GRADIVA

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

Multimodal Ensemble Learning Framework for Breast Cancer Diagnosis Using Transfer Learning and Clinical Machine Learning Models

Okeke, Ogochukwu C.[1], Mgbeafulike, Ike J.[2], Ugbor Ihechiluru C.[3], Udensi, Nkechi C.[4], Uba Chioma P.[5], Nwajiaku, Love I.[6] and Ebere Uzoka Chidi[7]

1,2,3,4,5,6.Department of Computer Science, Faculty of Physical Sciences,
Chukwuemeka Odumegwu Ojukwu University Uli, Anambra State.
3.Department of Cyber Security, School of Information and Communication Technology,
Federal University of Technology, Owerri. Nigeria.
4.Department of Computer Science, Faculty of Natural Sciences, Hezekiah University Umudi, Imo State, Nigeria.
6.Department of Computer Science, School of Sciences, Federal College of Education (Technical) Umunze,
Anambara, State, Nigeria.
7.African Centre of Excellence for Sustainable Power and Energy Development,
Department of Control and Instrumentation Engineering, University of Nigeria Nsukka, Enugu, Nigeria.
7.State University of Medical Science, Oba, Nsukka, Enugu State, Nigeria.
7.Gregory University, Uturu, Abia State, Nigeria.
Email: [1]Co.okeke@coou.edu.ng, [2]ij.mgbeafulike@coou.edu.ng, [3]ugbor.ihechiluru@futo.edu.ng,[4]nkudensi2014@gmail.com,
[5]Chiomapuba@gmail.com, [6]Nwajiakulove2017@ gmail.com, [7]steverolans@gmail.com

Abstract

Breast cancer remains one of the leading causes of cancer-related death among women all over the world. Early and accurate diagnosis provides potential for better management. From the literature reviewed, several existing systems relied mostly on single data modalities, while others present infrastructure gaps not compatible with health care settings in low and middle-income countries, especially in Africa. The aim of this paper is a multimodal ensemble learning framework for breast cancer diagnosis using transfer learning and clinical machine learning models. The study used three secondary datasets. This includes the mammogram image dataset from Mendeley. The sample size is 614 mammograms, and the classes are benign with 361 images and malignant with 253 images. The second dataset is the clinical records of the 614 participants with 31 attributes attached alongside the mammogram dataset. These first two datasets considered women in South Africa. University of Calabar Teaching Hospital (UCTH) breast cancer dataset, publicly available at Kaggle repository with 11 attributes, was the third dataset. The mammogram dataset was applied to train MobileNetV3, EfficientNet-B0, and SqueezeNet, while two independent clinical datasets from Nigeria and South Africa were applied to train Logistic Regression (LR), Decision Tree (DT), Multi-Layer Perceptron (MLP), and XGBoost. Comparative analysis identified MobileNetV3 as the best transfer learning model with 90% accuracy, LR as the best model for the Nigerian dataset with 93.02% accuracy, and MLP as the best clinical classifier on the South African dataset with 97.56% accuracy. These three models were applied to curate the ensemble model using the hard voting technique and then develop the multimodal ensemble learning framework for breast cancer diagnosis. Experimental results when tested with images and clinical data from low-income country healthcare system showed successful classification robustness and reduced error because the strengths of the medical image and clinical data classifiers were leveraged. In conclusion, the study provides a reliable framework that is computationally efficient, compatible for resource constraint environment and suitable for timely diagnosis of breast cancer.

Keywords: *Breast Cancer, Transfer Learning, Mammograms, Clinical Data, Machine Learning.*

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

1. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer and a leading cause of cancer-related mortality among women all over the world (Sung et al., 2021). In 2020, the American Cancer Society (ACS) revealed that over 2.3 million new cases of breast cancer were reported worldwide (ACS, 2025). In the same vein, 685,000 deaths were attributed to breast cancer worldwide within the same year. In sub-Saharan Africa, Bray et al. (2022) revealed that 801392 new cancer cases were reported in 2020, with 520158 deaths. In this data, cancer of the breast recorded 129400 among women.

More recently, a study on Global Burden of Disease (GBD) from 1990 to 2023 by Eromonsele (2026) revealed that African countries are recording some of the fastest-growing breast cancer burden. Countries like Equatorial Guinea recorded the highest increase in new breast cancer cases at 312%, Ethiopia reported a 207% increase, Egypt increased by 189%, the Democratic Republic of Congo reported a 160% increase, Mauritania reported a 141% increase, and 108.8% increase for Nigeria, respectively. To help reduce this problem, Huang et al. (2023), Coskun et al. (2023), Carriero et al. (2024), and Ganesan et al. (2026) all agreed that early detection remains critical.

Conventional screening methods include the application of medical image analysis using mammography, ultrasound, and Magnetic Resonance Imaging (MRI). Traditionally, these images required expert radiologist analysis and manual interpretation, which are often inconsistent, prone to error, and time-consuming. In addition, the growing complexity and volume of medical images further underscore the need for automated, reliable, and efficient diagnostic tools (Smrity et al., 2022).

In recent years, Artificial Intelligence (AI), particularly Deep Learning (DL), has emerged as a transformative force in medical diagnostics. Techniques such as Convolutional Neural Networks (CNN) and vision transformers have shown good potential in the classification of different medical images and have recorded superior success rates when compared to expert radiologists (Abdullah et al., 2025). Moreso, models that fuse heterogeneous data derived from radiomic and clinical breast cancer records have shown better promise, as their decision leverages insights from both modalities (Ibrahim et al., 2026; Li et al., 2025). Despite these advancements, several gaps remain unresolved.

Many current models remain unimodal, focusing mainly on either imaging or clinical data, which restricts application diversity across the world. Another challenge is the infrastructure gap, because most of the advanced models are not compatible with current healthcare settings in several underdeveloped countries, where the machines used are outdated, and medical image output is of very poor quality. The risk is that with the low image quality, even advanced transfer learning models stand a chance of making wrong decisions, which then raises research questions.

RQ1: How does integrating clinical data with radiomic MRI features in a deep learning and machine learning model address the limitations of unimodal approaches in breast cancer detection?

RQ2: Which lightweight transfer learning and machine learning models achieved the diagnostic best diagnostic performance for breast cancer classification?

RQ3: Does the multimodal ensemble learning framework improve breast cancer diagnostic robustness and reliability?

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

To answer these research questions, this paper proposes a compatible multimodal ensemble learning framework for breast cancer diagnosis. The paper will experiment with lightweight pre-trained CNN models that are compatible with healthcare infrastructure in low-income countries, train them with MRI data from medical images, and then combine them with a machine learning model trained on a clinical breast cancer dataset. The decisions from the two models will be merged through a voting strategy to facilitate reliable diagnosis of breast cancer.

1.1 The problem statement in a nutshell

Despite the huge advancement in deep learning technology for breast cancer detection, a serious barrier in medium-income and low-income countries is deployment and infrastructure gaps. This is because their healthcare centres and public hospitals, especially in rural localities, rely on outdated imaging equipment, which often produces low-quality MRI and mammograms characterized by artifacts, low resolution, and noise. This poor image quality substantially increases the risk of misdiagnosis, even with advanced deep learning models trained on primarily high-quality datasets. At the same time, hospitals with modern, high-resolution equipment are often unaffordable for poor patients, limiting access to reliable screening. Consequently, unimodal deep learning approaches that depend heavily on imaging data alone become unreliable in such resource-constrained environments. To address this critical gap, the integration of robust clinical data with radiomic MRI features through hybrid deep learning models is essential, as clinical variables can compensate for degraded image quality and improve diagnostic reliability across diverse socioeconomic settings.

1.2 The paper contributions

1) Develop multimodal ensemble learning framework for breast cancer diagnosis through the integration of mammographic image classification with two independent clinical classification models using hard voting strategy.
2) Experiment on lightweight deep learning and machine learning models to identify the optimal classifier for each data modality before multimodal fusion.
3) Develop a multimodal clinical diagnostic software for early detection of breast cancer in medium-low-income African countries.
4) Demonstrate through practical experiments the software reliability

2. REVIEW OF RELATED WORKS

This section reviews related papers that applied deep learning and transfer learning for the diagnosis of breast cancer. Among the studies is Zhang et al. (2023), who proposed a custom fusion model of VGG11 and auto-encoders. The data modality used s diffused tomography and ultrasound. The area under the curve reported 0.931. Alshehri et al. (2022) applied an attention mechanism as a Convolutional Neural Network (CNN) based thermogram classifier and compared it with a standalone CNN. The results reported 99.46% accuracy as against 92.3% reported in the CNN. Mohamed et al. (2022) trained a U-Net-based breast cancer segmentation model using thermograms and recorded 99.3% accuracy. Civiliban et al. (2023) trained a Mask R-CNN-based model with ResNet-50 as the backbone for the segmentation of breast lesions on thermograms. The model, when tested, recorded an average precision score of 0.921 and an overlap segmentation score of 0.868. Sinh et al. (2021) developed an AI-based breast screening device using thermography data. The results reported an AUC of 0.845. While

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

these studies reported high success rates in the classification of breast cancer, it is unsure their reliability in classifying low-quality medical images obtained from outdated machines.

In a different study, Gu et al. (2022) developed a deep learning classifier using a multicentre dataset that contains data of 5000 participants and reported an AUC of 0.913. Zheng et al. (2020) trained a deep learning model for the classification of axillary lymph nodes in patients during the early stages of breast cancer. The results reported an AUC of 0.902. Dan et al. (2024) present a comprehensive literature review on the diagnostic performance of deep learning models for breast cancer detection using ultrasound images. The findings showed that DL is dominating studies in this field of medical image analysis due to the high accuracy it records; however, most of these models may struggle to classify images generated on outdated machines dominant in several low-income country rural health care centres. Secondly, these studies are trained on foreign datasets that do not characterize genetic traits and behavioural patterns of breast cancer in Africa in particular. For instance, Jedy (2012) and Ntekim et al. (2022) revealed that breast cancer in black women is often present at a younger age and more aggressive, compared to white women, leading to bias in trained AI model performance. Therefore, it is important to develop a breast cancer classification model using a localized dataset which model the problem in Africa.

In 2024, Carriero et al. discussed the state-of-the-art advancements in early 2024 for best cancer management. From the thorough review made in the study, considering deep learning and transfer learning via different modalities such as mammography, ultrasound, and MRI. Several challenges were identified, including generalizability issues, multimodal interpretation, privacy, and reliability. Li et al. (2023) combined DL radiomic analysis with clinical characteristics of breast cancer to predict the response of patients to Neoadjuvant chemotherapy. However, while this approach is very good, it is more of a reactive approach. There is need for a proactive solution that facilitates early detection and optimizes the response of patients to breast cancer diagnostic therapy.

The impact of dataset size on model performance was discussed by Abdullah et al. (2025). The analysis revealed that studies trained on small datasets will always perform better compared with those trained on large datasets, thus highlighting the trade-off between dataset sizes. For instance, Hizukuri et al. (2021) reported an AUC of 0.95 when trained on a dataset obtained from 38 patients, while Zhou et al. (2019) reported an AUC of 0.86 after training on a dataset from over 600 patients. Furthermore, Kumari et al. (2023) integrated three deep learning models, which are VGG-16, Xception, and DenseNet-201, for the classification of breast cancer. The experimental findings demonstrate impressive classification accuracies of 99.42% for the IDC dataset and 99.12% for the BreaKHis dataset. Ishtyaq et al. (2023) used several transfer learning models, which are ResNet50, ResNet101, VGG16, and VGG19. Pati et al. (2023) trained several deep learning models, which are ResNet50, InceptionV3, AlexNet, VGG16, and VGG19 were trained and compared them for the classification of breast cancer using mammography images. The model reported an average accuracy of 0.9799, precision of 0.9951, and F1-score of 0.9897. Arooj et al. (2022) trained only AlexNet and reported 99.40% accuracy in classifying breast cancer. Jakkaladiki and Maly (2023) compared ResNet and DenseNet models trained on an augmented breast cancer dataset. Both models reported an average of 98.3% accuracy, while Alruwaili et al. (2022) reported 89.5% accuracy after training ResNet50 as a breast cancer classifier.

Research Gap: Overall, the reviewed literature has shown that deep learning and transfer learning models remain a major player in the development of smart diagnostic systems for early

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

breast cancer detection. The models, after evaluation reported high accuracies and precision averaging over 90%; however, beyond these metrics, there is a need to look into the reliability of the overall system. First, these models are mostly trained with foreign datasets not obtained from patients with basic African genetic traits. Secondly, these models may not be compatible with the hardware and software present in the current infrastructure settings at health care centres in low-income countries. Thirdly, there is a need to fix the potential error that can arise due to poor image quality. This problem will be solved by training clinical records using machine learning and then merging with the trained deep learning techniques through a hard voting ensemble approach. Collectively, this will reinforce decision-making and provide a reliable system for early diagnosis of breast cancer.

3. METHODOLOGY AND SYSTEM DESIGN

The methodology for this paper is quantitative. The system design is made up of two modular streams which are the image diagnostic stream and the clinical diagnostic stream as shown in Figure 1. The process began with three independent data sources which are the medical image dataset comprising of mammographs which contains visual information of breast tissues and structural abnormally. Then two different clinical datasets containing patient's specific information like tumour size, age, hormonal receptor status, etc. These three heterogenous datasets were applied to diagnose breast cancer and counter the limitation associated with unimodal models. The each of the dataset were processed and then split into 80% for training and then 20% for testing. To prevent overfitting and ensure optimization of the model, the training dataset was further divided into training 80% and then validation 20%. The validation was applied during the training process, while the 20% unseen dataset was applied to test the model and evaluate the performance. For the two clinical datasets, four machine learning algorithms were separately trained respectively and then compared to identify the best model. In this case two different best models were produced from each of the clinical datasets.

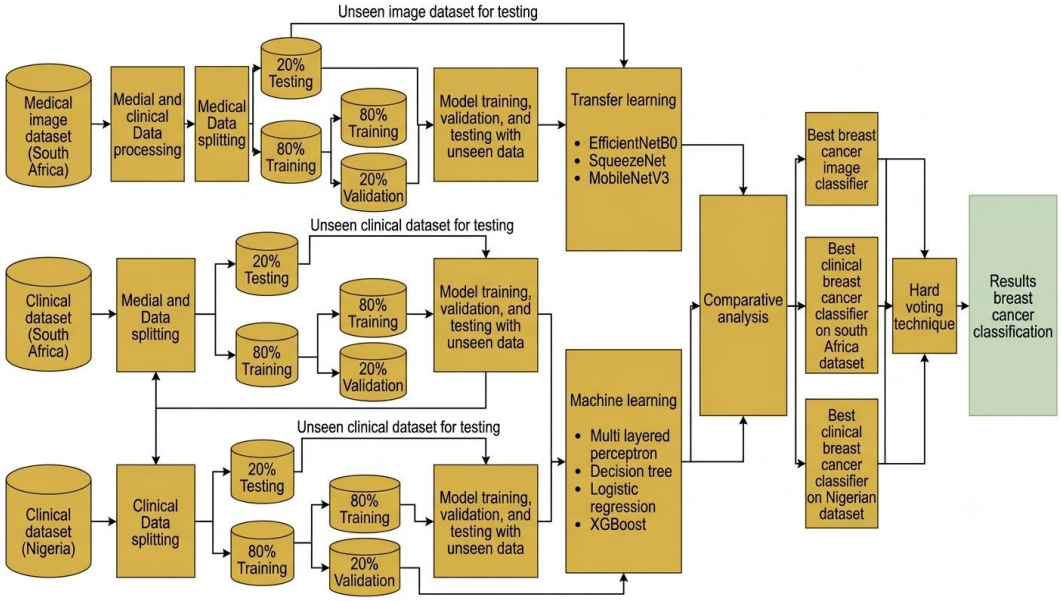


Figure 1: The Proposed Hybrid Deep Learning-Based Breast Cancer Detection System

Consequently, the image-based dataset was applied to train three transfer learning techniques and compared. The best classifier from each dataset was selected and then merged

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

as an ensemble model using hard voting technique with the two independent clinical classifiers. Unlike feature-level fusion, this hard voting approach performs decision-level fusion, where each of the three classifiers independently predicts whether a breast lesion is benign or malignant. Each classifier contributes one vote, and the final diagnosis is determined based on the majority decision.

3.1 Data Collection

Three different secondary datasets were used for this study. Mammogram image dataset was collected from Mendeley (Mattiev et al., 2025). The dataset contained 614 original mammograms sourced from tertiary hospitals at Polokwane, and Limpopo, South Africa. The image size is 1920 * 1080 pixels. The classes are benign with 361 images and malignant with 253 images. The participants are identified patient biopsy records at the hospitals from January 2023 to October 2023. The equivalent clinical health records of the patients in the image dataset were also collected to form the second dataset with 31 attributes as shown in table 1. Finally, the third dataset is the University of Calabar Teaching Hospital (UCTH) breast cancer dataset publicly available at Kaggle repository. The dataset contains 11 attributes shown in table 2. The number of participants is 213 patients admitted in the hospital from January 2019 to August 2021. The sample size is 213 records, span across benign and malignant classes.

Table 1: Clinical data description for the South African Patients

Variable	Data Type	Description
Radius Mean	Continuous	Mean distance from the center of the nucleus to points on its perimeter
Texture Mean	Continuous	Standard deviation of gray-scale pixel intensities
Perimeter Mean	Continuous	Mean perimeter of the cell nucleus
Area Mean	Continuous	Mean area of the cell nucleus
Smoothness Mean	Continuous	Local variation in radius lengths
Compactness Mean	Continuous	Calculated as ($Perimeter^2/Area - 1.0$)
Concavity Mean	Continuous	Severity of concave portions of the nucleus boundary
Concave Points Mean	Continuous	Number of concave portions on the nucleus boundary
Symmetry Mean	Continuous	Degree of symmetry of the nucleus
Fractal Dimension Mean	Continuous	"Coastline approximation" of nucleus boundary complexity
Radius SE	Continuous	Standard error of the radius measurement
Texture SE	Continuous	Standard error of texture measurement
Perimeter SE	Continuous	Standard error of perimeter measurement
Area SE	Continuous	Standard error of area measurement
Smoothness SE	Continuous	Standard error of smoothness
Compactness SE	Continuous	Standard error of compactness
Concavity SE	Continuous	Standard error of concavity
Concave Points SE	Continuous	Standard error of concave points
Symmetry SE	Continuous	Standard error of symmetry
Fractal Dimension SE	Continuous	Standard error of fractal dimension
Radius Worst	Continuous	Largest (worst) radius value
Texture Worst	Continuous	Largest texture value
Perimeter Worst	Continuous	Largest perimeter value
Area Worst	Continuous	Largest area value
Smoothness Worst	Continuous	Largest smoothness value
Compactness Worst	Continuous	Largest compactness value
Concavity Worst	Continuous	Largest concavity value
Concave Points Worst	Continuous	Largest number of concave points
Symmetry Worst	Continuous	Largest symmetry value
Fractal Dimension Worst	Continuous	Largest fractal dimension value
Diagnosis	Categorical	Target class indicating whether the tumour is benign or malignant

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

Table 2: Description of the Nigerian clinical data

Variable	Data Type	Description
S/N	Integer	Unique identifier assigned to each patient record
Year	Integer	Year in which the patient's record was documented
Age	Continuous	Age of the patient at diagnosis
Menopause	Binary	Menopausal status of the patient
Tumor Size	Continuous	Maximum diameter of the detected breast tumour
Inv-Nodes	Integer	Number of axillary lymph nodes invaded by cancer
Breast	Categorical	Breast affected by the tumour
Metastasis	Binary	Indicates whether the cancer has spread to distant organs
Breast Quadrant	Categorical	Anatomical location of the tumour within the breast
History	Family history of breast cancer	Binary
Diagnosis Result	Final clinical diagnosis of the breast lesion	Categorical (Target Variable)

Table 1 presents the data description of the clinical breast cancer patients obtained from South African hospitals, while Table 2 presents the clinical records of breast cancer patients obtained from Nigeria. Figure 2 presents the sample images of benign breast cancer, while Figure 3 presents the sample images of Malignant breast cancer.

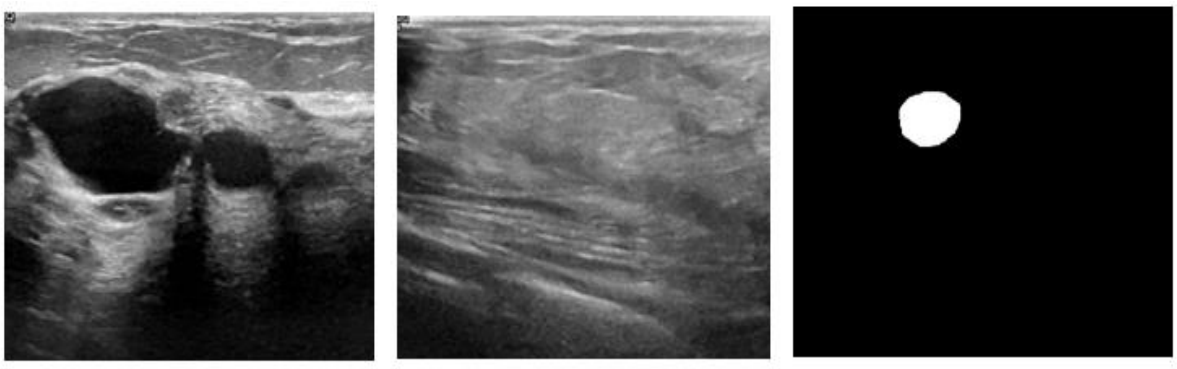


Figure 2: Sample three mammograms of benign breast cancer

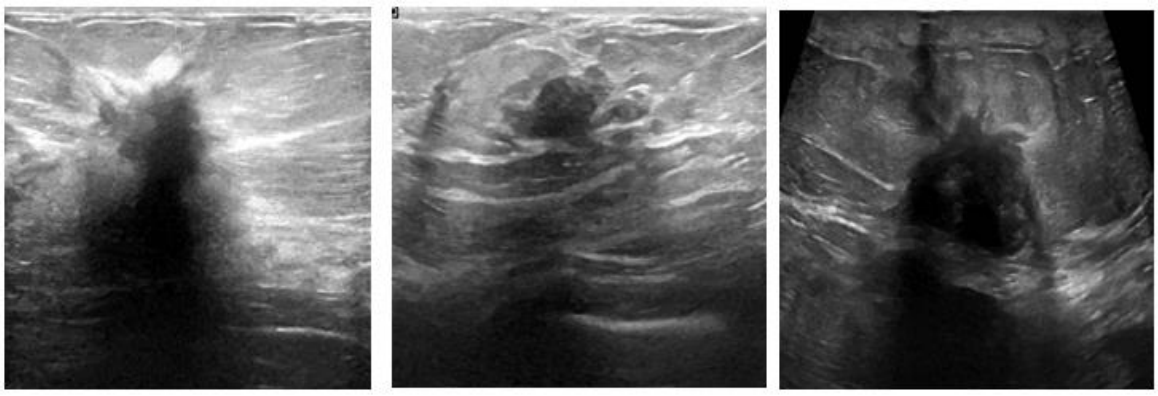


Figure 3: Sample three mammograms of malignant breast cancer

3.2 Deep Learning Models

The deep learning models used for this paper are lightweight models such as MobileNet-V3 small, SqueezeNet, and EfficientNetB0. SqueezeNet is a lightweight CNN designed with fewer CNN parameters, while maintaining high performance accuracy. This model is

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

specifically designed for low computation capacity applications and hence makes it suitable for the current infrastructural setting in rural healthcare centres (Shynybay et al., 2026). MobileNetV3 small combines depthwise separable convolutions, squeeze-and-excitation attention mechanisms, and optimized activation functions to achieve high classification accuracy with significantly lower computational complexity (Yalim et al., 2026). EfficientNetB0 is a type of CNN that uses a compound coefficient scaling strategy to optimize the width, depth, and input resolution of the network (Tan and Lee, 2020).

3.3 The Machine Learning Model

The machine learning models for this study are Multi-Layered Perceptron (MLP) (Hong et al., 2020), logistic regression (LR) (Tang et al., 2021), extreme gradient boosting (XGB), and decision tree (DT) (Tang et al., 2021). The MLP is a type of neural network that models complex clinical variables relationships through the interconnection of neurons structured in multiple hidden layers (128-64-32 neurons).

The activation function used is rectified linear unit. DT recursively parition clinical features space into increasingly heterogeneous subsets based on feature importance, thereby producing easily interpretable decision rules. Logistic Regression estimates the probability of malignancy using a logistic function and serves as an effective statistical baseline for binary classification.

XGBoost constructs an ensemble of sequential decision trees in which each new tree corrects the prediction errors of previous trees through gradient boosting, resulting in highly accurate nonlinear classification (Yavuz et al., 2023). Evaluating multiple algorithms allows the framework to identify the classifier that best captures the relationships among the clinical variables.

3.4 Training of the Machine Learning and Deep Learning Models

The multi-model breast cancer classification system was developed through the independent training of the three transfer learning techniques (MobileNetV3, SqueezeNet, EfficientNetB0) on the image-based breast cancer dataset and four machine learning techniques on the clinical breast cancer dataset.

The tools used are Python 3.11, TensorFlow, Scikit-learn, XGBoost, NumPy, OpenCV, and Pandas. For the transfer learning models, first the images were resized to 224*224 pixels, then normalized to [0,1], and augmented to balance the classes through random horizontal flipping at $\pm$15°, rotation of $\pm$15°, and zooming (0.2).

The pre-trained models were initialized with ImageNet-trained weights and fine-tuned using the processed breast cancer image dataset. Adam optimizer was used to train the networks at a 0.0001 learning rate, a 32-batch size, and 50 epochs, while the cross-entropy measured the loss. Figure 4 presents the flow chart of the image breast cancer classifier.

The two clinical datasets were also trained using MLP, DT, and LR. The MLP was trained with Adam optimization, while the LR, DT, and XGBoost were configured using the optimal parameters obtained as they were trained. All the experiments were carried out with an Intel Core i9 processor, 64 GB RAM, and NVIDIA Tesla GPU in Google Colab. Figure 5 presents the flow chart of the clinical data breast cancer classifier.

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

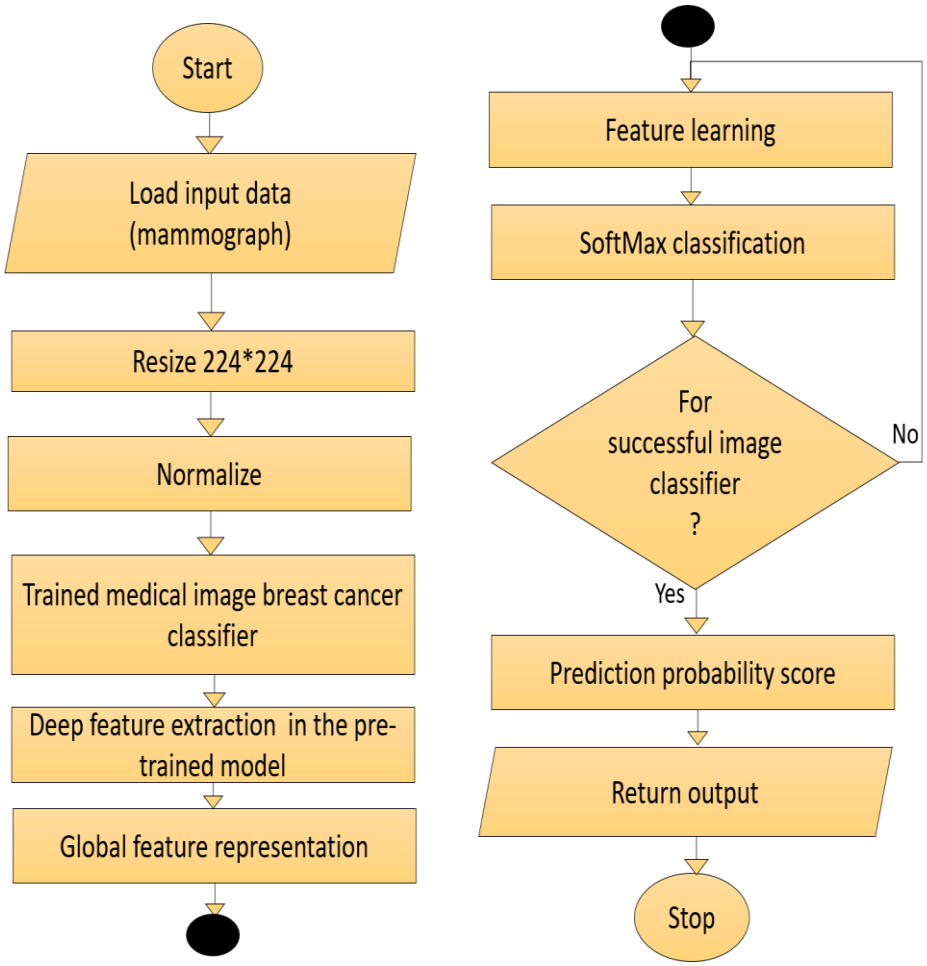


Figure 4: Flow Chart of the Image-Based Breast Cancer Classifier

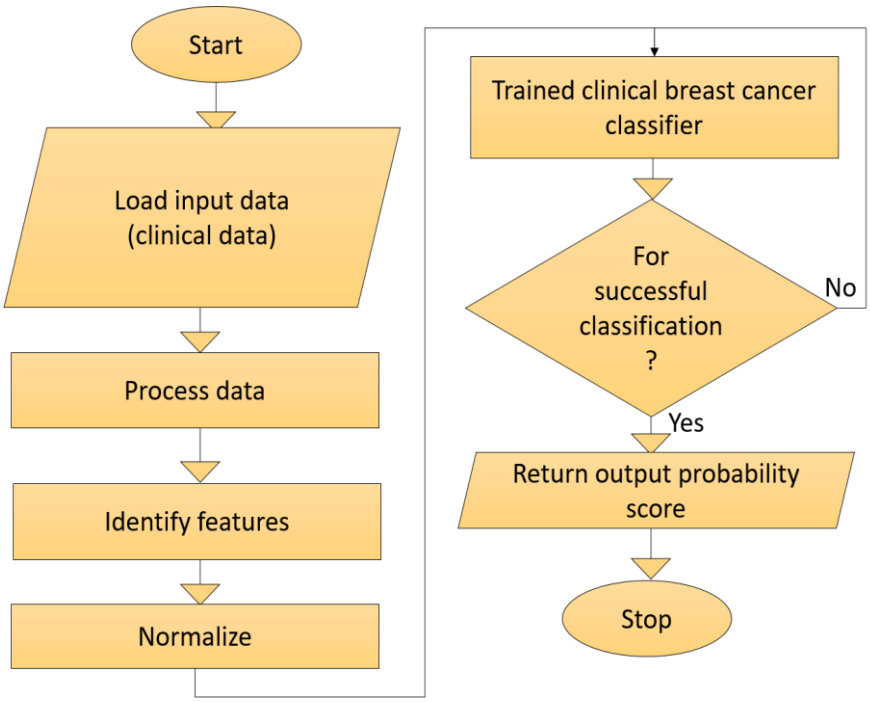


Figure 5: Flow Chart of the Clinical Record Classifier for Breast Cancer

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

Figure 4 shows the transfer learning-based breast cancer classification flow chart. Initially, the mammogram is loaded, automatically resized, normalized, and then fed to the trained model, which then extracts deep features, represents them globally, and learns to make classification using the SoftMax function.

Successful classification is reported with the probability score at the output. Figure 5 begins with the initial input of the clinical records collected from patients. The information is processed through feature identification and normalization, then the trained clinical record classifier decides the class of breast cancer and outputs the probability score.

3.6 Mathematical Modelling of the Hard Voting ensemble

Hard voting is a type of ensemble model that is used to make the final decision to classify breast cancer using input from clinical and medical images.

Let the mammogram image of the breast be represented asX_i, the first clinical dataset be represented as $X_{c1}$and the second clinical dataset be X_{c2}. the three classifiers are therefore presented in equation 6 as;

$$\begin{bmatrix} f_1(X_i) \\ f_2(X_{c1}) \\ f_3(X_{c2}) \end{bmatrix} \quad (6)$$

Each classifier predicts either malignant (1) or benign class (0) of breast cancer, while $\hat{y}$ is the predictor. Thus, the three respective predictors are reported in equation 7-9.

$$\hat{y}_1 = f_1(X_i) \quad (7)$$

$$\hat{y}_2 = f_2(X_{c1}) \quad (8)$$

$$\hat{y}_3 = f_3(X_{c2}) \quad (9)$$

The total vote by each classifier is represented as equation (10) for benign and then equation (11) for malignant.

$$N_0 = \sum_{i=1}^{3} \delta\ (\hat{y}_i = 0) \quad (10)$$

$$N_1 = \sum_{i=1}^{3} \delta\ (\hat{y}_i = 1) \quad (11)$$

The ensemble classifier predicts the class with the most votes. The decision function, which is the hard voting, is represented in equation 12.

$$\hat{y} = \arg \max_{k \in \{0,1\}} \sum_{i=1}^{3} \delta\ (\hat{y}_i = k) \quad (12)$$

The equivalent binary decision rule considering the three classifiers is presented as equation (13).

$$\hat{y} = \begin{cases} 1 & \hat{y}_2 + \hat{y}_2 + \hat{y}_3 \geq 2 \\ 0 & \hat{y}_2 + \hat{y}_2 + \hat{y}_3 < 2 \end{cases} \quad (13)$$

Equation 13 states that if at least two classifiers predict malignant, then the final diagnosis is malignant; otherwise, the diagnosis is benign. Figure 6 presents the block diagram of the multi-model ensemble framework for breast cancer classification.

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

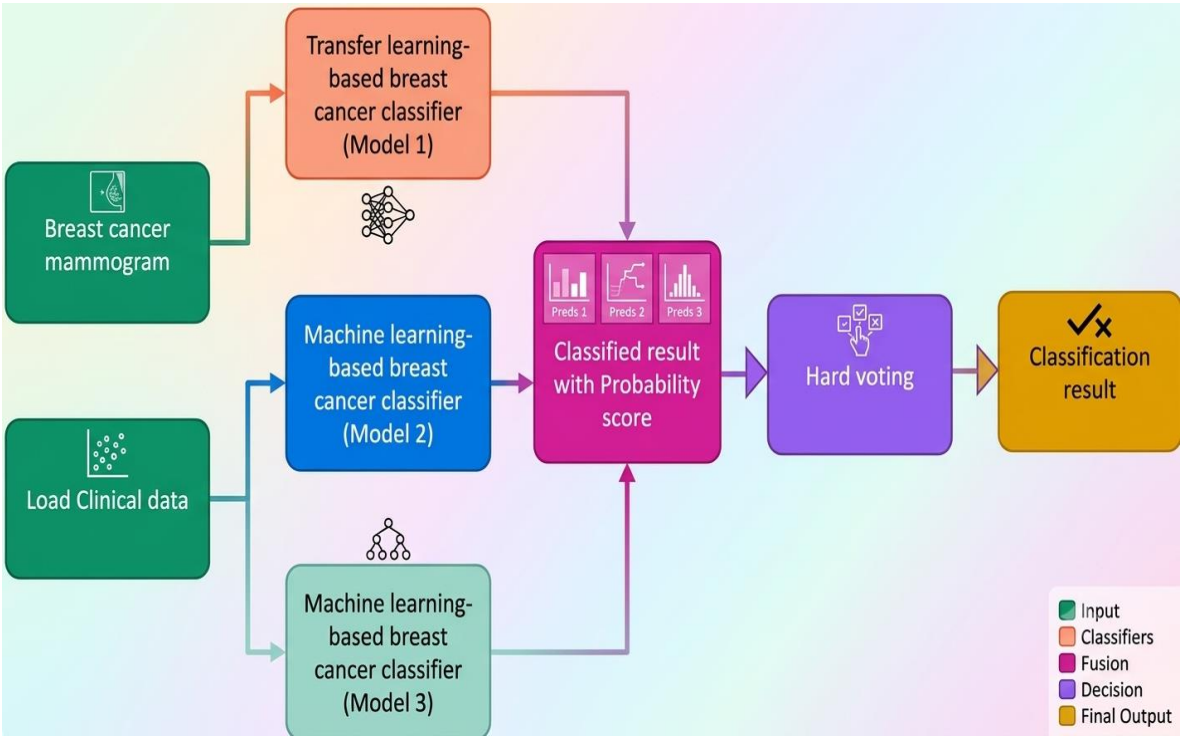


Figure 6: Block diagram of multi-model ensemble framework for breast cancer classification

Figure 6 presents the multi-model ensemble framework for the classification of breast cancer. The model utilized two sets of input, which are the mammograms and the clinical dataset. For the mammograms, transfer learning was used as the classifier. The clinical records are classified by two trained independent classifiers.

The classified results of the three models are processed through a hard voting technique to determine correctly if the patient has malignant or benign. The models were trained and validated with 80% of the original dataset, while 20% of unseen data were used to test the model and confirm performance.

4. RESULTS AND DISCUSSION

This section presents the training, validation, and testing results of the machine learning and transfer learning classifiers using the three different datasets.

4.1 Results of Machine Learning Models on the South African (SA) Clinical Dataset

The results of the machine learning models' training and validation on the SA clinical dataset are reported in Table 3.

Table 3: Validation Results of Machine Learning Models on the South African Clinical Dataset

Model	Accuracy	Precision	Recall	F1-Score
MLP	0.9495	0.9286	0.9512	0.9398
Logistic Regression	0.9899	1.0000	0.9756	0.9877
XGBoost	0.9596	0.9302	0.9756	0.9524
Decision Tree	0.9091	0.8810	0.9024	0.8916

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

Table 3 presents the different results of the machine learning models trained on the clinical breast cancer dataset obtained from SA. The results showed that logistic regression recorded the best validation accuracy of 98.99%, which is very good. The same model also reported the best F1-score of 98.77%. Although the XGBoost and MLP compete among the best, the DT reported the least validation performance.

The result implied that the LR is the best in our model for the classification of clinical records from the SA dataset. To confirm these results, unseen clinical records from the SA hospital were applied to test the models. The results obtained are reported in Table 3.

Table 4: Testing Results of Machine Learning Models on the SA Clinical Dataset

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
MLP	0.9756	1.0000	0.9412	0.9697	0.9975
Logistic Regression	0.9675	0.9796	0.9412	0.9600	0.9965
XGBoost	0.9675	1.0000	0.9216	0.9592	0.9937
Decision Tree	0.9350	0.9388	0.9020	0.9200	0.9301

Table 4 compares the results of the models when tested on unseen clinical records of breast cancer patients from SA. From the results, it was observed that the MLP reported the best overall generalization performance, recording 97.56% accuracy, a precision score of 94.12, an F1-score of 96.97%, and ROC-AUC score of 99.75%, while the LR and XGBoost reported very close competing results. Therefore, the MLP was selected as the optimal classifier for the SA clinical dataset and integrated into the multimodal hard voting ensemble framework.

4.2 Results of Machine Learning Models on the Nigerian (NG) Clinical Dataset

This section presents the results of the machine learning classifiers trained on the Nigerian clinical breast cancer dataset. Table 5 presents the training and validation results with the seen dataset.

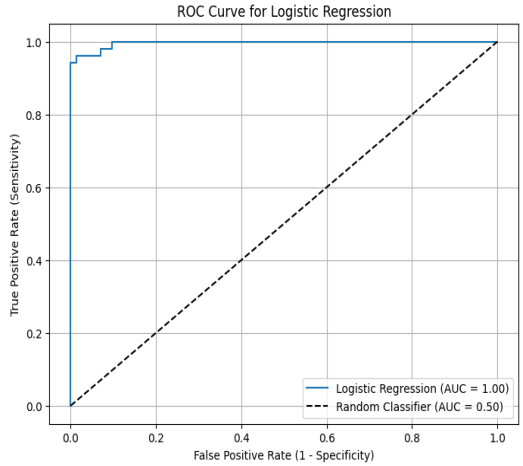
Table 5: Validation Results of Machine Learning Models on the Nigerian Clinical Dataset

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.9706	0.9375	1.0000	0.9677
Decision Tree	0.9412	0.8824	1.0000	0.9375
MLP	0.9706	0.9375	1.0000	0.9677
XGBoost	0.9118	0.8333	1.0000	0.9091

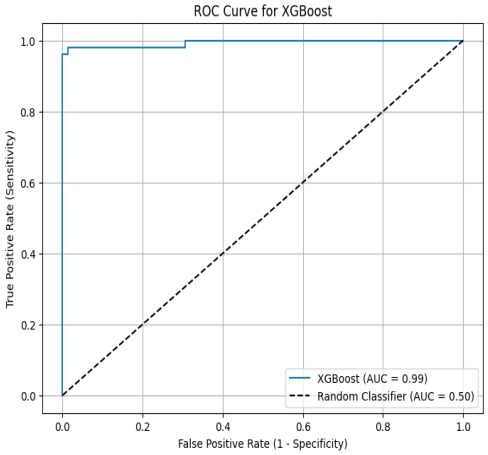
Table 5 presents the training and validation results of the NG clinical dataset on machine learning models. The results showed that the LR and MLP compete as the two best classifiers with 97.06% accuracy, a precision score of 93.75%, recall of 100%, and F1-score of 96.77%. The DT and XGBoost, on the other hand, produced comparatively lower performance, with XGBoost reporting the lowest validation accuracy score of 91.18%.

To get more clarity on the best model, an unseen dataset from NG Hospital containing clinical records of patients with breast cancer was used to test the models. The results of the ROC are reported in Figure 7, while the results of other test metrics like accuracy, precision, recall, and F1-score are all summarized in Table 6.

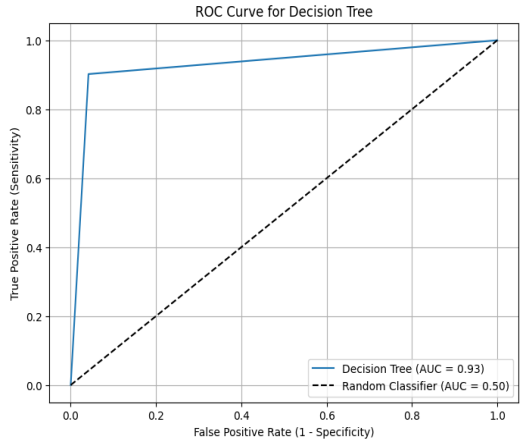
Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891



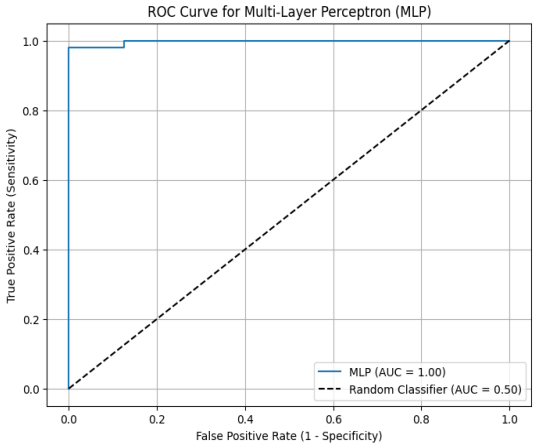
(a)ROC for the LR testing



(b) ROC for the XGBoost testing



(c)ROC for the DT testing



(d) ROC for the MLP testing

Figure 7: Test ROC of the Trained Machine Learning Clinical Breast Cancer Classification (All the ROC Results Are Reported In Table 6 and Discussed)

Table 6: Testing Performance of Machine Learning Models on the Nigerian Clinical Dataset

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.9302	1.0000	0.8421	0.9143	0.9693
Decision Tree	0.8372	0.8000	0.8421	0.8205	0.8377
MLP	0.9302	1.0000	0.8421	0.9143	0.9496
XGBoost	0.8140	0.7391	0.8947	0.8095	0.9364

Table 6 presents the testing results of the models on an unseen dataset. From the results, it was observed that again the MLP and LR reported the best results with an accuracy of 93.02%, a precision score of 100%, and F1-score of 91.43%. However, Logistic Regression recorded the highest ROC-AUC (96.93%) compared with the MLP (94.96%), demonstrating better discrimination between benign and malignant breast cancer cases on previously unseen data. The results also showed that although the XGBoost reported the best recall score of 89.47%, the lower accuracy and precision when compared to the other models indicate a higher false positive rate. To this end, the LR was selected as the optimal model for the development of the multimodal hard voting ensemble framework.

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

4.3 Results Of Transfer Learning Models On The Mammogram Dataset

This section presents the results of the transfer learning-based classifiers trained with the mammogram datasets. The models were trained and validated on the two classes, which are benign and malignant. The results were reported in Table 7.

Table 7: Testing Results of Mobilenetv3 on the Breast Cancer Image Dataset

Class	Precision	Recall	F1-Score	Support
Benign (B)	0.83	1.00	0.91	61
Malignant (M)	1.00	0.80	0.89	62
Accuracy			0.90	123
Macro Average	0.92	0.90	0.90	123
Weighted Average	0.92	0.90	0.90	123

Table 7 presents the MobileNetV3 results after training and validation. From the results, it was observed that the MobileNetV3 recorded an average accuracy of 90% when classifying benign and malignant breast cancer. The precision recorded 92% average score, recall was 90%, and f1-sore 90%. These results showed that the trained MobileNetV3 reported a good classification success rate in correctly distinguishing between malignant and benign breast cancer mammograms. In Table 8, the results of the EfficientNetB0 training were reported.

Table 8: Testing Performance of Efficientnet-B0 on the Breast Cancer Image Dataset

Class	Precision	Recall	F1-Score	Support
Benign (B)	0.71	1.00	0.83	61
Malignant (M)	1.00	0.60	0.75	62
Accuracy			0.80	123
Macro Average	0.86	0.80	0.79	123
Weighted Average	0.86	0.80	0.79	123

Table 8 presents the result of the EfficientNetB0 trained as the breast cancer classifier. From the results, it was observed that when 123 images were used to test the model, 80% accuracy was reported. The average F1-score is 79%, recall is 80%, and precision is 86%. Table 9 also presents the results of the SqueezeNet training and validation. From the results, it was observed generally that the SqueezeNet model performed average with 54% accuracy. The average precision score is low at 0.30, the precision score is 0.54, f1score of 0.38.

Table 9: Testing Performance of Squeezenet on the Breast Cancer Image Dataset

Class	Precision	Recall	F1-Score	Support
Benign (B)	0.54	1.00	0.71	61
Malignant (M)	0.00	0.00	0.00	62
Accuracy			0.54	123
Macro Average	0.27	0.50	0.35	123
Weighted Average	0.30	0.54	0.38	123

In general, MobileNetV3 in Table 7 achieved the best overall performance, recording an accuracy of 90.0%, weighted precision of 92.0%, weighted recall of 90.0%, and weighted F1-score of 90.0%. It also maintained balanced performance across both benign and malignant classes, correctly identifying all benign cases and 80.0% of malignant cases. EfficientNet-B0 in Table 8 ranked second with an accuracy of 80.0%, demonstrating excellent benign detection but reduced sensitivity to malignant lesions, with a malignant recall of 60.0%. In contrast, SqueezeNet in Table 9 produced the lowest performance, achieving only 54.0% accuracy and failing to correctly classify any malignant cases, resulting in a malignant recall of 0.0%. Based

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

on its superior accuracy, balanced class performance, and highest weighted evaluation metrics, MobileNetV3 was selected as the optimal transfer learning model for the image classification branch of the proposed multimodal hard voting ensemble framework.

4.4 Results of the Transfer Learning Model Testing With Unseen Mammograms

This section utilized 10 random mammogram images of benign and malignant to evaluate the model performance. The results are reported considering the sample ID of the image from the dataset, the actual class the image belongs to, the predicted class by the classifier, and the probability of Benign prediction, the probability of malignant prediction, and the overall confidence score. Table 10-12 presents the independent testing results of the transfer learning models. The result for the MobileNetV3 was reported in Table 10. The result for the EfficientNetB0 was reported in Table 11. The result for the SqueezeNet was reported in Table 12.

Table 10: Testing Result of the Mobilenetv3 with Unseen Mammogram

Sample ID	Actual Class	Predicted Class	Benign Probability	Malignant Probability	Confidence (%)
Benign (101)	Benign	Benign	0.987	0.013	98.7
Malignant (101)	Malignant	Malignant	0.045	0.979	97.9
Benign (102)	Benign	Benign	0.654	0.039	96.1
Malignant (102)	Malignant	Malignant	0.083	0.917	91.7
Benign (103)	Benign	Benign	0.942	0.058	94.2
Malignant (103)	Malignant	**Benign**	0.624	0.376	62.4
Malignant (104)	Malignant	Malignant	0.087	0.817	91.6
Benign (104)	Benign	Benign	0.973	0.027	97.3
Malignant (105)	Malignant	Malignant	0.094	0.906	90.6
Benign (105)	Benign	Benign	0.981	0.019	98.1

Table 11: Testing result of the EfficientNetB0 with unseen mammogram

Sample ID	Actual Class	Predicted Class	Benign Probability	Malignant Probability	Confidence (%)
Benign (101)	Benign	Benign	0.946	0.054	94.6
Malignant (101)	Malignant	Malignant	0.124	0.876	87.6
Benign (102)	Benign	Benign	0.918	0.082	91.8
Malignant (102)	Malignant	**Benign**	0.673	0.327	67.3
Benign (103)	Benign	Benign	0.893	0.107	89.3
Malignant (103)	Malignant	Malignant	0.211	0.789	78.9
Malignant (104)	Malignant	**Benign**	0.641	0.359	64.1
Benign (104)	Benign	**Malignant**	0.557	0.043	61.3
Malignant (105)	Malignant	Malignant	0.182	0.818	81.8
Benign (105)	Benign	Benign	0.934	0.066	93.4

Table 12: Testing Result of the Squeezenet with Unseen Mammogram

Sample ID	Actual Class	Predicted Class	Benign Probability	Malignant Probability	Confidence (%)
Benign (101)	Benign	Benign	0.832	0.168	98.7
Malignant (101)	Malignant	**Benign**	0.731	0.269	77.9
Benign (102)	Benign	Benign	0.791	0.209	96.1
Malignant (102)	Malignant	Benign	0.624	0.376	62.4
Benign (103)	Benign	Benign	0.744	0.256	94.2
Malignant (103)	Malignant	**Benign**	0.683	0.317	62.4
Malignant (104)	Malignant	**Benign**	0.604	0.396	94.3
Benign (104)	Benign	Malignant	0.557	0.043	61.3
Malignant (105)	Malignant	**Benign**	0.671	0.219	59.4
Benign (105)	Benign	Benign	0.813	0.187	98.1

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

Table 10-12 presents the independent testing results of the transfer learning models. For MobileNetv3, it was observed that out of the five malignant images used to test the model, four images were correctly classified as malignant, while one was misclassified. The result also showed that all the benign images used to test the model were correctly classified. These results mean that the model was able to correctly classify malignant and benign with a high success rate; however, it poses a 10% risk of misclassification for malignant. Table 11 evaluated the performance of the EfficientNetB0 using mammograms of malignant and benign classes. From the results, it was observed that the model misclassified five malignant images out of the five applied to test the model, and then misclassified a benign image out of the five. Overall, it was observed that out of the ten images used to test the model, seven were classified correctly and three images were misclassified.

Table 12 also presents the SqueezeNet. From the results, it was observed that the model performed on average. The results reported four misclassifications out of the ten images used to test the model. Overall, the comparative analysis demonstrates a clear performance hierarchy among the evaluated transfer learning models, with MobileNetV3 consistently outperforming EfficientNet-B0 and SqueezeNet in terms of prediction accuracy, confidence, and generalization on previously unseen data. The superior performance of MobileNetV3 can be attributed to its ability to achieve an effective balance between model complexity and feature extraction efficiency, allowing it to capture discriminative mammographic patterns while minimizing overfitting.

These findings are consistent with the overall testing results, where MobileNetV3 achieved the highest accuracy and F1-score among all evaluated image classifiers. Based on its superior diagnostic capability and consistent performance across both seen and unseen datasets, MobileNetV3 was selected as the optimal image classification model and integrated with the best-performing clinical classifiers through the proposed hard voting ensemble framework.

4.4 Multi-model ensemble framework for breast cancer classification

This section presents the results of the multi-model ensemble framework for breast cancer classification using mammograms and clinical records of verified patients from Nigerian hospitals. Figure 8 presents the result of the practical test experiment, while figure 4.19 showed the second experimental result.

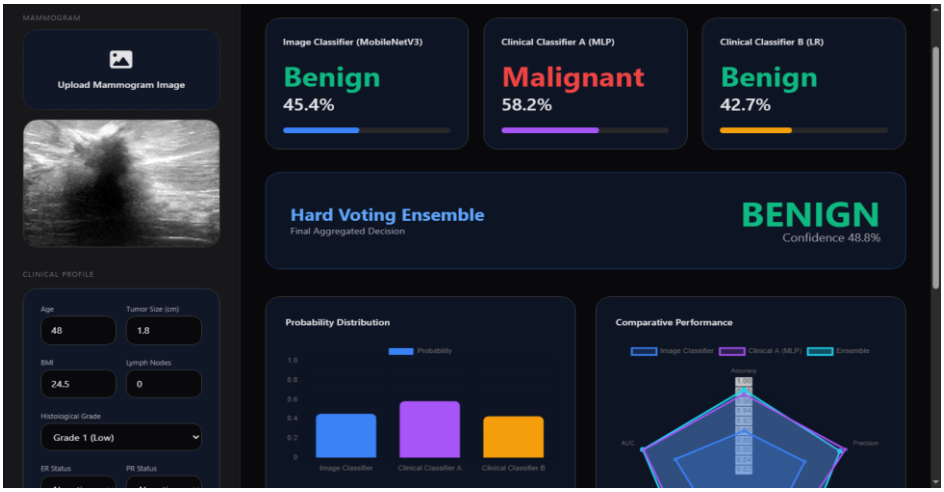


Figure 8: Experimental Test 1

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

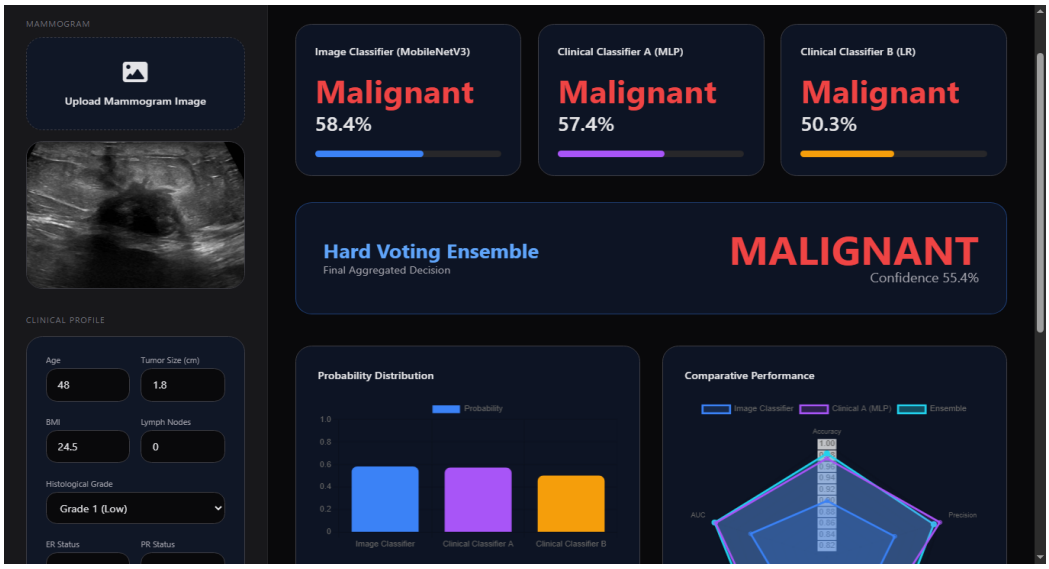


Figure 9: Experimental Test 2

The result in figure 8 showed that the transfer learning model classified the image as benign with a 45.4% confidence score; the MLP-based classifier classified the image as malignant with 58.2% confidence; however, the LR classifier classified the image as benign with 42.7 confidence. The three classifiers were analysed using the hard voting ensemble techniques to make the final decision and classify the output as benign with a 48.4% success rate.

In Figure 9, the result revealed that the three models agreed on the same classified result. The transfer learning model reported 58.4% malignant classification confidence; 57.4% confidence was recorded for the MLP classifier and 50.3% confidence for the LR classifier. The final result produced by the hard voting is 55.4%. figure 10 showed the result of experiment 3, while figure 11 is the final experiment.

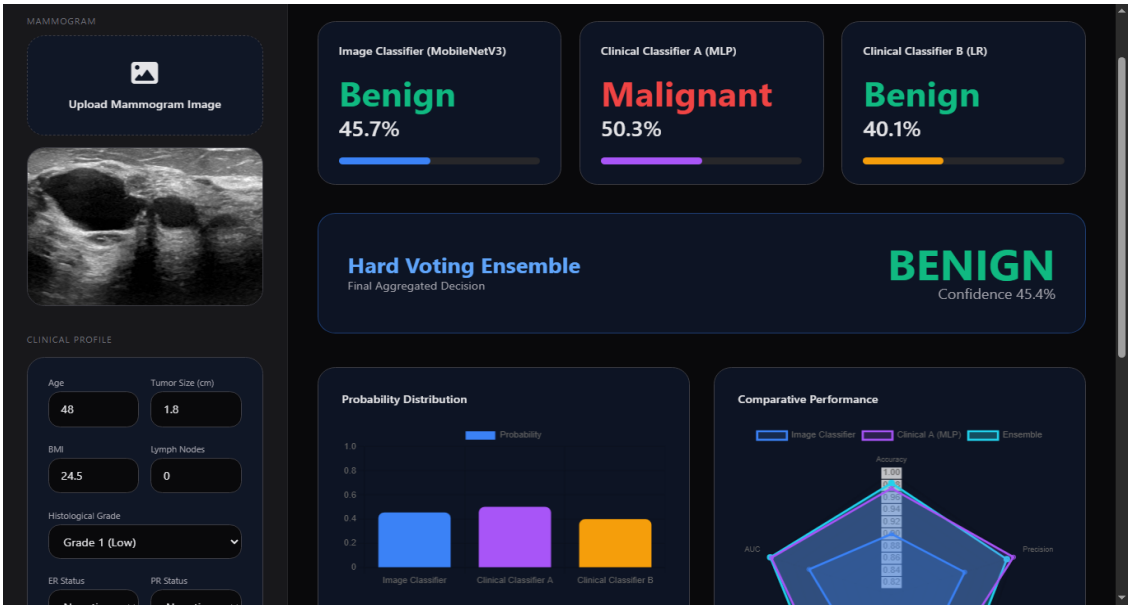


Figure 10: Experimental Test 3

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

Figure 11: Experimental Test 4

In Figure 10, the experiment showed that the transfer learning model classified benign with 45.7%, while the classification model with LR also classified benign with 40.1%; however, the MLP classifier reported malignant as the classified class with 50.3%. The final voting results classified benign as the patient condition with 45.45% confidence. Finally, in the experiment reported in Figure 10, the transfer learning model and the LR classified the image as benign with 54.7% confidence and 10.1% confidence, respectively, while the MLP classified it as malignant with 50.3% confidence. The final voting of the ensemble model reported a 50.9% confidence classification of malignant as the final result.

In general, this paper has demonstrated that the application of a multi-model classifier counters the limitation of single models for breast cancer diagnosis. The study has also shown that integration of clinical and image classifiers provides a reliable solution to potential errors that might occur in the future when classifying the risk of breast cancer in patients.

5. CONCLUSION

This paper presents a multimodal ensemble learning framework for breast cancer diagnosis through the integration of clinical records and mammographic images, to improve reliability of clinical diagnosis. Three transfer learning models which are MobileNetV3, EfficientNet-B0, and SqueezeNet were evaluated for image classification, while LR, MLP, XGBoost, and DT were trained and evaluated for clinical breast cancer classification using two independent datasets from NG and SA. For the SA clinical dataset, comparative analysis identified MLP as the best classifier while LR was identified as the best for Nigerian clinical dataset. For the transfer learning models, MobileNetV3 was identified as the best. These three models were integrated through a hard voting strategy for final breast cancer diagnosis. The experimental results demonstrated that the proposed multimodal framework effectively leveraged the complementary strengths of imaging and clinical information, resulting in improved diagnostic performance and enhanced robustness compared to individual classifiers. The integration of heterogeneous data sources reduced the likelihood of misclassification and produced more reliable predictions, particularly for challenging breast cancer cases. Furthermore, the framework offers a practical and computationally efficient approach for

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

intelligent clinical decision support, making it suitable for deployment in resource-constrained healthcare environments. Overall, the proposed multimodal ensemble learning framework represents a promising solution for accurate breast cancer diagnosis and has the potential to support clinicians in early detection, timely intervention, and improved patient outcomes.

5.1 Limitation and future directions

Despite the promising performance achieved by the multimodal ensemble learning framework, several limitations were acknowledged. First, the Nigerian dataset used for this work is very small and may affect decision of the of the LR. Expanding the dataset with more diverse breast cancer cases could further improve the generalization capability of the classifier when diagnosing Nigerian patients. Another weakness is that the image dataset is only mammograms. Future studies may also incorporate additional imaging modalities, including ultrasound and magnetic resonance imaging (MRI), together with genomic, histopathological, and biomarker data to develop a more comprehensive breast cancer diagnosis system.

5.2 Data availability statement

The Nigerian data can be obtained from
https://www.kaggle.com/datasets/anoushkaduggal/ucth-breast-cancer-dataset

The South African data source is cited in the body of the work.

5.3 AI declaration statement

Pro version of Grammarly tool was used to fix typographical error and also applied for grammar check.

Reference

1) Abdullah, K. A., et al. (2025). Deep learning-based breast cancer diagnosis in breast MRI. Journal Name, Volume (Issue), pages. (PMC12226709)

2) Alruwaili M, Gouda W. (2022). Automated breast cancer detection models based on transfer learning. Sensors. 22(3):876.

3) Alshehri, A.; AlSaeed, D. (2022). Breast Cancer Detection in Thermography Using Convolutional Neural Networks (CNNs) with Deep Attention Mechanisms. Appl. Sci. 12, 12922

4) Arooj S, Zubair M, Khan MF, Alissa K, Khan MA, Mosavi A. (2022). Breast cancer detection and classification empowered with transfer learning. Front Public Health. 10:924432.

5) Bray F, Parkin DM. (20220. African Cancer Registry Network. Cancer in sub-Saharan Africa in 2020: a review of current estimates of the national burden, data gaps, and future needs. Lancet Oncol. 2022 Jun; 23(6):719-728. doi: 10.1016/S1470-2045(22)00270-4. PMID: 35550275.

6) Carriero, A.; Groenhoff, L.; Vologina, E.; Basile, P.; Albera, M. (2024). Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. Diagnostics 14, 848. https:// doi.org/10.3390/diagnostics1408084

7) Civilibal, S.; Cevik, K.K.; Bozkurt, A. (2023). A Deep Learning Approach for Automatic Detection, Segmentation and Classification of Breast Lesions from Thermal Images. Expert Syst. Appl. 212, 118774

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

8) Coşkun D, Karaboğa D, Baştürk D, Akay A, Nalbantoğlu B, Doğan ÖU, Paçal S, İ. And, Karagöz MA. (2023). A comparative study of YOLO models and a transformer-based YOLOv5 model for mass detection in mammograms. Turkish J Electr Eng Comput Sci. 31(7):1294–313.

9) Dan, Q.; Xu, Z.; Burrows, H.; Bissram, J.; Stringer, J.S.A.; Li, Y. Diagnostic Performance of Deep Learning in Ultrasound Diagnosis of Breast Cancer: A Systematic Review. Npj Precis. Oncol. 2024, 8, 21

10) Eromonsele, F. (2026). *Nigeria records 543% surge in breast cancer cases as Africa tops global rise – Study*. Premium Times. https://www.premiumtimesng.com/health/health-news/861130-nigeria-records-543-surge-in-breast-cancer-cases-as-africa-tops-global-rise-study.html

11) Ganesan, J., Krishnan, V., Rathinavel, T. *et al.* (2026). Enhancing breast cancer detection accuracy through machine learning, deep learning and transfer learning techniques for clinical practice. *Discov Artif Intell* **6**, 29 https://doi.org/10.1007/s44163-025-00649-3

12) Gu, Y.; Xu, W.; Lin, B.; An, X.; Tian, J.; Ran, H.; Ren, W.; Chang, C.; Yuan, J.; Kang, C.; et al. (2022). Deep Learning Based on Ultrasound Images Assists Breast Lesion Diagnosis in China: A Multicenter Diagnostic Study. Insights Imaging 13, 124.

13) Hizukuri A, Nakayama R, Nara M, Suzuki M, Namba K (2021). Computer-aided diagnosis scheme for distinguishing between benign and malignant masses on breast DCE-MRI images using deep convolutional neural network with Bayesian optimization. J Digit Imaging 34:116–123. https:// doi.org/10.1007/s10278-020-00394-2 34.

14) Hong, H.; Tsangaratos, P.; Ilia, I.; Loupasakis, C.; Wang, Y. (2020). Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. *Sci. Total Environ*, *742*, 140549

15) Huang, P. W., Ouyang, H., Hsu, B. Y., Chang, Y. R., Lin, Y. C., Chen, Y. A., … Pai,T. W. (2023). Deep-learning based breast cancer detection for cross-staining histopathology images. Heliyon, 9(2).

16) Ibrahim, A. M., et al. (2026). A multi-modal deep learning framework for enhanced breast cancer diagnosis using mammograms and clinical data. Scientific Reports, 16, Article 17175. https://doi.org/10.1038/s41598-026-48085-2

17) Ishtyaq Mahmud M, Mamun M, Abdelgawad A. (2023). A Deep Analysis of Transfer Learning Based Breast Cancer Detection Using Histopathology Images. *arXiv e-prints*, arXiv-2304.

18) Jakkaladiki SP, Maly F. (2023). An efficient transfer learning based cross model classification (TLBCM) technique for the prediction of breast cancer. PeerJ Comput Sci. 9:e1281.

19) Jedy-Agba, E., Curado, M. P., Ogunbiyi, O., Oga, E., Fabowale, T., Igbinoba, F., Osubor, G., Otu, T., Kumai, H., Koechlin, A., Osinubi, P., Dakum, P., Blattner, W., & Adebamowo, C. A. (2012). Cancer incidence in Nigeria: A report from population-based cancer registries. *Cancer Epidemiology, 36*(5), e271–e278. https://doi.org/10.1016/j.canep.2012.04.007

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

20) Kumari V, Ghosh R. (20230. A magnification-independent method for breast cancer classification using transfer learning. Healthcare Analytics;. p. 100207.

21) Li, T., et al. (2025). Deep learning in multi-modal breast cancer data fusion: A literature review. Quantitative Imaging in Medicine and Surgery, 15(11), 11578–11610. https://doi.org/10.21037/qims-2024-2903

22) Li, Y.; Fan, Y.; Xu, D.; Li, Y.; Zhong, Z.; Pan, H.; Huang, B.; Xie, X.; Yang, Y.; Liu, B. (2023). Deep Learning Radiomic Analysis of DCE-MRI Combined with Clinical Characteristics Predicts Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer. Front. Oncol. 12, 1041142

23) Mattiev, Jamolbek; Baloyi, Vekani; Mokwena, Sello (2025), "Mammogram Image Dataset for Breast Cancer Detection", Mendeley Data, V1, doi: 10.17632/88vzgys5vg.1

24) Mohamed, E.A.; Rashed, E.A.; Gaber, T.; Karam, O. (2022). Deep Learning Model for Fully Automated Breast Cancer Detection System from Thermograms. PLoS ONE,17, e0262349

25) Ntekim, A., Oluwasanu, M., & Odukoya, O. (2022). Breast cancer in adolescents and young adults less than 40 years of age in Nigeria: A retrospective analysis. *International Journal of Breast Cancer,* Article 9943247. https://doi.org/10.1155/2022/9943247

26) Pati, A., Parhi, M., Pattanayak, B. K., Singh, D., Singh, V., Kadry, S., Kang, B.G. (2023). Breast Cancer Diagnosis Based on IoT and Deep Transfer Learning Enabled by Fog Computing. Diagnostics, 13(13), 2191.

27) Shynybay, Z.; Georgieva, T.; Nedelcheva, E.; Alikhanov, J.; Moldazhanov, A.; Zinchenko, D.; Bakytova, M.; Sapargali, A.; Daskalov, P. (2026). Comparative Study of the Performance of SqueezeNet and GoogLeNet CNN Models in the Identification of Kazakhstani Potato Varieties. *AgriEngineering 8*, 17. https://doi.org/10.3390/agriengineering8010017

28) Singh, A.; Bhat, V.; Sudhakar, S.; Namachivayam, A.; Gangadharan, C.; Pulchan, C.; Sigamani, A. (2021). Multicentric Study to Evaluate the Effectiveness of Thermalytix as Compared with Standard Screening Modalities in Subjects Who Show Possible Symptoms of Suspected Breast Cancer. BMJ Open, 11, e052098

29) Smrity A., Muntaqim Z., Kafi M. (2022). Recent advancements in machine learning and deep learning for early detection of breast cancer: A comprehensive review; Innovate practices in breast health; https://doi.org/10.1016/j.ibreh.2026.100050

30) Sung H. et al., (2021)"Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021. DOI: 10.3322/caac.21660.

31) Tan, M.; Le, Q.V. (2020). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv*, arXiv:1905.11946.

32) Tang, R.-X., Yan, E.-C., Wen, T., Yin, X.-M., & Tang, W. (2021). Comparison of Logistic Regression, Information Value, and Comprehensive Evaluating Model for Landslide Susceptibility Mapping. *Sustainability*, *13*(7), 3803. https://doi.org/10.3390/su13073803

Volume 65 | Issue 07 | July 2026
DOI: 10.5281/zenodo.21698891

33) Yalım, A., Aytugar, E., Kalabalık, F., & Akdağ, İ. (2026). Comparative Analysis of Deep Learning Architectures for Automatic Tooth Segmentation in Panoramic Dental Radiographs: Balancing Accuracy and Computational Efficiency. *Diagnostics*, *16*(2), 336. https://doi.org/10.3390/diagnostics16020336

34) Yavuz Ozalp, A., Akinci, H., & Zeybek, M. (2023). Comparative Analysis of Tree-Based Ensemble Learning Algorithms for Landslide Susceptibility Mapping: A Case Study in Rize, Turkey. *Water*, *15*(14), 2661. https://doi.org/10.3390/w15142661

35) Zhang, M.; Xue, M.; Li, S.; Zou, Y.; Zhu, Q. (2023). Fusion Deep Learning Approach Combining Diffuse Optical Tomography and Ultrasound for Improving Breast Cancer Classification. Biomed. Opt. Express 14, 1636

36) Zheng, X.; Yao, Z.; Huang, Y.; Yu, Y.; Wang, Y.; Liu, Y.; Mao, R.; Li, F.; Xiao, Y.; Wang, Y.; et al. (2020). Deep Learning Radiomics Can Predict Axillary Lymph Node Status in Early-Stage Breast Cancer. Nat. Commun. 11, 1236.

37) Zhou J, Luo LY, Dou Q et al (2019) Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J Magn Reson Imaging 50:1144–1151. https://doi.org/10.1002/jmri.26721 58.