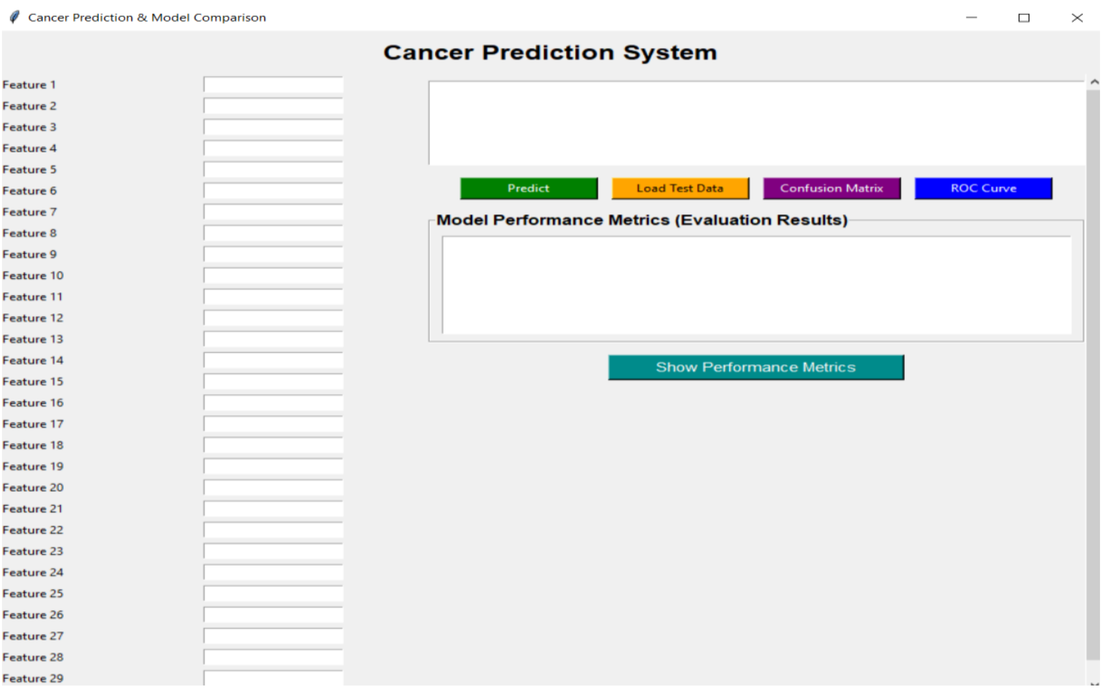
$$\bar{y} = mode\{T1(x), T2(x), \ldots, Tn(x)\} \quad (3)$$

where $Ti(x)T\_i(x)Ti(x)$ represents the prediction of the i^{th} decision tree.

3.4 GUI-Based System Implementation

To enhance usability and practical deployment, a graphical user interface (GUI) was developed using the Tkinter framework in Python. The GUI allows users to input clinical feature values, load test samples automatically, and obtain real-time predictions from all three trained models. The interface also displays model performance metrics dynamically, identifies the most efficient model, and provides visualization of ROC curves and confusion matrices through interactive pop-up windows. The Sample GUI environment for Breast Cancer Prediction System is shown in Figure 2.

Figure 2: GUI Environment for Breast Cancer Prediction System



4. Empirical Results Analysis

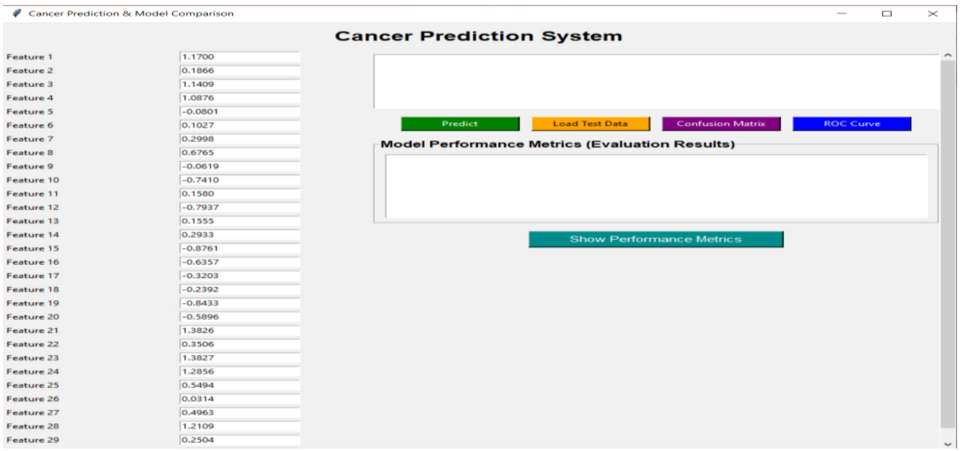
4.1 Experimental Setup

The experimental evaluation of the developed GUI application for breast cancer prediction was executed using the Python programming language and associated packages including Tensorflow/Keras, Matplotlib, and Tkinter, thus being guaranteed a compatibility with machine learning and deep learning toolkits. The experiment was performed on a Windows OS platform running on a 64-bit operating system. The computer configuration included an Intel® Core™ i3 7020U processor clocked at 2.30 GHz frequency with a memory of 4 GB RAM capacity. The neural network structure is trained using the Adam algorithm with adaptive learning rates. A total of 50 training epochs are set along with the batch size of 16. Binary cross entropy function is chosen to be applicable in medical cases when classifying patients as healthy or having a disease. For the logistic regression model, the liblinear solver is used with a limit of 1000 iterations. The number of decision trees for random forest is specified to 100, the Gini impurity split criterion.

4.2 Results and Discussion

The empirical study shows that ANN and Logistic Regression have better performance in comparison with Random Forest regarding recall and F1 score (Jothi et al., 2013). Though Logistic Regression has simple design, the model shows poor capacity of modeling nonlinear interactions between features. Random Forest model is rather robust and stable but it is inferior regarding its predictive power because of the ensemble technique of decision making. Implementation of empirical results in a GUI system makes it more user-friendly thanks to graphical representation of predictions and results in form of performance measures, ROC curves, and confusion matrices. The developed models were assessed on the basis of several performance criteria to ensure comprehensive comparison. The sample under analysis has been split randomly into 80% training data and 20% testing data to avoid biased evaluation. All models were trained using one preprocessed and standardized dataset in order to make comparison consistent and fair (Jothi et al., 2022). After that, the preprocessed dataset is input into the model. Figure 3 shows the breast cancer prediction system result for load the test data to assess models.

Figure 3: Load Test Data



The following metrics are evaluated to analyse the performance of the models. In a breast cancer prediction system, model performance is evaluated based on the classification of tumors as malignant (Class 0) or benign (Class 1).

$$Accuracy = \frac{Tumors\ correctly\ classified\ as\ Class\ 0\ and\ Class\ 1}{Total\ no.of\ samples} \quad - (4)$$

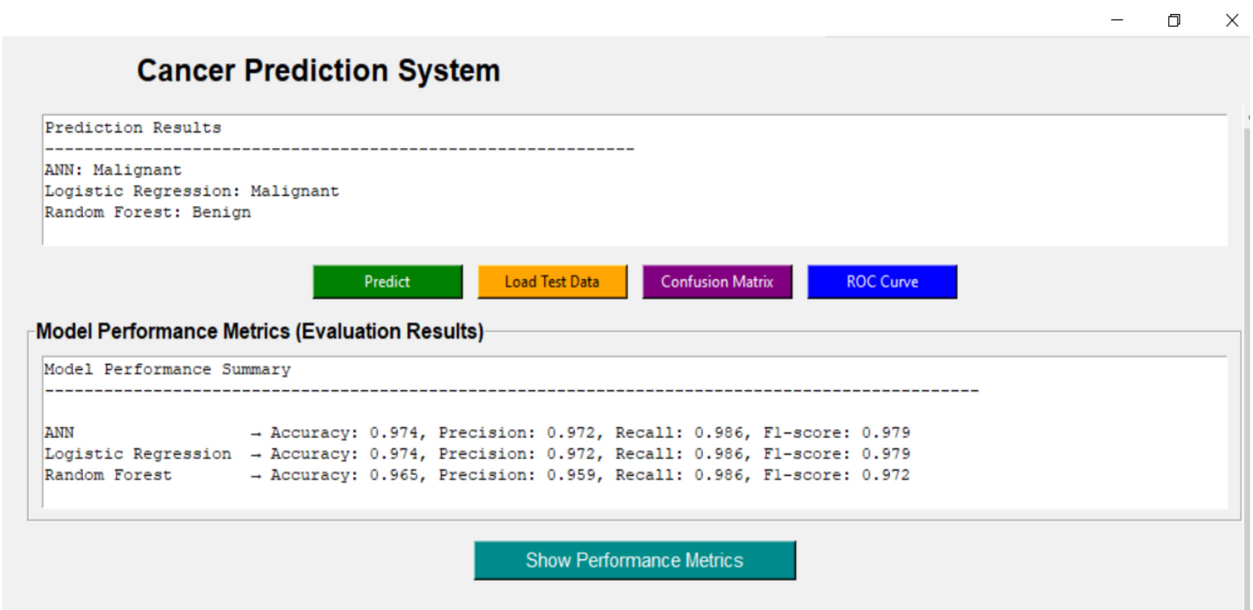
$$Precision = \frac{Class\ 0\ correctly\ classified\ as\ malignant}{Class\ 0\ incorrectly\ classified\ as\ malignant + Class\ 1\ incorrectly\ classified\ as\ malignant} \quad - (5)$$

$$Recall = \frac{Class\ 0\ correctly\ classified\ as\ malignant}{Class\ 0\ incorrectly\ classified\ as\ malignant + Class\ 0\ incorrectly\ classified\ as\ benign} \quad - (6)$$

$$F1 - Score\ = 2\ \times \frac{Precision\ \times Recall}{Precision + \ Recall} \quad - (7)$$

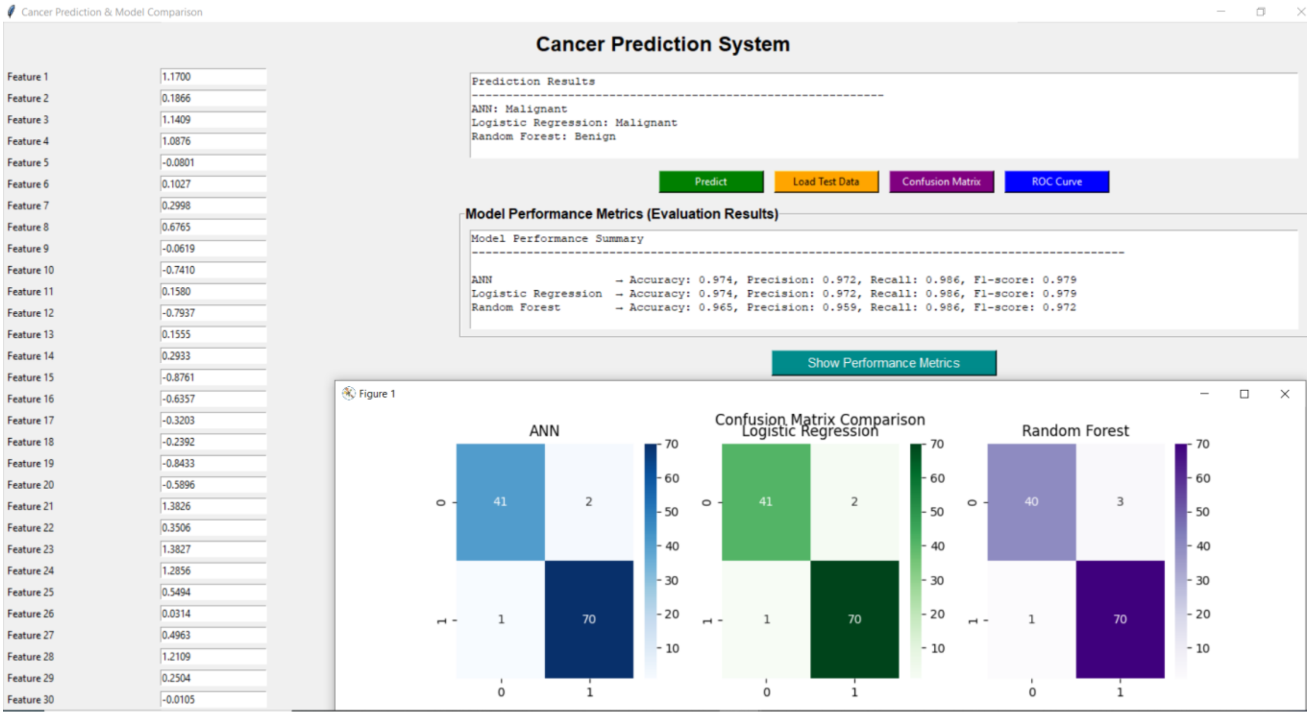
Figure 4 shows the performance results from all three models. The performance of the Artificial Neural Network (ANN) was very accurate with 97.4% of accuracy, 97.2% precision, 98.6% recall, and 97.9% F1 scores. The model performed well in the identification of malignant tumors because of its higher recall rate. On the other hand, the Logistic Regression had performance measures similar to those of ANN with accuracy of 97.4%, precision of 97.2%, recall of 98.6%, and 97.9% F1 score. It may be noted that despite the fact that the method used here is a linear algorithm, it performs well because of the discriminative power of the selected features through pre-processing and normalization. Lastly, the Random Forest had 95.6% accuracy, 95.8% precision, 97.2% recall, and 96.5% F1 score.

Figure 4: Model Prediction Results



A comparison of results shows that ANN and logistic regression yield less false-negative cases than the random forest. It is highly important in medical applications since the reduction of false negatives decreases the likelihood of misclassifying malignancies as benign. Figure 5 illustrates the confusion matrix for the breast cancer classification problem.

Figure 5: Confusion matrix for the breast cancer detection

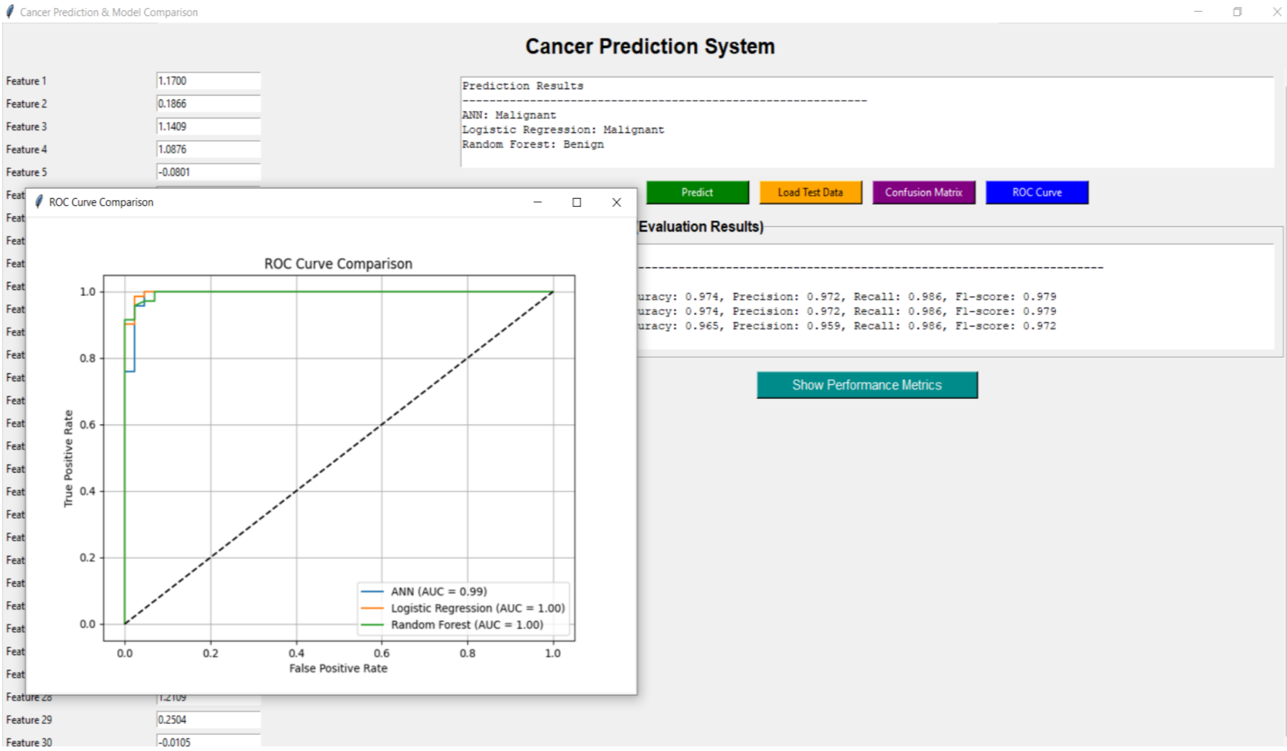


Based on the experimental results, one concludes that the models of machine learning and deep learning can be efficiently utilized for diagnosing breast cancer at its early stages. ANN is the most efficient classifier within the proposed scheme.

4.3. ROC Curve Analysis

The ROC curves also indicate the differences between the classifiers' performances. All three classifiers yield AUC scores close to 1. This means that all classifiers have a high discriminative ability. ANN and logistic regression demonstrate better ROC curves. It indicates their robustness for distinguishing the classes. Figure 6 illustrates the ROC Curve analysis for breast cancer detection.

Figure 6: ROC Curve analysis



5. Conclusion

The study successfully created a GUI based breast cancer prediction system using machine learning and deep learning algorithms. Three artificial intelligence models including Artificial Neural Network (ANN), Logistic Regression, and Random Forest were used for classifying tumors as either benign or malignant from the Breast Cancer Wisconsin (Diagnostic) dataset. According to the findings, the accuracy, precision, recall, and F1 scores were all found to be high. Specifically, ANN and Logistic Regression models have produced identical findings with respect to accuracy at 97.4%, precision at 97.2%, recall at 98.6%, and F1 score at 97.9%, showing reliability and consistency in their ability to diagnose tumors. However, the findings of the Random Forest model have been relatively less accurate with respect to accuracy and F1 score which stand at 95.6% and 96.5% respectively. ROC curves and confusion matrix also confirmed that ANN and Logistic Regression had excellent discriminative abilities. Their accuracies were higher and they exhibited very few cases of misclassifications especially when the actual classes were malignant.

Further enhancements of the study involve using convolutional neural networks (CNN) algorithms for diagnosing images of cancers. Explainable artificial intelligence (XAI) models may be employed in the process.

Conflict of Interest: All the authors no conflict of interest.

References

1. Fatima, N., Liu, L., Hong, S., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. IEEE Access, 8, 150360-150376.

2. Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., ... & Siegel, R. L. (2022). Breast cancer statistics, 2022. CA: a cancer journal for clinicians, 72(6), 524-541.
3. Abdikenov, B., Zhaksylyk, T., Shortanbaiuly, O., Orazayev, Y., Makhanov, N., Karibekov, T., ... & Seitova, A. (2025). Future of breast cancer diagnosis: A review of DL and ML applications and emerging trends for multimodal data. IEEE Access.
4. Rabiei, R., Ayyoubzadeh, S. M., Sohrabei, S., Esmaeili, M., & Atashi, A. (2022). Prediction of breast cancer using machine learning approaches. Journal of biomedical physics & engineering, 12(3), 297.
5. Wei, Y., Zhang, D., Gao, M., Tian, Y., He, Y., Huang, B., & Zheng, C. (2023). Breast cancer prediction based on machine learning. J. Softw. Eng. Appl, 16(8), 348-360.
6. Agrawal, N. (2025). Predicting breast cancer risk: A machine learning approach. In Emerging Trends in Computer Science and Its Application (pp. 584-587). CRC Press.
7. Arravalli, T., Chadaga, K., Muralikrishna, H., Sampathila, N., Cenitta, D., Chadaga, R., & Swathi, K. S. (2025). Detection of breast cancer using machine learning and explainable artificial intelligence. Scientific Reports, 15(1), 26931.
8. Khan, M. Z., Salahuddin, M., Khan, A. A., Sultan, S. A., Bhatti, S. D., Ullah, S., & Mohamed, M. A. (2026). Innovative Hybrid Machine Learning Methods for Breast Cancer Predictive Optimization. International Journal of Computational Intelligence Systems.
9. Jothi, G, Krishnan, V., Rathinavel, T., Shalini, K. S., Selvankumar, T., Gomathi, M., ... & Ameer, K. (2025). Enhancing breast cancer detection accuracy through machine learning, deep learning and transfer learning techniques for clinical practice. Discover Artificial Intelligence.
10. Asuncion, A., & Newman, D. (2007, November). UCI machine learning repository.
11. Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5), 717-727.
12. LaValley, M. P. (2008). Logistic regression. Circulation, 117(18), 2395-2399.
13. Rigatti, S. J. (2017). Random forest. Journal of insurance medicine, 47(1), 31-39.
14. Jothi, G., Inbarani, H. H., & Azar, A. T. (2013). Hybrid tolerance rough set: PSO based supervised feature selection for digital mammogram images. International Journal of Fuzzy System Applications (IJFSA), 3(4), 15-30.
15. Jothi, G., Akilandeswari, J., David Samuel Azariya, S., & Naveenkumar, A. (2022, April). Unsupervised feature selection approaches for medical dataset using soft computing techniques. In Evolution in Computational Intelligence: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021) (pp. 105-114). Singapore: Springer Nature Singapore.