Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 2, Issue 5, 2024

BREAST CANCER DETECTION USING MACHINE LEARNING AND EXPLAINABLE AI TECHNIQUES

Shujaat Ali Rathore*[1], Muhammad Hammad u Salam[2], Dr. Mohd Yaqoob Wani[3], Mehmood Ashraf[4], Muhammad Irfan[5]

*[1]*Department of Computer Science & Information Technology, University of Kotli, Azad Jammu and Kashmir.*
Corresponding Author,
[2]*Department of Computer Science & Information Technology, University of Kotli, Azad Jammu and Kashmir.*
[3]*Dean, Faculty of Computer Sciences, Ibadat International University, Islamabad.*
[4]*Department of Communication and Cyber Security, Bahauddin Zakariya University, Multan, 60000, Pakistan,*
[5]*Department of Computer Science, NCBA&E, Sub-Campus Multan, 60000, Pakistan*

shujaat.ali@uokajk.edu.pk

DOI: https://doi.org/10.5281/zenodo.17173272

Keywords
Breast Cancer Detection, Machine Learning Algorithms, Explainable Artificial Intelligence (XAI), SHAP, LIME, ELI5, Ensemble Learning, Stacked Model, Feature Importance, Attribute Selection, Medical Diagnosis, Predictive Analytics, Clinical Decision Support

Article History
Received: 11 October 2024
Accepted: 21 November 2024
Published: 18 December 2024

Corresponding Author: *
Shujaat Ali Rathore

Abstract

Breast cancer arises from the uncontrolled growth of abnormal cells within breast tissue, often progressing into malignant tumors capable of spreading and becoming life-threatening. A complex interplay between environmental influences and individual genetic makeup contributes to the development of this condition. With the increasing prevalence of breast cancer, enhancing early detection and treatment methods is critically important. In recent years, machine learning (ML) has shown significant promise in improving diagnostic accuracy in the medical field. This study explores the use of patient diagnostic features alongside various ML classifiers to identify breast cancer cases effectively. Explainable Artificial Intelligence (XAI) methods were integrated to provide interpretability and transparency into the model predictions. Among the models evaluated, the Random Forest classifier achieved the highest F1-score of 84%, while a stacked ensemble approach—combining multiple model outputs—achieved an F1-score of 83%. The study further examined explanation techniques such as SHAP (SHapley exPlanations), LIME (Local Interpretable Model-agnostic Explanations), ELI5, Anchor, and QLattice to interpret model behavior and feature importance. These interpretable approaches have the potential to support clinicians in making more accurate diagnoses, minimizing errors, and improving decision-making processes in breast cancer care

INTRODUCTION

Breast cancer occurs when breast cells grow abnormally and form tumors capable of becoming malignant. After skin cancer, it is the most frequently diagnosed cancer in women. In 2022, around 2.3 million women worldwide received a breast cancer diagnosis, and roughly 670,000 died as a result. While breast cancer is much more common in women, men can also develop it. Breast cancers are classified according to where they originate in the breast—either invasive or non-invasive. Primary risk factors include aging, family history, inherited genetic mutations, hormonal shifts, lifestyle behaviors, and environmental exposures. Without treatment, the disease can spread to other organs, leading to multi-organ failure and potentially death. Research indicates that early detection significantly improves survival rates. Key symptoms that warrant urgent medical attention include breast or underarm swelling, nipple discharge (often with blood), breast pain, and changes in breast size or shape.

Screening tools like mammography reveal breast density, calcifications, abnormalities, and lumps. For more detailed diagnosis, tests such as MRI, PET scan, CT scan, gene panels, receptor status tests (e.g., estrogen or progesterone), and biopsy are used. Treatment options include surgery, chemotherapy, radiation therapy, hormone therapy, targeted therapy, and immunotherapy. Although genetic predispositions cannot be altered, lifestyle changes—limiting alcohol and tobacco use, maintaining healthy body weight, exercising, breast-feeding when possible, and avoiding hormone therapy post-menopause—can help reduce risk.

Despite the availability of diagnostic tools, limitations exist: imaging may not always detect abnormalities, symptoms may be overlooked, tests can be expensive, and diagnostic processes may be slow. According to WHO data, diagnostic errors are a leading source of adverse outcomes. Machine learning (ML) offers a way to mitigate human error by learning patterns from data and assisting clinicians. Many studies have applied ML models to breast cancer prediction, aimed at accelerating diagnosis, reducing errors, and improving treatment planning. Explainable AI (XAI) methods make these models' decisions more transparent and trustworthy, enabling validation of results and identification of potential mistakes.

Several researchers have explored ML-based breast cancer detection. For example, Sharma et al. used the WDBC dataset (569 samples, 32 features), finding k-Nearest Neighbors (kNN) yielded ~95.90% accuracy. Thilaka et al. focused on five major features using the same dataset; their best model was the Support Vector Classifier (SVC), achieving ~93% accuracy. Silva et al. studied a dataset from Indonesian women (400 cases, half diagnosed with cancer), where XGBoost reached ~85% accuracy; they also used SHAP and CRISP-ML(Q) for explainability and quality control. Munshi et al. combined Support Vector Machine (SVM) and Random Forest models in an ensemble, achieving near perfect accuracy (99.99%). Islam et al. developed a predictive model with data from 500 patients in Bangladesh (Dhaka Medical College Hospital); among five ML algorithms tested, XGBoost performed best at ~97%, with SHAP used for result transparency. Furthermore, Oztekin et al. compared an XAI model with radiologist assessments using a novel breast cancer dataset comprising 752 instances.

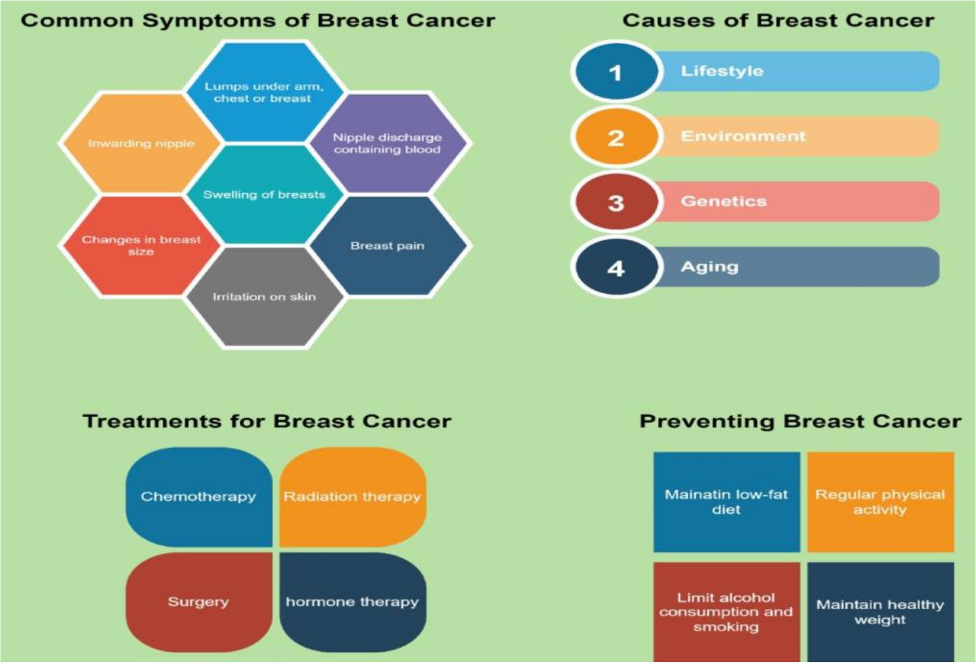


Fig. **Common symptoms, causes, treatments, and methods for breast**

	Author	Dataset	Best ML model	XAI techniques	Accuracy
1.	Sharma et al.[11]	Wisconsin Diagnosis Breast Cancer dataset	kNN	None	95.90%
2.	Thilaka et al.[12]	WBCD dataset	SVC	None	93%
3.	Silva et al.[13]	Mendeley data	XGBoost	SHAP	85%
4.	Munshi et al.[14]	Breast Cancer Wisconsin dataset	(RF + SVM) ensemble model	SHAP	99.99%
5.	Islam et al.[15]	Patient data from Dhaka Medical College Hospital	XGBoost	SHAP	97%
6.	Oztekin et al.[9]	Newly created Breast-XD dataset	XGBoost	SHAP LIME	94.43%
7.	Imouokhome et al.[16]	BreakHis dataset from Kaggle Challenge	ResNet50	Integrated Gradient, GradientShap And Occlusion	96.84%

Table 1. Articles that use ML for the diagnosis of breast cancer.

S.No	Feature name	Description
1.	Age	The patient's age at the time of diagnosis.
2.	Menopause	The menopausal status of the patient.
3.	Tumor size	The size of the excised tumor, measured in centimeters.
4.	Involved nodes	The number of axillary lymph nodes affected by tumor spread.
5.	Breast	Whether the tumor is located on the left or right side.
6.	Metastasis	Whether the cancer has metastasized to other parts of the body or organs.
7.	Breast	The breast is divided into four sections, with the nipple at the center,

	quadrant	indicating the tumor's precise location.
8.	History	Whether the patient has a personal or family history of cancer.
9.	Diagnosis result	Classifying the tumor into benign and malignant.

Table 2. Explanation of the dataset's features.

They labeled patients' lesions as either benign or malignant and tested six different classifiers; **XGBoost** achieved the highest accuracy at **94%**. In comparison, a radiologist's diagnosis reached **83.95%**, showing that ML models can be very reliable. They used **SHAP** and **LIME** to highlight which features were most influential. In another study, *Imouokhome et al.* classified breast tumors via histopathological images with a **pre-trained ResNet50**; using a dataset of 7,909 images (from 82 patients via open surgical biopsies), they achieved **96.84%** accuracy. Deep learning outputs in that study were interpreted using advanced XAI methods. A detailed survey of these and related studies is provided in Table 1.

We observed that most prior studies do *not* apply multiple diverse XAI tools to interpret their results. Also, important predictors identified in those studies are rarely validated by other methods. To address these gaps, this research makes the following contributions:

- We measure the dependency of features on diagnosis via **mutual information**, so we can understand how strongly each marker contributes to breast cancer detection.
- We develop a custom **stacked ensemble classifier**, combining multiple models to deliver more robust predictions.
- We employ **five XAI techniques**—LIME, Anchor, SHAP, QLattice, and ELI5—to determine feature importance, and further validate the key markers using mutual information together with inferential statistics.

Major limitations noted include: a relatively small dataset and lack of consistency in which features are selected across different studies

Materials and methods Dataset description

In this study, the "UCTH Breast Cancer Dataset" was utilized for conducting machine learning analysis. This dataset was made available on the Mendeley Data repository in 2023 and originates from the University of Calabar Teaching Hospital (UCTH) in Nigeria. It was compiled through clinical observations of 213 patients over a period of two years. The dataset includes nine key features relevant to breast cancer diagnosis: age, menopausal status, tumor size, number of involved lymph nodes, the side of the breast affected, presence of metastasis, the specific breast

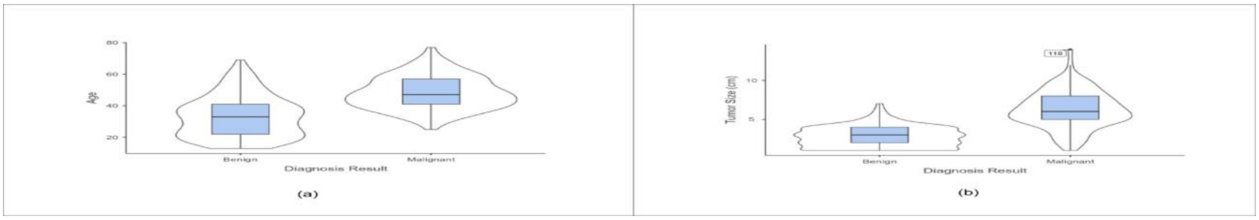


Fig. 2 Violin plots**a)**(Age(**b)** Tumor size.

quadrant involved, and previous history of cancer. Among these, age and tumor size are continuous variables, while the remaining features are categorical. The target variable, labeled as 'diagnosis result,' is also categorical, where '0' denotes abenign tumor and '1' indicates a malignant one. These attributes collectively contribute to a comprehensive dataset used to train and evaluate machine learning models for breast cancer prediction. A detailed overview of these features is presented in Table 2

	Diagnosis result	*N*	Missing	Mean	Median	SD	Range	Minimum	Maximum
Tumor size (cm)	Benign	120	0	2.72	3	1.34	6	1	7
	Malignant	92	1	6.29	6	2.38	13	1	14
Age	Benign	120	0	33.21	33	13.04	56	13	69
	Malignant	93	0	48.27	47	10.44	52	25	77

Table 3 Descriptive analysis of continuous variables.

		Statistic	df	*p*
Age	Student's t	-9.11	211	< 0.001
Tumor Size (cm)	Student's t	-13.85	210	< 0.001

Table4. Independent samples t-test. Inference: Age and tumor size are significant features

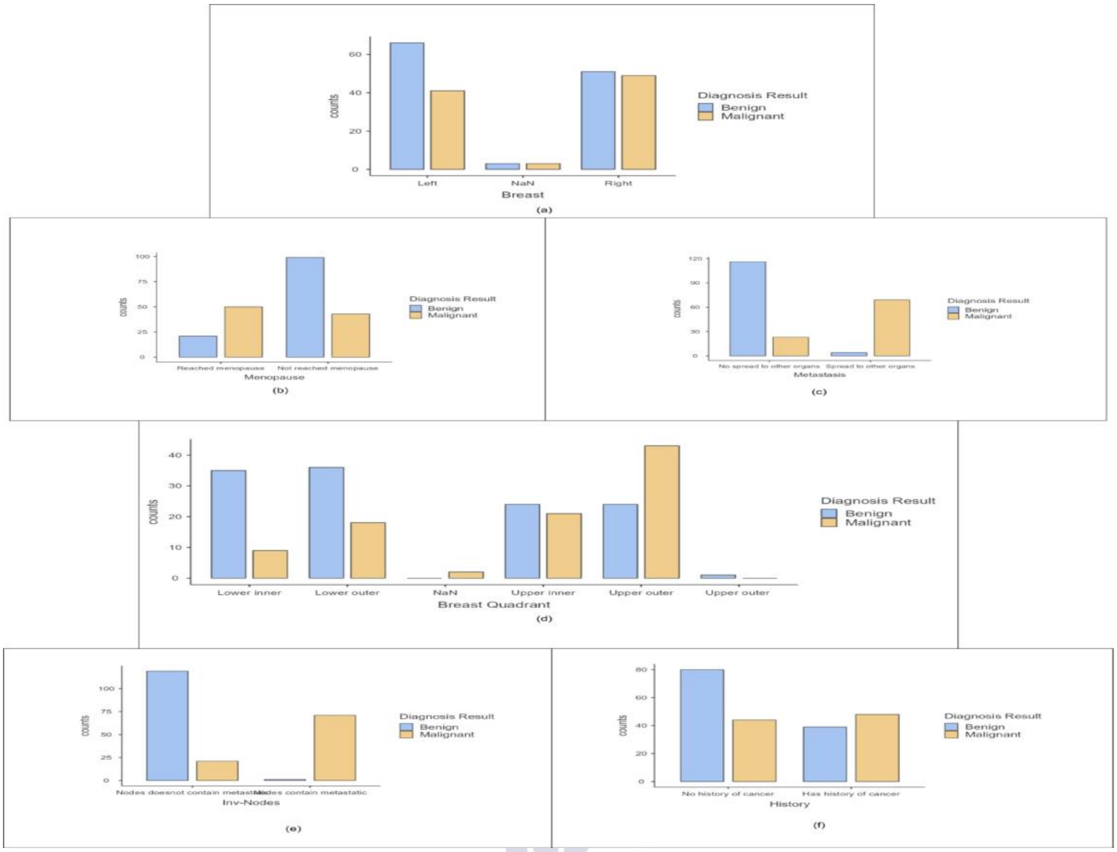
Statistical preprocessing

Statistical preprocessing was conducted using Jamovi to derive both statistical and descriptive insights. Descriptive analysis of the continuous variables is summarized in Table 3, while violin plots in Figure 2 illustrate the distribution of numerical data. The plots indicate that older women have a higher likelihood of developing malignant breast tumors compared to younger women, and larger tumor sizes are associated with malignancy. T-tests were applied to assess the significance of continuous features, with a threshold p-value of less than 0.001 considered significant. Based on the results shown in Table 4, both age and tumor size were identified as important features.

Categorical variables were examined through bar plots depicted in Figure 3, showing the distribution of benign and malignant cases across different feature categories. The visualizations suggest that breast cancer is less common in women who have not yet reached menopause. Cancer diagnoses are more frequent when tumors have spread to auxiliary lymph nodes, and metastasis is notably present in breast cancer cases. Malignant tumors tend to occur more often in the upper outer quadrant of the breast. Additionally, patients with a prior history of cancer show an increased risk of developing breast carcinoma. These bar plots provide a detailed understanding of the dataset's characteristics. A chi-square test was performed to determine the significance of categorical variables, with results displayed in Table 5. The features menopause, involved nodes, breast quadrant, and metastasis were found to be significant predictors of breast cancer according to the chi-square analysis.

Data preprocessing

Data preprocessing is essential for transforming raw data into a format that is easier to analyze. In this study, preprocessing was applied to handle missing values and outliers, as well as to optimize the number of input features for analysis. Initially, the data was shuffled to prevent the model from learning any sequence patterns. The dataset contained 13 missing values labeled as 'NaN,' which were removed to ensure consistency. Label encoding was employed to convert categorical text data into numerical values, as machine learning algorithms require numerical inputs. Each category was assigned a unique integer for processing. To prevent bias caused by varying value ranges, data scaling was performed using Max-Abs scaling, which scales all values to fall between -1



and 1 based on their maximum absolute value. Feature importance was assessed using mutual information and Pearson's correlation. Pearson's correlation coefficients between feature pairs were visualized using a heatmap, facilitating the identification of relationships among variablesof relationships among variables

representing the categorical variables: (a) Breast, (b) Menopause, (c) Metastasis, (d) Breast quadrant, (e) Involved nodes, and (f) History. These visualizations illustrate the distribution of patients across benign and malignant tumor diagnoses for each feature, aiding in understanding their impact on breast cancer diagnosis.

S.No	Attribute	*p* value
1	Menopause	< 0.001
2	Involved nodes	< 0.001
3	Breast	0.287
4	Breast quadrant	< 0.001
5	Metastasis	< 0.001
6	History	0.005

Table 5 presents the results of the chi-square test, indicating that Menopause, Involved Nodes, Breast Quadrant, and Metastasis are significant features influencing breast cancer diagnosis.

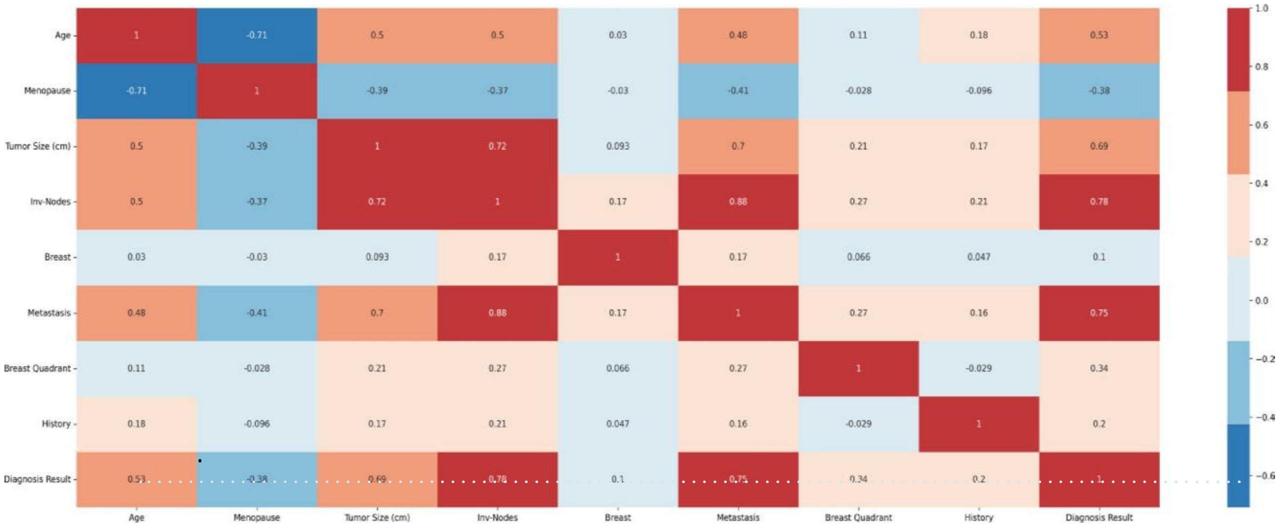
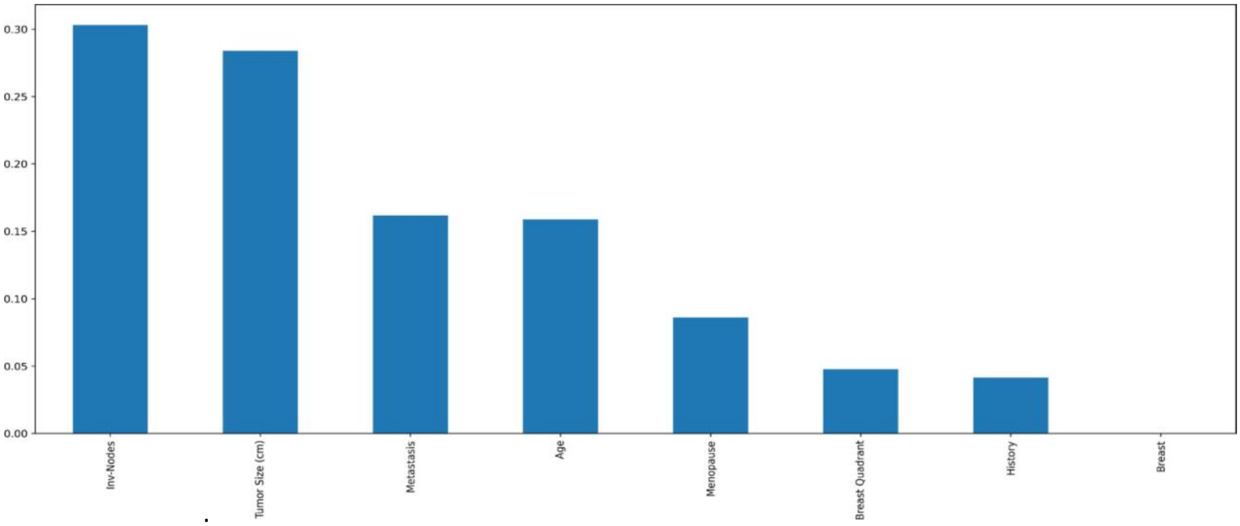


Fig. 4 Pearsons correlation heatmap.



Mutual information of features.

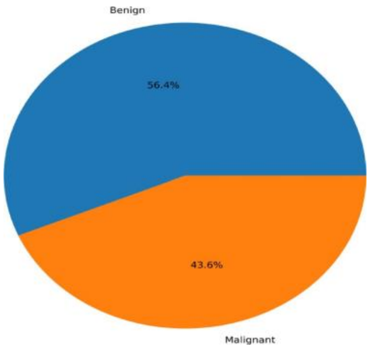


Fig. Diagnosis

Values of 1, 0, and −1 represent positive, no, and negative correlations, respectively. The heatmap shown in Figure 4 reveals that Involved Nodes, Metastasis, Tumor Size, and Age have a strong correlation with the diagnosis result. Mutual Information, a univariate filtering method, evaluates the importance of each feature independently by measuring the dependency between two variables using entropy. As illustrated in Figure 5, the features ranked by mutual information importance include Involved Nodes, Tumor Size, Metastasis, Age, Menopause, Breast Quadrant, and History. Figure 6 presents a pie chart depicting the distribution of the target variable (diagnosis result), showing a slight imbalance between the classes.

There is an imbalance in the dataset, which could lead to biased model performance. To address this, Borderline-SMOTE was applied to the training data to generate synthetic samples, balancing the classes equally at 50% each. Subsequently, the dataset was split into training and testing sets with a ratio of 70:30.

Machine Learning and Explainable AI in Breast Cancer Diagnosis

he study utilized a **stacking algorithm** to combine predictions from eight different machine learning classifiers: XGBoost, LightGBM, CatBoost, AdaBoost, KNN, Decision Tree, Logistic Regression, and Random Forest. This ensemble approach combines the strengths of individual models, improving overall performance and reducing overfitting and biases. Hyperparameter tuning was performed using **GridSearchCV with 5-fold cross-validation** to optimize the models' performance on unseen data.

To ensure the model's transparency and reliability, five **Explainable Artificial Intelligence (XAI)** techniques were employed:

- **SHAP (Shapley Additive exPlanations):** This method assigns a **Shapley value** to each feature, quantifying its impact on the model's predictions. It's a model-agnostic technique that provides detailed, individualized explanations, helping to identify unexpected feature interactions and refine the model.
- **LIME (Local Interpretable Model-agnostic Explanations):** LIME explains the predictions for individual data samples by slightly altering the data and observing the changes in the output. This helps identify the most relevant features for a specific prediction, making the results understandable to patients and non-technical stakeholders.
- **Eli5 (Explain Like I'm Five):** Eli5 provides both global and local explanations of classifier weights and predictions. It helps in understanding the model's inner workings, identifying biases, and troubleshooting issues like overfitting by showing how different features contribute to a prediction. It's particularly useful for text data.
- **QLattice:** This technique explores a wide range of simple, interpretable models to find patterns and connections in data. It generates a set of mathematical expressions that explain the relationships between input and output, providing clear and understandable explanations unlike many complex models.

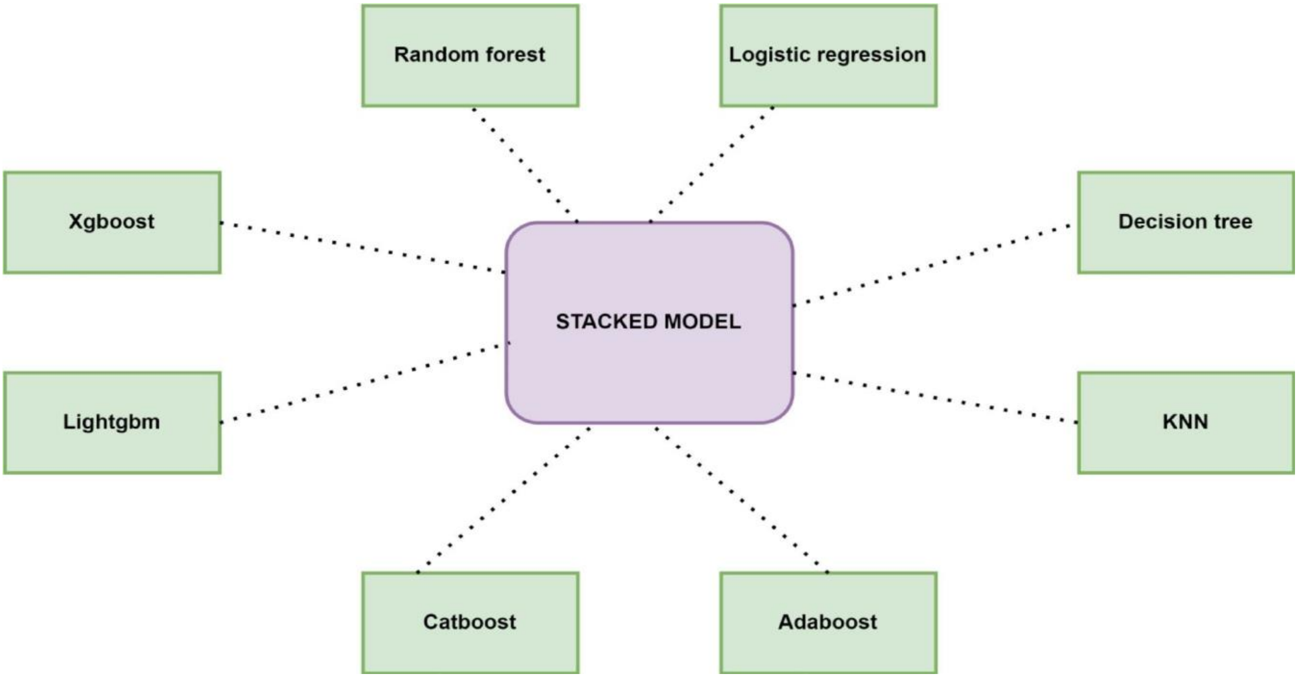


Fig. 7 Stacking architecture.

QLattice is a technique that focuses on discovering simple, understandable formulas to show users how input data is converted into output. The models it creates are inherently transparent, identifying key features and their interactions to influence the final result. It is also adaptable and robust to new data, as it can explore new models when data patterns change.

Anchor

Anchor, another model-agnostic method, uses "if-then" rules, or anchors, to explain specific decisions. In a medical context, an anchor might state that if certain symptoms and test results are present, a particular diagnosis will always be given. This method is locally faithful, accurately justifying a specific prediction even if the overall model is complex. By clearly outlining the reasoning behind each prediction, Anchor helps users build confidence in the model's decisions.

The study used all five XAI techniques (SHAP, LIME, Eli5, QLattice, and Anchor) together to cross-verify explanations and leverage each method's unique strengths, enhancing the reliability and versatility of the interpretive outputs for both doctors and patients.

Findings and Model Evaluation

This study applied multiple classifiers to predict breast cancer in patients, with optimization through ensemble stacking. To thoroughly assess model performance, a range of evaluation metrics were used, including Jaccard score, F1-score, log loss, Matthews correlation coefficient, accuracy, average precision, AUC, precision, recall, and Hamming loss. The performance of each classifier according to these metrics is summarized in Table 7. Among all classifiers, Random Forest, Logistic Regression, and CatBoost achieved the highest accuracy of 89%. Logistic Regression had the highest precision at 87%, indicating it produced fewer false positives compared to other models. Random Forest, CatBoost, and LightGBM each achieved a recall of 84%, demonstrating their effectiveness in correctly identifying true positive cases. Random Forest achieved the best balance between precision and recall, with a maximum F1-score of 84%. All models performed well in terms of AUC, with Random Forest attaining the highest value of 0.96, showing excellent capability in distinguishing between classes. Logistic Regression and CatBoost both reached an average precision of 0.94, reflecting a favorable balance between precision and recall. While the stacking classifier delivered consistent and reliable results, it did not surpass the top-performing models such as Random Forest, Logistic Regression, or CatBoost. The ensemble stack's performance, illustrated in Figure 9, shows AUC, average precision, and a confusion matrix highlighting a low number of false predictions, which contributes to improved classification accuracy. Hyperparameters optimized via grid search for each algorithm are detailed in Table 8.

Explainable AI analysisTo ensure transparency and clarity for all stakeholders, the machine learning predictions were interpreted using explainable AI (XAI) methods. This study applied five different XAI techniques: LIME, Anchor, SHAP, QLattice, and Eli5. The interpretations were conducted on the stacked ensemble model, as it combines the outputs of multiple classifiers. The SHAP (Shapley Additive Explanations) visualizations, shown in Figure 10, help reveal the influence of various features on the model's predictions, enabling a better understanding of how each characteristic contributes to the final decision.

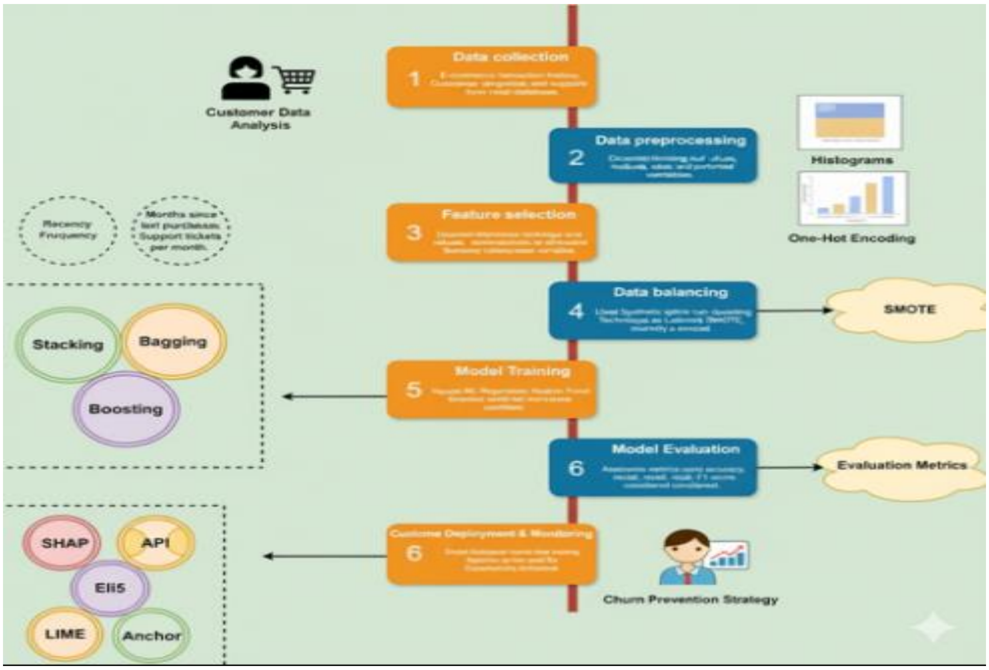


Fig 8. Flow diagram-based methodology.

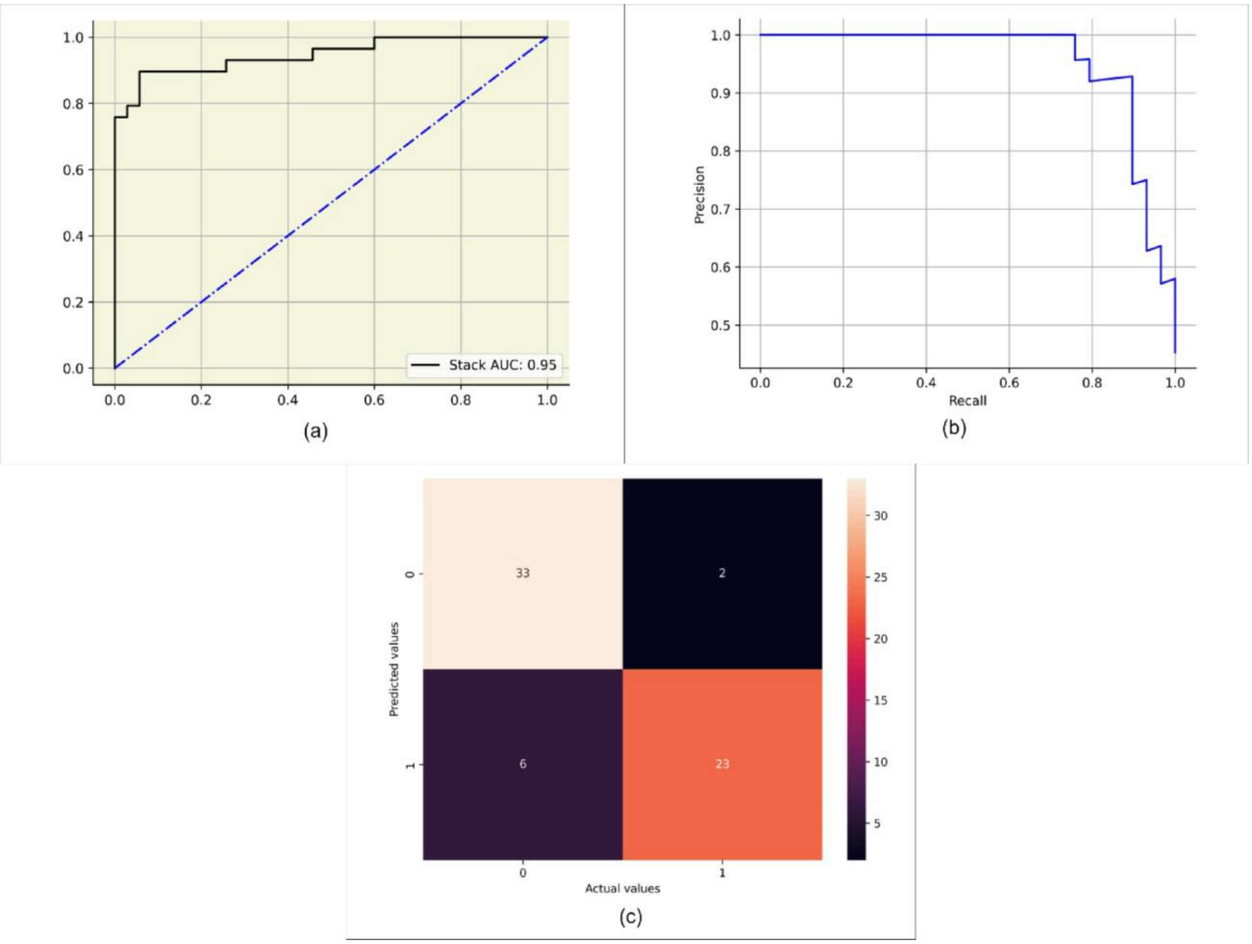
XAI Technique	Characteristics
SHAP	SHAP is preferred as it is not computationally intensive on small datasets. It is model agnostic. It provides both local and global interpretability[25]
LIME	LIME is preferred as it is not computationally intensive on small datasets. It is model agnostic. It provides local interpretability. Easy to interpret for technical and non-technical stakeholders. Allows quick adaptability to new data. It is an excellent choice for text data, as in the study
Eli5	It is typically fast compared to other models. Easy to interpret for technical and non-technical stakeholders. It provides both local and global interpretability. It is an excellent choice for text data, as in the study
QLattice	It can quickly adapt to new data and discover new model relationships. It provides both local and global interpretability. It doesn't identify bias, unlike other techniques
Anchor	Easy to interpret for technical and non-technical stakeholders. It provides coverage values to trust the specific rule. It provides local interpretability. No feature selection is involved

Table 6. Characteristics of five XAI techniques used.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Area under curve	Average precision	Hamming loss	Jaccard score	Log loss	Mathew's correlation coefficient
Random Forest	**89**	84	**93**	**84**	**0.96**	0.93	0.11	**0.79**	**3.94**	**0.78**
Logistic Regression	**89**	**87**	90	83	0.94	**0.94**	0.11	**0.79**	**3.94**	**0.78**
Decision Tree	80	71	**93**	79	0.92	0.9	0.2	0.68	7.32	0.63
KNN	78	76	76	83	0.78	0.8	0.22	0.61	7.88	0.56
Adaboost	86	81	90	81	0.93	0.9	0.14	0.74	5.06	0.72
Catboost	**89**	84	**93**	80	0.95	0.91	**0.1**	**0.79**	**3.94**	**0.78**
Lightgbm	83	76	90	80	0.93	0.91	0.17	0.7	6.2	0.67
Xgboost	86	79	**93**	81	0.94	0.92	0.14	0.75	5.07	0.73
STACK	84	81	86	83	0.93	0.92	0.16	0.71	5.63	0.69

Table 7. Performance analysis of the machine learning classifiers using different metrics.

The prediction output of the machine learning model for breast cancer diagnosis is explained using SHAP values. The beeswarm plot illustrates how SHAP values are distributed and their influence for each feature across all data points. The mean bar plot summarizes the average effect of each feature over the entire dataset. These Shapley values, grounded in game theory, ensure a fair allocation of the "payout" (here, the model's prediction) among the "players" (features). A hyperplane at SHAP value zero divides the two classes: values to the right indicate malignant cases (class 1), while values to the left correspond to benign cases (class 0). The Y-axis lists the features, with their values color-coded from blue (low) to red (high). Each dot on the plot represents an individual data instance.



(a) (b) (c)

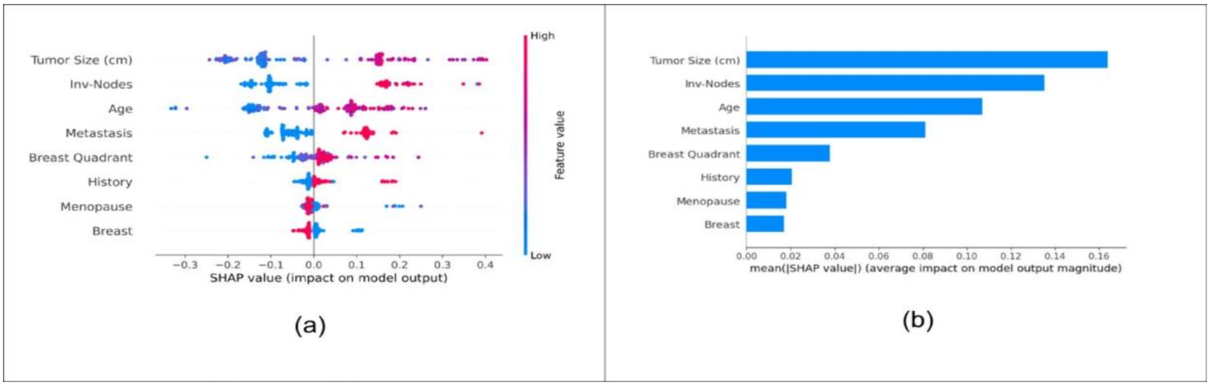
Fig. 9 Performance of stacking classifier **(a)** AUC curve **(b)** precision-recall curve **(c)** Confusion matrix.

The SHAP value for an individual prediction represents the contribution of a specific feature, with the color indicating the feature's actual value for that instance. For example, in the menopause feature, many data points with low values appear on the side with higher SHAP values, suggesting that patients who have not yet reached menopause are more likely to have benign tumors. The horizontal spread of dots reveals the variation in how much each feature influences the model's predictions. The mean bar plot condenses this information by showing the average absolute SHAP value for each feature, where longer bars signify greater influence on the prediction. Features such as breast quadrant, history, menopause, and breast have a smaller effect on the model's outcomes. Overall, malignant tumors are associated with larger tumor size, more involved nodes, older age, and presence of metastasis. SHAP provides both individual patient-level explanations and a global understanding of the model, helping to prioritize clinical attention and potentially guiding biopsy or imaging choices.

Algorithm	Hyperparameter values chosen
Random forest	{'bootstrap': Enabled,
	'maximum_depth': 80,
	'maximum_features': 2,
	'minimum_samples_leaf': 3,
	'minimum_samples_split': 8,
	'n_estimators': 200}
Logistic regression	{'C': 1000, 'penalty': 'l2'}

Decision tree	{'criterion': 'gini',
	'maximum_depth': 90,
	'maximum_features': 'log2',
	'minimum_samples_leaf': 5,
	'minimum_samples_split': 10,
	'splitter': 'best'}
KNN	{'no_of_neighbors': 1}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 1000}
Catboost	{'border_count': 5,
	'depth': 3,
	'iterations': 250,
	'l2_leaf_reg': 1,
	'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 0,
	'lambda_l2': 1,
	'minimum_data_in_leaf': 30,
	'num_leaves': 31,
	'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3,
	'gamma': 0.0,
	'learning_rate': 0.15,
	'maximum_depth': 6,
	'minimum_child_weight': 1}
STACK	{ use_probas = Enabled,
	average_probas = Disabled,
	meta_classifier = Logistic regression}

Table8.Fine-Tuning Model Parameters via Grid Searc



(a)

(b)

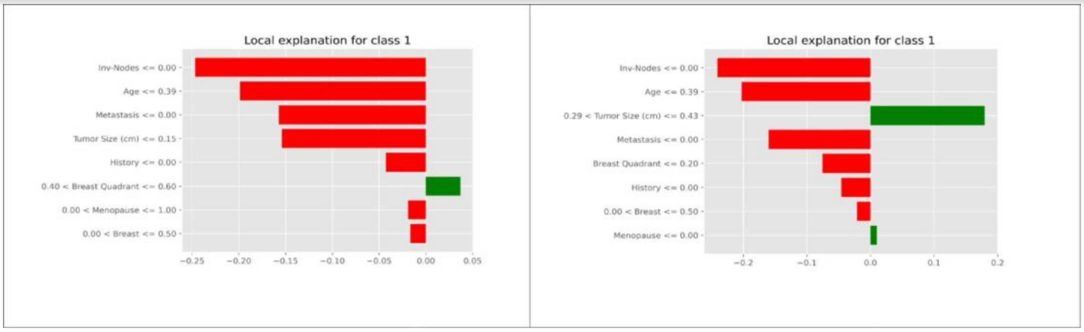
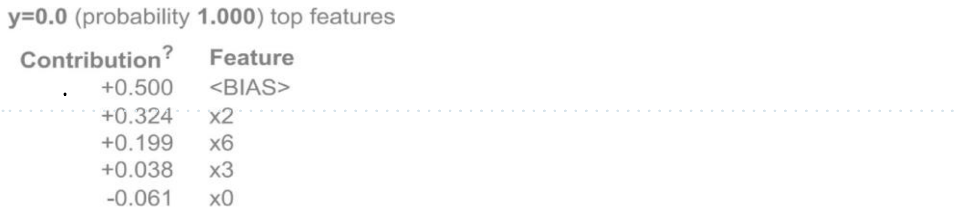


Fig. 11LIME analysis for different



y=0.0 (probability 1.000) top features

Contribution?	Feature
+0.500	<BIAS>
+0.324	x2
+0.199	x6
+0.038	x3
-0.061	x0

Fig. 12 Interpreting the Eli5 Model for Breast Cancer Prediction

LIME plots offer a detailed perspective on how each feature influences the prediction for an individual case by approximating the complex model locally with a simpler, interpretable one. This surrogate model generates explanations by creating new samples around the specific instance, slightly altering its features. The importance of each feature in this simple model explains the prediction, with larger absolute weights indicating greater influence on that instance's classification. The local explanation for a malignant (class 1) case is shown in Fig. 11, where the X-axis reflects prediction probability and each bar represents a feature. Positive bars indicate features that increase the likelihood of a malignant prediction, while negative bars decrease it. For example, a tumor size value of less than or equal to zero appears as a negative bar, meaning it reduces the probability of malignancy. Features are ordered by their contribution, and LIME highlights involved nodes, age, metastasis, and tumor size as key predictors. This method allows clinicians to see how individual features affect a patient's diagnosis, helping them understand how small changes might impact outcomes and guiding follow-up or monitoring decisions.

ELI5 complements this by offering insight into the model's overall reasoning through aggregated feature contributions across many predictions. In Fig. 12, each row shows a feature's contribution—such as age (x0), tumor size (x2), involved nodes (x3), and breast quadrant (x6)—along with a bias term. Positive values indicate strong support for the prediction, while negative values suggest less influence. For instance, tumor size contributes +0.324, signifying a strong impact. ELI5 thus identifies these features as significant predictors. By examining global feature importance, clinicians can evaluate which factors generally influence the model, detect biases, and verify the model's logic against clinical knowledge, increasing trust and facilitating integration into clinical practice.

Figure 13 presents an analysis using QLattice, which explores potential model structures and feature interactions. The QGraph depicts a mathematical model connecting input features to the output through nodes and edges, where nodes perform operations (addition, multiplication, etc.) and edges represent data flow. The graph is directed, moving data from inputs toward the output node. Tumor size, involved nodes, and age emerge as important features. Activation strengths indicate how strongly each node affects the outcome, with moderate activation suggesting balanced contributions. The final combined value passes through a logistic function, converting it into a probability between 0 and 1. Overall, QGraph provides a mathematical representation best fitting the predictions, offering interpretability distinct from other techniques.

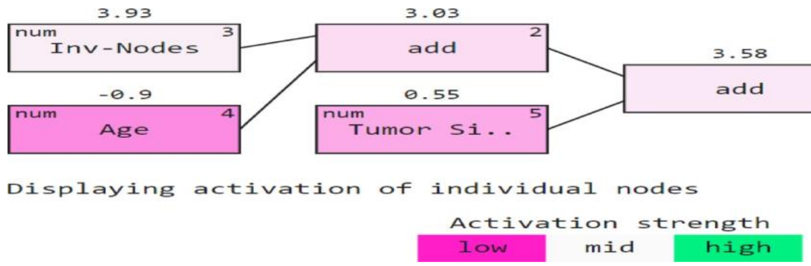


Fig. 13 .QGraph analysis.

Anchors	Prediction	Precision	Coverage
Age < = 0.40 AND Tumor Size (cm) < = 0.29	Benign	0.97	0.23
Tumor Size (cm) < = 0.14 AND Inv-Nodes < = 0.00	Benign	0.9	0.27
Age < = 0.40 AND Tumor Size (cm) < = 0.14	Benign	1	0.12
Tumor Size (cm) < = 0.14 AND Inv-Nodes < = 0.00	Benign	0.89	0.28
Tumor Size (cm) < = 0.29 AND Inv-Nodes < = 0.00	Benign	0.87	0.55
Inv-Nodes > 0.00 AND Tumor Size (cm) > 0.29	Malignant	0.98	0.32
Inv-Nodes > 0.00 AND Age > 0.64	Malignant	0.99	0.15
Tumor Size (cm) > 0.29 AND Inv-Nodes > 0.00	Malignant	0.99	0.31
Age < = 0.53 AND Tumor Size (cm) < = 0.14	Malignant	0.99	0.21
Inv-Nodes > 0.00 AND Age > 0.53	Malignant	1	0.26

Table 9. Anchors with precision and coverage.

Unlike other XAI techniques, QLattice does not offer straightforward interpretability for non-technical users. However, its adaptability to new data makes it a valuable tool for generating meaningful interpretations. This method supports clinical workflows by enabling physicians to quickly identify key indicators, facilitating the integration of AI insights into standard diagnostic procedures.

Anchors represent specific conditions—combinations of feature values—that reliably predict an outcome when met. Table 9 presents anchor explanations, with each row showing an anchor rule, its precision, and coverage. Precision reflects the probability that the prediction is correct under the given condition, while coverage indicates how many cases in the dataset satisfy that condition. For example, the rule "Inv-Nodes > 0.00 AND Age > 0.64" predicts malignancy with 99% accuracy based on the dataset. This approach enhances transparency and interpretability

by highlighting feature combinations that strongly influence predictions. Such clarity helps ensure consistent and trustworthy clinical decisions by providing doctors with a clear checklist of influential factors. Collectively, these XAI techniques work together to foster transparency and confidence in the model among healthcare professionals and patient.

Discussion

This research utilized the "UTCH Breast Cancer Dataset" obtained from Mendeley Data, originally collected at the University of Calabar Teaching Hospital, Nigeria. The dataset comprised 213 patient records with nine features. The study evaluated the performance of nine classifiers for breast cancer prediction and applied five explainable AI (XAI) methods to interpret the results. The evaluation primarily focused on the F1-score, a critical metric in medical contexts as it balances precision and recall. Higher precision reduces false positives, thereby minimizing unnecessary medical interventions and patient anxiety, while higher recall lowers false negatives, ensuring fewer missed cancer cases. The random forest classifier achieved the best result with an F1-score of 84%. The stacking ensemble model combined the strengths of individual classifiers and balanced their biases, reaching an F1-score of 83%, with only two false positives and six false negatives out of 65 test samples. Previous research comparing radiologist assessments reported an F1-score of 84.88% on a larger dataset of 752 patients. The comparable performance of the stacking model suggests its reliability relative to advanced diagnostic methods, potentially assisting radiologists in faster, more accurate decision-making with reduced human error. Across all five XAI techniques, age, involved nodes, and tumor size emerged as the most important features, with LIME also highlighting breast quadrant as significant. Older age showed a negative correlation with cancer prediction, while affected nodes and larger tumors increased risk. These XAI methods provide valuable insights throughout clinical workflows: SHAP and LIME help identify key predictors during initial screening, ELI5 and QLattice reveal diagnostic pathways for detailed analysis, and Anchors ensure critical conditions are met before confirming a diagnosis. These markers aid precise detection and align with findings from prior medical studies.

The study compared feature importance using statistical hypothesis tests, mutual information, and XAI methods, summarized in Table 10. Tumor size, age, and involved nodes were consistently significant across all analyses. Mutual information and statistical tests also flagged metastasis as important, while chi-square tests identified menopause and breast quadrant as relevant markers for breast cancer diagnosis. These features help machine learning models better predict breast cancer status. Cross-validation of features through different analytical approaches strengthens confidence in these markers. While statistical tests measure associations or mean differences—often limited to linear relationships and dataset biases—mutual information quantifies shared information between variables. Both methods assess individual feature significance but may miss interactions between features affecting predictions. XAI techniques complement these by capturing complex feature interactions influencing the model.

Several studies have applied machine learning to breast cancer diagnosis. Sharma et al. and Thilaka et al. used the same dataset but varied feature selection and classifiers, showing machine learning's adaptability in tailoring models to specific clinical goals and improving practical utility. Adding XAI increases transparency, enabling clinicians to understand and trust model outputs, reducing misdiagnosis risk and reassuring patients about treatment decisions. Many researchers integrate XAI with ML to enhance medical diagnosis effectiveness. This study made efforts to develop a superior model with XAI support. Few prior works combined ensemble models with multiple XAI methods. Aravena et al. focused on demographic-specific breast cancer data from Indonesian women, identifying features like age at first pregnancy and diet as impactful. Another study applied ensemble learning with SHAP explanations on the Wisconsin dataset containing 32 image-derived features. Islam et al. used geometric image

Features	Chi-square/t-tests	Mutual Information	XAI
Tumor size	✓	✓	✓
Age	✓	✓	✓
Menopause	✓	×	×
Involved nodes	✓	✓	✓
Breast Quadrant	✓	×	×
Metastasis	✓	✓	×
Breast	×	×	×
History	×	×	×

Table 10. Key features identified through statistical analysis, mutual information, and XAI methods

the F1-score to address the critical balance between false positives and false negatives in sensitive breast cancer diagnosis.

features and SHAP for interpretability, while Imouokhome et al. targeted cellular-level image classification using ResNet50 and explainability techniques like Integrated Gradient and GradientShap. Oztekin et al. evaluated models with nine image-based features and compared results to radiologist assessments, using SHAP and LIME for explanation. Table 11 compares these studies with the current work,

which uniquely employs multiple XAI techniques alongside statistical and mutual information analyses, offering a comprehensive understanding of feature importance to improve decision-making, model transparency, and performance. Throughout, emphasis was placed on the F1-score to address the critical balance between false positives and false negatives in sensitive breast cancer diagnosis

References	Dataset size	No. of features	ML models used	Performance	XAI Techniques
11	569 instances	32 attributes	Three classifiers	95.90% accuracy, 98.27% precision, 94.20% F1 score	None
12	500 samples	5 features	Two classifiers	93% accuracy	None
13	400 records of Indonesian Women	Not mentioned	Four classifiers	85% accuracy, 85.4% precision	SHAP
14	569 instances	32 features	Several classifiers with ensemble models	99.99% F1 score, 99.99% recall, 99.99% precision, 99.99% accuracy	SHAP
15	500 records of Bangladeshi patients	7 features	Five classifiers	96% F1 score, 95% recall, 94% precision, 97% accuracy	SHAP
16	7907 histopathological images	No numerical features	ResNet50 CNN model	96.84% accuracy	Integrated Gradient, GradientShap,

					Occlusion
9	752 patients	11 features	Six classifiers along with radiologist reviews	94.34% accuracy, 92.11% precision, 93.96% F1 score	SHAP and LIME
This research	213 patients	9 features	Nine classifiers + Custom Stack ensemble model	84% F1 score	**SHAP, LIME, Eli5, Qlattice and Anchor**

Table 11. Summary of comparable studies..

The study also utilized GridSearchCV for optimizing hyperparameters. Unlike many other works, this research incorporated multiple XAI techniques, each bringing unique advantages, which together offer a robust and reliable interpretation of the model. This combination effectively bridges the gap between global and local model explanations. Some of these XAI methods are model-agnostic, while others are specific to certain models, providing flexibility in drawing accurate conclusions. Additionally, the chosen XAI approaches are relatively simple and easy to apply, making them accessible to a broad audience. The integration of a stacked ensemble model alongside five XAI techniques underscores the significant contribution of this work to breast cancer detection. However, there are some limitations. The model's accuracy could be enhanced by employing deep learning methods [32,33], and increasing the dataset size would help mitigate overfitting and reduce variance. The dataset used contained only numerical features, which restricts feature diversity [34]. Moreover, the study focused on a specific demographic group (Nigeria), which may cause domain mismatch issues—meaning the model's effectiveness could decline when applied to more diverse populations. Future improvements include expanding the dataset with samples from varied demographics and validating the model on external datasets to ensure broader applicability [35]. Comparing machine learning model outputs with clinical expert assessments can help minimize misdiagnoses. Further research could increase both dataset size and diversity to improve generalizability. Real-time diagnostic validation through biopsy or radiological imaging reviews could strengthen clinical relevance. Incorporating raw image data may offer greater flexibility in feature selection. Lastly, standardizing feature sets across studies would promote collaborative research and accelerate advancements.

Conclusion

Breast cancer is characterized by the abnormal proliferation of cancer cells within breast tissue. If left untreated, the disease can spread to nearby tissues, posing serious risks including death. This study applied machine learning methods alongside explainable AI techniques to predict breast cancer presence. Several machine learning algorithms were employed to classify tumors as benign or malignant. Among these, the stacked ensemble model demonstrated superior generalization and robustness compared to individual models, and showed resilience to bias. To address the black-box nature of these models, multiple explainability methods were implemented, focusing on innovative yet easily interpretable techniques suitable for healthcare professionals. The explanations generated by XAI methods were further validated using mutual information and statistical hypothesis testing. These classification models have the potential for integration into clinical settings such as hospitals, diagnostic laboratories, and research studies. Early detection facilitated by these models can contribute to lowering breast cancer mortality rates .

REFERENCES

Ravi, S. & Saranya, A. Breast Cancer detection using machine learning in medical Imaging–A survey. *Procedia Comput. Sci.* **239**, 2235–2242 (2024).

Kapila, R. & Saleti, S. An efficient ensemble-based machine learning for breast cancer detection. *Biomed. Signal Process. Control.* **86**, 105269 (2023).

Das, A. K., Biswas, S. K., Mandal, A., Bhattacharya, A. & Sanyal, S. Machine learning based intelligent system for breast cancer prediction (MLISBCP). *Expert Syst. Appl.* **242**, 122673 (2024).

Talukdar, P. & Ray, R. Analysis of breast cancer classification using machine learning techniques and hyper parameter tuning. *Biocatal. Agric. Biotechnol.* **25**, 103195 (2024 Apr).

Yadav, R. K., Singh, P. & Kashtriya, P. Diagnosis of breast cancer using machine learning techniques-a survey. *Procedia Comput. Sci.* **218**, 1434–1443 (2023).

Darwich, M. & Bayoumi, M. An evaluation of the effectiveness of machine learning prediction models in assessing breast cancer risk. *Inf. Med. Unlocked*. **14**, 101550 (2024 Jul).

Ramay, S. A., Kanwal, K., Javid, H. A., Abbas, T., Ansari, G. J., & Irfan, M. (2023). Enhancing Fruit Quality Detection with Deep Learning Models. *Journal of Computing & Biomedical Informatics*, 6(01), 28-40.

Singh, A., Kaur, S., Singh, D. & Singh, G. Technical Review of Breast Cancer Screening and Detection using Artificial Intelligence and Radiomics. In2024 11th International Conference on Computing for Sustainable Global Development (INDIACom) 2024 Feb 28 (pp. 1171–1176). IEEE.

Adedayo-Ajayi, V. O., Ogundokun, R. O., Tunbosun, A. E., Adebiyi, M. O. & Adebiyi, A. A. Metastatic breast cancer detection using deep learning algorithms: A systematic review. In *International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)* 2023 Apr 5 (Vol. 1, pp. 1–5). IEEE. (2023).

Oztekin, P. S. et al. Comparison of explainable artificial intelligence model and radiologist review performances to detect breast Cancer in 752 patients. *J. Ultrasound Med.* (2024).

Bai, S. et al. Breast Cancer diagnosis: A comprehensive exploration of explainable artificial intelligence (XAI) techniques. ArXiv Preprint ArXiv:2406.00532. Jun 1. (2024).

Janjua, J. I., Nadeem, M., Khan, Z. A., & Khan, T. A. (2022, April). Computational Intelligence Driven Prognostics for Remaining Service Life of Power Equipment. In *2022 IEEE Technology and Engineering Management Conference (TEMSCON EUROPE)* (pp. 1-6). IEEE.

Alqarafi, A., Batool, H., Abbas, T., Janjua, J. I., Ramay, S. A., & Ahmed, M. (2024, December). Estimating Uncertainty in Deep Learning Methods and Applications. In *2024 International Conference on Computer and Applications (ICCA)* (pp. 1-6). IEEE.

Abbas, T., Fatima, A., Shahzad, T., Alharbi, M., Khan, M. A., & Ahmed, A. (2024). Multidisciplinary cancer disease classification using adaptive FL in healthcare industry 5.0. *Scientific Reports*, *14*(1), 18643.

Sharma, S., Aggarwal, A. & Choudhury, T. Breast cancer detection using machine learning algorithms. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* 2018 (pp. 114–118). IEEE.

Abbas, T., Khan, A. H., Kanwal, K., Daud, A., Irfan, M., Bukhari, A., & Alharbey, R. (2024). IoMT-Based Healthcare Systems: A Review. *Computer Systems Science & Engineering*, *48*(4).

Al-Tarawneh, A. M., AlOmoush, R. A., ul Islam, T., Janjua, J. I., Abbas, T., & Ihsan, A. (2024, December). Current Trends in Artificial Intelligence for Educational Advancements. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1-6). IEEE.

Janjua, J. I., Ahmad, R., Abbas, S., Mohammed, A. S., Khan, M. S., Daud, A., ... & Khan, M. A. (2024). Enhancing smart grid electricity prediction with the fusion of intelligent modeling and XAI integration. *International Journal of Advanced and Applied Sciences*, *11*(5), 230-248

Thilaka, A. & Sundaravalli, E. Breast Cancer forecasting using machine learning algorithms. *Int. J. Data Inf. Intell. Comput.* **2** (3), 11–20 (2023).

Rehman, A., Noor, F., Janjua, J. I., Ihsan, A., Saeed, A. Q., & Abbas, T. (2024, December). Classification of Lung Diseases Using Machine Learning Technique. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1-7). IEEE.

Silva-Aravena, F., Núñez Delafuente, H., Gutiérrez-Bahamondes, J. H. & Morales, J. A hybrid algorithm of Ml and Xai to prevent breast cancer: A strategy to support decision making. *Cancers* **15** (9), 2443 (2023).

. Munshi, R. M. et al. A novel approach for breast cancer detection using optimized ensemble learning framework and XAI. *Image Vis. Comput.* **142**, 104910 (2024).

. Islam, T. et al. Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Sci. Rep.* **14** (1), 8487 (2024).

Imouokhome, F. A., Ehimiyein, O. G. & Chete, F. O. Diagnosis and interpretation of breast Cancer using explainable artificial intelligence. *NIPES-J. Sci. Technol. Res.* **5**(2). (2023).

Eteng, I. et al. UCTH Breast Cancer Dataset, V2, (2023). https://doi.org/10.17632/63fpbc9cm4.2

Şahin, M. & Aybek, E. Jamovi: an easy to use statistical software for the social scientists. *Int. J. Assess. Tools Educ.* **6** (4), 670–692 (2019).

Khan, T. (2023). AN AGRICULTURAL INTERNET OF THINGS (A-IOT) BASED INTELLIGENT SYSTEM FOR DISEASE PREDICTION USING TRANSFER LEARNING, A CASE STUDY. Lahore Garrison University Research Journal of Computer Science and Information Technology, 7(3).

Almansour, B. Y., Almansour, A. Y., Janjua, J. I., Zahid, M., & Abbas, T. (2024, December). Application of Machine Learning and Rule Induction in Various Sectors. In 2024 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 1-8). IEEE.

Rattan, V., Mittal, R., Singh, J. & Malik, V. Analyzing the application of SMOTE on machine learning classifiers. In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)* 2021 Mar 5 (pp. 692–695). IEEE.

Nguyen, H. T., Cao, H. Q., Nguyen, K. V. & Pham, N. D. Evaluation of explainable artificial intelligence: Shap, lime, and cam. InProceedings of the FPT AI Conference 2021 May (pp. 1–6).

. Khater, T., Hussain, A., Mahmoud, S. & Yasen, S. Explainable ai for breast cancer detection: A lime-driven approach. In *2023 16th International Conference on Developments in eSystems Engineering (DeSE)* 2023 Dec 18 (pp. 540–545). IEEE.

. Vishwarupe, V. et al. Explainable AI and interpretable machine learning: A case study in perspective. *Procedia Comput. Sci.* **204**, 869–876 (2022).

. Wenninger, S., Kaymakci, C. & Wiethe, C. Explainable long-term Building energy consumption prediction using qlattice. *Appl. Energy*. **308**, 118300 (2022).

Ribeiro, M. T., Singh, S., Guestrin, C. & Anchors High-precision model-agnostic explanations. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1). (2018).

Ghasemi, A., Hashtarkhani, S., Schwartz, D. L. & Shaban-Nejad, A. Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review. *Cancer Innov.* **3** (5), e136 (2024).

Novielli, P. et al. Personalized identification of autism-related bacteria in the gut Microbiome using explainable artificial intelligence. *Iscience* ;**27**(9). (2024).

Hussain, S. et al. Breast cancer risk prediction using machine learning: A systematic review. *Front. Oncol.* **14**, 1343627 (2024).

Luo, L. et al. Deep learning in breast cancer imaging: A decade of progress and future directions. *IEEE Rev. Biomed. Eng.* (2024).

Grinet, M. A., Gouveia, A. I. & Gomes, A. J. Machine learning in breast cancer imaging: A review on data, models and methods. *Comput. Methods Biomech. Biomed. Eng. Imaging Visual.* **11** (7), 2302387 (2024).

30 . Anastasi, G. et al. Machine learning techniques in breast cancer preventive diagnosis: A review. *Multimed. Tools Appl.* **12**, 1–44 (2024).

Frasca, M., La Torre, D., Pravettoni, G. & Cutica, I. Explainable and interpretable artificial intelligence in medicine: A systematic bibliometric review. *Discov. Artif. Intell.* **4** (1), 15 (2024).

Rehman, A., Mahmood, T. & Saba, T. Robust kidney carcinoma prognosis and characterization using Swin-ViT and DeepLabV3 + with multi-model transfer learning. *Appl. Soft Comput.* **170**, 112518 (2025).

Ali, A. M., Benjdira, B., Koubaa, A., Boulila, W. & El-Shafai, W. TESR: Two-stage approach for enhancement and super-resolution of remote sensing images. *Remote Sens.* **15** (9), 2346 (2023).

Mahmood, T., Rehman, A., Saba, T., Wang, Y. & Alamri, F. S. Alzheimer's disease unveiled: Cutting-edge multi-modal neuroimaging and computational methods for enhanced diagnosis. *Biomed. Signal Process. Control*. **97**, 106721 (2024).

Mahmood, T., Saba, T., Rehman, A. & Alamri, F. S. Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. *Expert Syst. Appl.* **249**, 123747 (2024).