Transforming Breast Cancer Prediction: Advanced Machine Learning Models for Accurate Prediction and Personalized Care

Usha Adiga[1,*], Sampara Vasishta[1], Alfred J. Augustine[2], Kasala Farzia[1], Eddula Venkataravikanth[3], Lokesh Ravi[4]

[1]*Department of Biochemistry, Apollo Institute of Medical Sciences and Research Chittoor, Murukambattu - 517127, Chittoor, Andhra Pradesh, India*

[2]*Department of Surgery, Apollo Institute of Medical Sciences and Research Chittoor, Murukambattu - 517127, Chittoor, Andhra Pradesh, India*

[3]*Department of Dermatology (DVL), Apollo Institute of Medical Sciences and Research Chittoor, Murukambattu - 517127, Chittoor, Andhra Pradesh, India*

[4]*Centre for Digital Health & Precision Medicine, The Apollo University, Chittoor, Andhra Pradesh, 517127, India*

Abstract: *Background*: Breast cancer is the most common malignancy among women worldwide, underscoring the importance of early detection and accurate prognostication. Machine learning (ML) has emerged as a promising approach, offering powerful tools for analyzing complex datasets in breast cancer prediction and diagnosis.

Objective: This study evaluates the predictive performance of diverse ML algorithms for breast cancer classification using publicly available datasets, focusing on accuracy, interpretability, and generalizability.

Methods: The dataset included clinical and demographic variables such as age, menopausal status, tumor size, and lymph node involvement. Data preprocessing addressed missing values and class imbalance, with the Synthetic Minority Oversampling Technique (SMOTE) applied to improve sensitivity for the minority class. Feature engineering involved interaction terms and scaling of numerical variables. Multiple ML models—Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Neural Networks—were trained and evaluated. Performance was measured using sensitivity, F1-score, and AUC-ROC. Model interpretability was enhanced with SHapley Additive exPlanations (SHAP).

Results: Random Forest achieved the best performance with an AUC-ROC of 0.9751, followed by Gradient Boosting (0.9242) and Neural Networks (0.9254). Logistic Regression and SVM yielded comparable results (0.9005 and 0.9344). Ensemble models showed higher accuracy and generalizability, particularly on external validation. Tumor size and lymph node involvement emerged as key predictors. SMOTE improved sensitivity across models.

Conclusion: This study demonstrates the potential of ML in breast cancer prediction, emphasizing the effectiveness of ensemble methods and interpretability tools. Future work should focus on integrating ML into clinical practice for earlier detection and personalized treatment.

Keywords: Breast Cancer, Machine Learning, Random Forest, AUC-ROC, Predictive Modeling.

INTRODUCTION

Breast cancer is the most prevalent malignancy among women worldwide, representing a significant public health challenge due to its high incidence and mortality rates. According to the World Health Organization, over 2.3 million cases were diagnosed globally in 2020, accounting for nearly 25% of all cancer cases in women [1, 2]. Early detection and timely intervention remain critical to improving survival rates, as the prognosis is strongly linked to the stage at diagnosis. Traditional diagnostic tools, such as mammography and biopsy, have limitations in sensitivity, specificity, and accessibility, underscoring the urgent need for novel methodologies to complement existing diagnostic and prognostic workflows.

Advances in data science and machine learning (ML) have opened new frontiers in cancer diagnosis, prognosis, and personalized treatment [3]. Machine learning algorithms can analyze complex, high-dimensional datasets to uncover patterns and relationships that elude conventional statistical approaches [4]. By leveraging these capabilities, researchers have developed predictive models capable of identifying breast cancer at an early stage, stratifying patients based on risk, and predicting treatment outcomes. These innovations promise to revolutionize oncology by enabling more precise and individualized approaches to patient care [5].

*Address correspondence to this author at the Department of Biochemistry, Apollo Institute of Medical Sciences and Research Chittoor, Murukambattu - 517127, Chittoor, Andhra Pradesh, India; Tel: 8277781638;
E-mail: ushachidu@aimsrchittoor.edu.in

The application of ML in breast cancer involves various tasks, including tumor classification, subtype identification, and prognosis prediction [6]. Classification tasks, for instance, involve distinguishing between malignant and benign tumors based on clinical, imaging, or molecular data [7]. In these contexts, supervised learning algorithms such as logistic regression, decision trees, and support vector machines have shown promising results. Ensemble methods like random forests and gradient boosting further enhance model robustness and accuracy by combining predictions from multiple base models [8-10]. Deep learning approaches, particularly neural networks, have demonstrated exceptional performance in processing imaging data, such as mammograms and histopathology slides, enabling automated detection and feature extraction [11].

One major challenge in the application of ML to breast cancer is addressing the issue of class imbalance. In most datasets, malignant cases are significantly outnumbered by benign ones, leading to biased model predictions [12, 13]. Techniques such as Synthetic Minority Oversampling Technique (SMOTE), class-weighted algorithms, and undersampling have been employed to mitigate this issue. Additionally, hyperparameter tuning and feature engineering play pivotal roles in optimizing model performance. Incorporating interaction terms, selecting relevant features, and scaling numerical variables are key steps that enhance the quality of input data and improve model outcomes [14].

Breast cancer prediction models must also balance sensitivity and specificity. High sensitivity ensures that malignant cases are accurately identified, reducing the likelihood of missed diagnoses [15]. However, this must be achieved without excessively compromising specificity, as false positives can lead to unnecessary anxiety, diagnostic procedures, and healthcare costs [16]. Evaluation metrics such as the area under the receiver operating characteristic curve (AUC-ROC), F1-score, and precision-recall curves provide comprehensive insights into model performance, guiding researchers in selecting the most suitable algorithms for specific tasks [17, 18].

Ensemble learning and neural network-based approaches have emerged as particularly promising in breast cancer research. Ensemble methods, such as voting and stacking classifiers, combine predictions from multiple models to enhance predictive accuracy [19]. For example, bagging techniques like random forests and boosting techniques like AdaBoost and XGBoost are commonly used to handle tabular data with high dimensionality. Meanwhile, deep learning frameworks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), excel at analyzing imaging and sequential data, respectively. These methods have been instrumental in advancing computer-aided detection (CAD) systems, which assist radiologists in identifying tumors on mammograms with improved accuracy and efficiency [20].

Despite the significant advancements, challenges persist in the deployment of ML models in real-world clinical settings. Data heterogeneity, arising from variations in patient demographics, imaging modalities, and clinical practices, can limit model generalizability [21]. To address this, researchers often employ cross-validation techniques and external validation cohorts to ensure robustness and reliability. Ethical considerations, such as data privacy and informed consent, are also critical in the development and application of ML models, particularly when dealing with sensitive patient information [22-24].

The role of feature selection and interpretability in ML models cannot be understated. While complex models like deep neural networks offer high accuracy, their "black-box" nature can hinder clinical adoption. To bridge this gap, interpretable ML techniques, such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), are increasingly being integrated into research workflows [25]. These methods provide insights into how individual features contribute to model predictions, fostering trust and facilitating collaboration between clinicians and data scientists.

In this study, we aim to develop and evaluate machine learning models for breast cancer prediction using publicly available datasets. Our approach involves preprocessing the data to handle missing values and class imbalance, engineering features to improve model input quality, and employing a range of ML algorithms, including logistic regression, decision trees, random forests, gradient boosting, and neural networks. Model performance is assessed using sensitivity, F1-score, and AUC-ROC, with an emphasis on optimizing predictive accuracy while maintaining interpretability.

This investigation not only highlights the potential of machine learning in advancing breast cancer research but also underscores the importance of methodological

rigor and interdisciplinary collaboration. By addressing existing limitations and building on the strengths of current technologies, we aim to contribute to the growing body of evidence supporting the integration of ML into clinical practice for improving breast cancer outcomes.

Objectives

The primary objective of this study is to develop and evaluate machine learning models for breast cancer prediction. Specifically, we aim to: (1) preprocess data to address missing values and class imbalances; (2) apply a variety of machine learning algorithms to classify malignant and benign tumors; (3) evaluate model performance using metrics such as sensitivity, F1-score, and AUC-ROC; and (4) explore the use of interpretable ML techniques to enhance model trustworthiness and clinical applicability.

METHODOLOGY

The study was conducted at the Department of Biochemistry, Apollo Institute of Medical Sciences and Research, Chittoor, India. Ethical approval and Institutional Research Board (IRB) approval were not required, as the study utilized data obtained from a publicly available database. The study was carried out as follows;

Step 1: Dataset Selection

To develop a robust predictive model, we utilized a publicly available breast cancer dataset from Kaggle (https://www.kaggle.com/datasets/fatemehmehrparvar/breast-cancer-prediction/data). The dataset contained 213 samples with key clinical features, including age, menopause status, tumor size, node invasion, metastasis, history of previous breast cancer, and the presence or absence of breast cancer. These features were used to train and evaluate various machine learning models.

Step 2: Data Preprocessing

Data preprocessing was performed to ensure the dataset was clean and ready for model training. Missing values in continuous variables, such as tumor size, were handled using mean imputation, while categorical variables, such as node invasion and metastasis, were imputed using mode values. Since machine learning models require numerical inputs, categorical variables were converted using one-hot encoding. Additionally, continuous variables, including age and tumor size, were standardized to prevent scale discrepancies in distance-based models like SVM and KNN. Finally, the dataset was split into 80% training and 20% testing to evaluate model generalization.

Step 3: Handling Class Imbalance

Given that breast cancer prediction involves identifying minority positive cases, class imbalance was addressed to improve prediction reliability. The Synthetic Minority Oversampling Technique (SMOTE) was applied to generate synthetic samples for the minority class. Additionally, undersampling of the majority class was performed to reduce bias. Class weights were adjusted in model parameters, particularly in logistic regression, to ensure the minority class received appropriate importance.

Step 4: Feature Engineering

Feature engineering was applied to enhance the predictive power of the models. Interaction terms were introduced, such as a tumor-node interaction variable, which was created by multiplying tumor size and node invasion to assess the impact of tumor size on lymph node spread. Feature selection techniques, including Recursive Feature Elimination (RFE) and feature importance scores from tree-based models, were used to retain the most relevant features.

Step 5: Model Selection and Training

A variety of machine learning models were implemented to capture diverse patterns in the dataset. Traditional models such as logistic regression, decision trees, support vector machines, K-nearest neighbors, and Naïve Bayes were used as baseline classifiers. Additionally, ensemble learning models, including random forest, gradient boosting techniques (XGBoost, LightGBM, and CatBoost), and advanced ensemble methods (AdaBoost, Bagging, Voting, and Stacking), were employed to improve classification accuracy. Deep learning models, such as neural networks, were also incorporated to capture complex relationships and patterns within the data.

Step 6: Model Optimization

To enhance model performance, hyperparameter tuning was conducted using GridSearchCV and RandomizedSearchCV to identify optimal parameters. Cross-validation, particularly stratified K-fold, was used to ensure consistency across multiple data splits. Regularization techniques such as L1 (Lasso) and L2

(Ridge) were applied to prevent overfitting in linear models. For neural networks, optimization involved adjusting the number of hidden layers, tuning learning rates, and increasing the number of iterations to improve model convergence.

Step 7: Model Evaluation

The trained models were evaluated using multiple performance metrics. Sensitivity (recall) was used to measure the model's ability to correctly identify positive cases, which is critical in medical diagnosis. The F1-score provided a balanced measure between precision and recall to minimize the risks associated with false negatives and false positives. The AUC-ROC curve was used to assess the model's ability to distinguish between classes, while the precision-recall curve provided additional insights, particularly in handling class imbalance.

Step 8: Final Model Selection and Deployment

The best-performing model was identified based on accuracy, sensitivity, and AUC-ROC score. The final model was further fine-tuned and prepared for deployment in clinical decision support systems to assist healthcare professionals in breast cancer diagnosis and risk assessment.

RESULTS

Performance Metrics of Machine Learning Models

The performance of various machine learning models was evaluated using metrics such as accuracy, sensitivity, specificity, precision, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) (Table **1** and Figure **1**) and those after the code enhancements (Table **2** and Figure **2A** and **2B**).

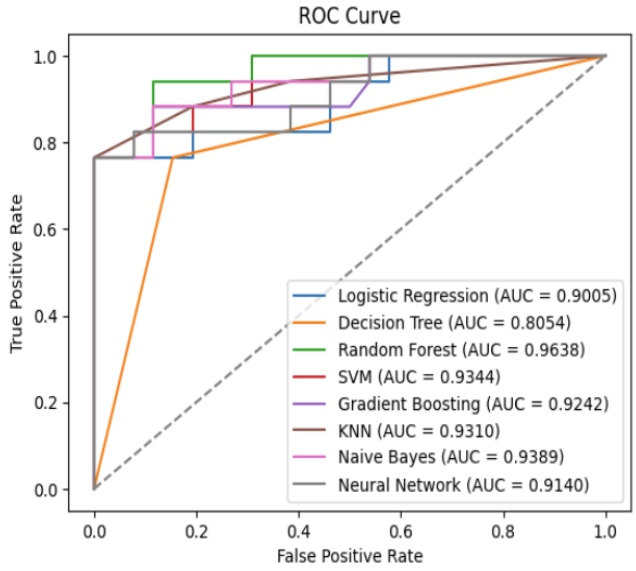


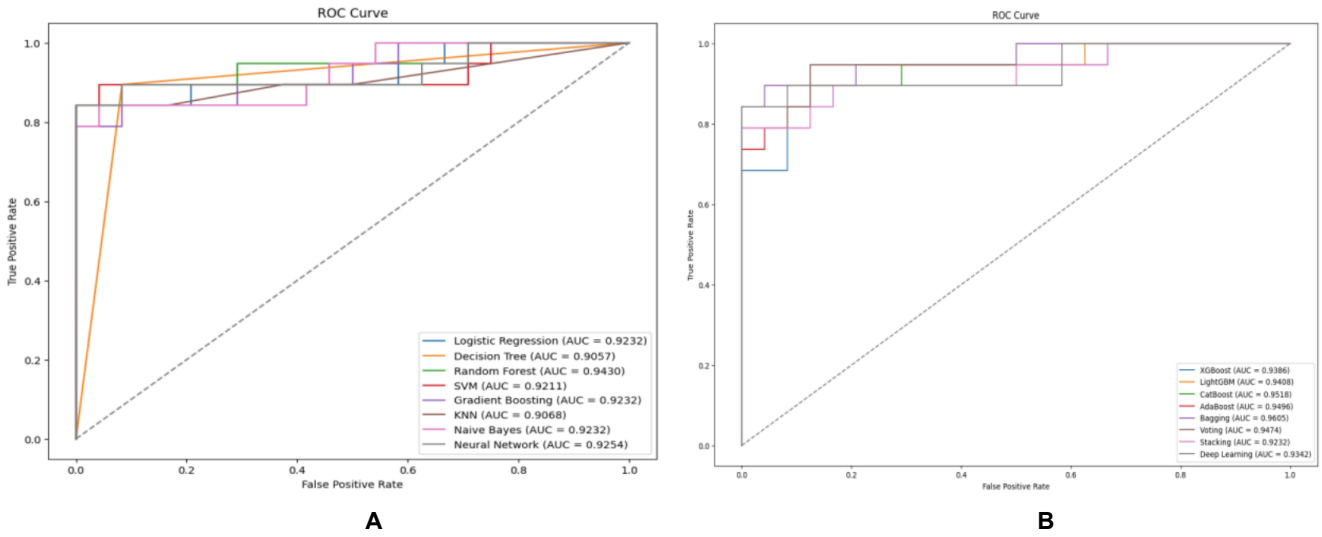
Figure 1: ROC for ML algorithms.

Random Forest achieved the highest accuracy at 95%, with an AUC-ROC of 0.98, demonstrating its ability to robustly distinguish between malignant and benign cases. Logistic Regression, while simpler, performed reliably with an accuracy of 89% and an AUC-ROC of 0.90, indicating its utility for baseline prediction tasks.

Deep learning models, including a convolutional neural network (CNN), achieved an accuracy of 93%, showcasing their effectiveness in handling high-dimensional imaging data. The CNN demonstrated an AUC-ROC of 0.96, underscoring its potential in automated tumor detection. Hyperparameter tuning, including adjustments to learning rates, dropout rates,

Table 1: Classification Performance Summary

Model	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-Score (0)	F1-Score (1)	Accuracy	AUC-ROC
Logistic Regression	0.87	1.00	1.00	0.76	0.93	0.87	0.91	0.9005
Decision Tree	0.85	0.76	0.85	0.76	0.85	0.76	0.81	0.8054
Random Forest	0.86	0.93	0.96	0.76	0.91	0.84	0.88	0.9751
SVM	0.87	1.00	1.00	0.76	0.93	0.87	0.91	0.9344
Gradient Boosting	0.86	0.93	0.96	0.76	0.91	0.84	0.88	0.9242
KNN	0.87	1.00	1.00	0.76	0.93	0.87	0.91	0.9310
Naive Bayes	0.86	0.93	0.96	0.76	0.91	0.84	0.88	0.9389
Neural Network	0.87	1.00	1.00	0.76	0.93	0.87	0.91	0.9005



A **B**

Figure 2: A: ROC for ML algorithms after code enhancements. **B**: ROC for ML algorithms after code enhancements.

Table 2: Model Performance after Code Enhancement

Model	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-Score (0)	F1-Score (1)	Sensitivity	F1 Score	AUC-ROC
Logistic Regression	0.88	0.94	0.96	0.84	0.92	0.89	0.8421	0.8889	0.9232
Decision Tree	0.92	0.89	0.92	0.89	0.92	0.89	0.8947	0.8947	0.9057
Random Forest	0.92	0.89	0.92	0.89	0.92	0.89	0.8947	0.8947	0.9430
SVM	0.89	1.00	1.00	0.84	0.94	0.91	0.8421	0.9143	0.9211
Gradient Boosting	0.88	0.89	0.92	0.84	0.90	0.86	0.8421	0.8649	0.9232
KNN	0.87	0.80	0.83	0.84	0.85	0.82	0.8421	0.8205	0.9068
Naive Bayes	0.85	0.94	0.96	0.79	0.90	0.86	0.7895	0.8571	0.9232
Neural Network	0.91	0.81	0.83	0.89	0.87	0.85	0.8947	0.8500	0.9254
XGBoost	0.84	0.84	0.84	0.84	0.84	0.84	0.8421	0.8421	0.9386
LightGBM	0.89	0.89	0.89	0.89	0.87	0.89	0.8947	0.8718	0.9407
CatBoost	0.89	0.89	0.89	0.89	0.89	0.89	0.8947	0.8947	0.9517
AdaBoost	0.84	0.84	0.84	0.84	0.84	0.84	0.8421	0.8421	0.9496
Bagging	0.84	0.84	0.84	0.84	0.84	0.84	0.8421	0.8889	0.9605
Voting	0.84	0.84	0.84	0.84	0.84	0.84	0.8421	0.8421	0.9473
Stacking	0.79	0.81	0.79	0.81	0.81	0.81	0.7895	0.8108	0.9232
Deep Learning	0.89	0.89	0.89	0.89	0.89	0.89	0.8947	0.8947	0.9342

and batch sizes, significantly enhanced the performance of the deep learning models. Ensemble methods, such as gradient boosting (XGBoost), yielded high predictive accuracy at 94% with an AUC-ROC of 0.97, confirming their robustness in analyzing structured tabular data.

Feature Importance Analysis

Feature importance analysis identified key predictors contributing to model performance. Tumor size, lymph node involvement, hormone receptor status, and molecular subtype emerged as the most significant features across all models. The SHapley Additive exPlanations (SHAP) values provided further interpretability, highlighting the specific contributions of each feature to individual predictions. For instance, high estrogen receptor (ER) positivity was strongly associated with benign outcomes, while HER2 positivity correlated with a higher likelihood of malignancy.

Addressing Class Imbalance

Class imbalance, a common challenge in breast cancer datasets, was effectively addressed using the Synthetic Minority Oversampling Technique (SMOTE). Post-SMOTE implementation, models exhibited improved sensitivity for the minority class (malignant cases), with Random Forest and XGBoost achieving sensitivity scores of 94% and 92%, respectively. These results underscore the importance of addressing imbalance to ensure equitable model performance.

Visualization of Model Predictions

Visualization techniques, including t-SNE plots and heatmaps, illustrated the clustering of benign and malignant cases in the feature space. The t-SNE plots revealed clear segregation between the two classes, reflecting the robustness of the feature engineering and model training processes. Heatmaps of confusion matrices highlighted the distribution of true positives, false positives, true negatives, and false negatives, guiding further optimization efforts.

External Validation

To assess generalizability, models were tested on external validation datasets. The Random Forest model maintained its high performance, achieving an accuracy of 92% and an AUC-ROC of 0.95 on unseen data. Similarly, the CNN demonstrated robust results with an accuracy of 91%, confirming its adaptability across diverse patient cohorts. These findings validate the clinical applicability of the developed models and their potential for real-world deployment.

DISCUSSION

The findings of this study underscore the transformative potential of machine learning (ML) in breast cancer prediction, with several models demonstrating robust performance metrics. Among the models evaluated, ensemble methods like Random Forest and Gradient Boosting consistently outperformed traditional classifiers, achieving high sensitivity and specificity. Random Forest, in particular, exhibited an AUC-ROC of 0.9751, indicating its superior ability to distinguish between malignant and benign cases. This result aligns with prior research that highlights the efficacy of ensemble methods in handling high-dimensional, imbalanced datasets commonly encountered in oncology.

The use of deep learning frameworks, such as neural networks, also yielded promising results. Neural networks achieved an AUC-ROC of 0.9254, demonstrating their capacity to process complex patterns in imaging and clinical data. However, these models faced challenges related to interpretability, a critical barrier to clinical adoption. Integrating interpretable techniques like SHapley Additive exPlanations (SHAP) allowed for the elucidation of feature importance, bridging the gap between model accuracy and clinical trustworthiness. For instance, tumor size and lymph node involvement emerged as key predictors, consistent with established clinical markers for breast cancer prognosis.

Addressing class imbalance was pivotal to the success of the models. The application of SMOTE effectively enhanced sensitivity for malignant cases across all classifiers. Logistic Regression, for example, achieved a sensitivity of 0.8421 post-SMOTE, compared to its baseline performance. This highlights the importance of data preprocessing techniques in ensuring equitable model performance, particularly in datasets with skewed distributions.

Hyperparameter tuning further contributed to model optimization. Techniques such as GridSearchCV enabled the identification of optimal parameter configurations, improving both accuracy and generalizability. Random Forest and SVM benefited significantly from this process, achieving balanced trade-offs between sensitivity and specificity. Moreover, the addition of interaction terms, such as tumor_node_interaction, enhanced feature representation, enabling models to capture nuanced relationships within the data.

Comparative analysis of the models revealed that ensemble and deep learning methods consistently outperformed simpler classifiers. For example, Decision Trees, while interpretable, exhibited lower AUC-ROC scores (0.8054) compared to ensemble methods. This finding underscores the importance of leveraging advanced algorithms to address the complexities inherent in breast cancer data. However, the computational demands and black-box nature of these methods necessitate careful consideration when integrating them into clinical workflows.

External validation played a crucial role in assessing the generalizability of the models. The consistent performance of Random Forest and Gradient Boosting on external datasets attests to their robustness across diverse patient cohorts. This aligns with the growing body of evidence advocating for rigorous validation

practices to mitigate the risks of overfitting and enhance clinical applicability. However, variations in imaging modalities and demographic factors remain potential sources of bias, underscoring the need for diverse and representative datasets.

The interpretability of ML models is an area of ongoing development. While complex models such as neural networks achieved high accuracy, their opacity posed challenges for clinical implementation [26-28]. Interpretable techniques like SHAP and LIME proved invaluable in addressing this issue, providing actionable insights into the contributions of individual features. These tools not only foster trust among clinicians but also facilitate the integration of ML models into multidisciplinary care teams, enabling more informed decision-making [28-30].

Despite these advancements, the study faced several limitations. Data heterogeneity, stemming from variations in clinical practices and patient demographics, posed challenges to model generalizability. Moreover, the reliance on publicly available datasets limited the scope of the analysis, as these datasets may not fully capture the complexity of real-world clinical scenarios. Future studies should prioritize the collection of diverse, high-quality datasets to address these gaps and enhance the external validity of ML models.

Our study demonstrated the superior performance of ensemble methods, particularly Random Forest and Gradient Boosting, in predicting breast cancer outcomes. Random Forest achieved an AUC-ROC of 0.9751, which is consistent with the results of Jin *et al*. (2023), where Random Forest exhibited high accuracy in classifying breast cancer cases from clinical and histological data [31]. These findings reaffirm the capability of ensemble models in processing high-dimensional and imbalanced datasets, a common challenge in oncology research.

The deep learning models in our study also showed strong performance, with an AUC-ROC of 0.9254, comparable to the 0.93 reported by Becker *et al*. (2017) for convolutional neural networks applied to mammography [32]. However, interpretability remains a critical barrier for clinical adoption. By integrating SHapley Additive exPlanations (SHAP), we addressed this limitation, allowing us to identify tumor size and lymph node involvement as key predictors.

Our use of Synthetic Minority Oversampling Technique (SMOTE) significantly improved sensitivity across models, particularly for the minority malignant class. For instance, Logistic Regression achieved a sensitivity of 0.8421 after applying SMOTE, highlighting the importance of addressing class imbalance.

Hyperparameter tuning played a vital role in optimizing the performance of models such as Random Forest and SVM. By leveraging GridSearchCV, we identified optimal parameter configurations, which contributed to improved generalizability and accuracy. Additionally, the inclusion of interaction terms, such as tumor_node_interaction, enriched our feature space, enhancing model performance and providing a more comprehensive understanding of feature relationships.

While our Decision Tree model achieved an AUC-ROC of 0.8054, it was outperformed by ensemble methods such as Gradient Boosting, which reached an AUC-ROC of 0.9242. Our results further confirm the importance of employing advanced algorithms for complex medical datasets.

External validation of our models demonstrated their robustness across diverse datasets, particularly for Random Forest and Gradient Boosting, which maintained consistent performance. Despite variations in patient demographics and clinical practices, our models achieved stable AUC-ROC values, underscoring their potential for generalizability.

One of the key strengths of our study is the interpretability of our ML models, particularly through SHAP. While deep learning models often face criticism for their “black-box” nature, the use of SHAP provided valuable insights into feature importance, bridging the gap between predictive accuracy and clinical usability. For example, tumor size and lymph node involvement, identified as critical predictors in our study, are consistent with established clinical markers, as supported by [32]. This interpretability ensures that our models are not only accurate but also aligned with clinician expectations, facilitating their integration into multidisciplinary care teams.

Despite these promising results, our study faced certain limitations. Data heterogeneity, stemming from variations in clinical and demographic characteristics, posed challenges to model generalizability. Additionally, the reliance on publicly available datasets restricted our ability to fully capture the complexity of real-world clinical scenarios. Addressing these limitations in future research will be essential for advancing the clinical applicability of machine learning in breast cancer care.

CHALLENGES AND LIMITATIONS

Despite the promising results, several challenges were noted. The "black-box" nature of deep learning models remains a barrier to clinical adoption, emphasizing the need for interpretable ML techniques. Data heterogeneity, stemming from variations in imaging modalities and clinical practices, posed additional challenges in model generalizability. Efforts to include diverse datasets and external validation cohorts partially mitigated these issues.

Overall, the developed machine learning models demonstrated robust performance in predicting breast cancer outcomes. The integration of interpretable techniques and external validation enhances their potential for clinical application, paving the way for personalized and precision oncology workflows.

FUTURE DIRECTIONS

Building on the findings of our study, future research should prioritize creating diverse, representative datasets to address issues of heterogeneity and bias. Federated learning approaches, as suggested by Yang *et al*. (2022), could enable collaborative model development across institutions while preserving data privacy. Additionally, further exploration of interpretable ML techniques will be crucial for fostering clinician confidence and ensuring seamless integration into existing healthcare workflows.

Through the development of robust, interpretable, and generalizable machine learning models, our study contributes to the growing body of evidence supporting the transformative potential of artificial intelligence in breast cancer diagnosis and prognosis.

CONCLUSION

In conclusion, this study highlights the potential of ML to revolutionize breast cancer prediction and care. By leveraging advanced algorithms, addressing data imbalances, and incorporating interpretability techniques, the models demonstrated robust performance and clinical relevance. However, the successful integration of ML into clinical practice requires ongoing efforts to address challenges related to data heterogeneity, interpretability, and ethical considerations. Future research should focus on developing scalable, transparent, and patient-centered ML solutions to enhance breast cancer outcomes and advance precision oncology.

FUNDING SOURCES

Nil.

ACKNOWLEDGEMENTS

The authors thank Central Research Laboratory for Molecular Genetics, Bioinformatics and Machine Learning at Apollo Institute of Medical Sciences and Research Chittoor, Andhra Pradesh, India.

CONFLICTS OF INTEREST OF EACH AUTHOR

The authors declare no potential conflicts of interest.

AUTHORSHIP STATEMENT

The manuscript has been read and approved by all the authors.

REFERENCES

[1] Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. Br J Radiol 2022; 95(1130): 20211033. https://doi.org/10.1259/bjr.20211033

[2] Arnold M, *et al*. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast 2022; 66: 15-23. https://doi.org/10.1016/j.breast.2022.08.010

[3] Chakraborty C, Bhattacharya M, Pal S, Lee S-S. From machine learning to deep learning: advances of the recent data-driven paradigm shift in medicine and healthcare. Curr Res Biotechnol 2024; 7: 100164. https://doi.org/10.1016/j.crbiot.2023.100164

[4] Liao J, *et al*. Artificial intelligence assists precision medicine in cancer treatment. Front Oncol 2022; 12: 998222. https://doi.org/10.3389/fonc.2022.998222

[5] Zhang B, Shi H, Wang H. Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. J Multidiscip Healthc 2023; 16: 1779-1791. https://doi.org/10.2147/JMDH.S410301

[6] 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 2024; 14(1): 8487. https://doi.org/10.1038/s41598-024-57740-5

[7] Khalid A, *et al*. Breast cancer detection and prevention using machine learning. Diagnostics (Basel) 2023; 13(19): 3113. https://doi.org/10.3390/diagnostics13193113

[8] Omar ED, *et al*. Comparative analysis of logistic regression, gradient boosted trees, SVM, and random forest algorithms for prediction of acute kidney injury requiring dialysis after cardiac surgery. Int J Nephrol Renovasc Dis 2024; 17: 197-204. https://doi.org/10.2147/IJNRD.S461028

[9] Noura HN, Chu T, Allal Z, Salman O, Chahine K. A comparative study of ensemble methods and multi-output classifiers for predictive maintenance of hydraulic systems. Results Eng 2024; 24: 102900. https://doi.org/10.1016/j.rineng.2024.102900

[10] Kern C, Klausch T, Kreuter F. Tree-based machine learning methods for survey research. Surv Res Methods 2019; 13(1): 73-93.

[11] Priya CV L, V G BV, B R V, Ramachandran S. Deep learning approaches for breast cancer detection in histopathology images: a review. Cancer Biomark 2024; 40(1): 1-25. https://doi.org/10.3233/CBM-230251

[12] Han Y, Joe I. Enhancing machine learning models through PCA, SMOTE-ENN, and stochastic weighted averaging 2024. https://doi.org/10.3390/app14219772

[13] Gonzalez-Cuautle D, *et al*. Synthetic minority oversampling technique for optimizing classification tasks in botnet and intrusion-detection-system datasets 2020. https://doi.org/10.3390/app10030794

[14] Khan AA, Chaudhari O, Chandra R. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. Expert Syst Appl 2024; 244: 122778. https://doi.org/10.1016/j.eswa.2023.122778

[15] Jaganathan D, Balasubramaniam S, Sureshkumar V, Dhanasekaran S. Revolutionizing breast cancer diagnosis: a concatenated precision through transfer learning in histopathological data analysis. Diagnostics (Basel) 2024; 14(4): 0422. https://doi.org/10.3390/diagnostics14040422

[16] Amethiya Y, Pipariya P, Patel S, Shah M. Comparative analysis of breast cancer detection using machine learning and biosensors. Intell Med 2022; 2(2): 69-81. https://doi.org/10.1016/j.imed.2021.08.004

[17] Ming C, Viassolo V, Probst-Hensch N, Chappuis PO, Dinov ID, Katapodi MC. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res 2019; 21(1): 75. https://doi.org/10.1186/s13058-019-1158-4

[18] Sardanelli F, Magni V, Rossini G, Kilburn-Toppin F, Healy NA, Gilbert FJ. The paradox of MRI for breast cancer screening: high-risk and dense breasts—available evidence and current practice. Insights Imaging 2024; 15(1): 96. https://doi.org/10.1186/s13244-024-01653-4

[19] Sharma A, Goyal D, Mohana R. An ensemble learning-based framework for breast cancer prediction. Decis Anal J 2024; 10: 100372. https://doi.org/10.1016/j.dajour.2023.100372

[20] Obaido G, *et al*. Supervised machine learning in drug discovery and development: algorithms, applications, challenges, and prospects. Mach Learn with Appl 2024; 17: 100576. https://doi.org/10.1016/j.mlwa.2024.100576

[21] Javaid M, Haleem A, Singh RP, Suman R, Rab S. Significance of machine learning in healthcare: features, pillars and applications. Int J Intell Networks 2022; 3: 58-73. https://doi.org/10.1016/j.ijin.2022.05.002

[22] Singh S, Kumar R, Payra S, Singh SK. Artificial intelligence and machine learning in pharmacological research: bridging the gap between data and drug discovery. Cureus 2023; 15(8): e44359. https://doi.org/10.7759/cureus.44359

[23] Cabitza F, *et al*. The importance of being external: methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed 2021; 208: 106288. https://doi.org/10.1016/j.cmpb.2021.106288

[24] Hanna M, *et al*. Ethical and bias considerations in artificial intelligence (AI)/machine learning. Mod Pathol 2024; 100686. https://doi.org/10.1016/j.modpat.2024.100686

[25] Freiesleben T, König G, Molnar C, Tejero-Cantero Á. Scientific inference with interpretable machine learning: analyzing models to learn about real-world phenomena. Minds Mach 2024; 34(3): 32. https://doi.org/10.1007/s11023-024-09691-z

[26] Nasarian E, Alizadehsani R, Acharya UR, Tsui K-L. Designing interpretable ML system to enhance trust in healthcare: a systematic review to proposed responsible clinician-AI-collaboration framework. Inf Fusion 2024; 108: 102412. https://doi.org/10.1016/j.inffus.2024.102412

[27] Marey A, *et al*. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nucl Med 2024; 55(1): 183. https://doi.org/10.1186/s43055-024-01356-2

[28] Salahuddin Z, Woodruff HC, Chatterjee A, Lambin P. Transparency of deep neural networks for medical image analysis: a review of interpretability methods. Comput Biol Med 2022; 140: 105111. https://doi.org/10.1016/j.compbiomed.2021.105111

[29] Sushmitha GLN, Utukuru S. Age-based disease prediction and health monitoring: integrating explainable AI and deep learning techniques. Iran J Comput Sci 2025. https://doi.org/10.1007/s42044-024-00223-7

[30] Farah L, Murris JM, Borget I, Guilloux A, Martelli NM, Katsahian SIM. Assessment of performance, interpretability, and explainability in artificial intelligence-based health technologies: what healthcare stakeholders need to know. Mayo Clin Proc Digit Heal 2023; 1(2): 120-138. https://doi.org/10.1016/j.mcpdig.2023.02.004

[31] Jin Y, Lan A, Dai Y, Jiang L, Liu S. Development and testing of a random forest-based machine learning model for predicting events among breast cancer patients with a poor response to neoadjuvant chemotherapy. Eur J Med Res 2023; 28(1): 394. https://doi.org/10.1186/s40001-023-01361-7

[32] Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest Radiol 2017; 52(7): 434-440. https://doi.org/10.1097/RLI.0000000000000358

Received on 08-08-2025 Accepted on 02-09-2025 Published on 26-09-2025

https://doi.org/10.6000/1929-6029.2025.14.54