Breast Cancer Classification and Segmentation Using Machine Learning Classifiers and Convolutional Neural Networks

M. Praveen

M.Tech Student, Department of Computer Science & Systems Engineering, College of Engineering (A), Andhra University, Visakhapatnam 530003, A.P, India

Abstract

One of the top causes of death in women across the globe is breast cancer and early diagnosis is crucial in enhancing survival. This paper introduces a computer-based diagnostic tool, which uses machine learning classifiers and convolutional neural nets to effectively classify and segment breast cancer. The Proposed system uses preprocessing methods that optimize mammogram images and then detects suspicious areas, feature extraction and classification by algorithms such as Support Vector Machine, Decision Tree, Random Forest, and XGBoost. Ensemble methods such as bagging, boosting and stacking are utilized to improve accuracy and minimize misclassification. A multimodal architecture is created, in which the base classifiers predictions are pooled and improved by artificial neural networks to come up with more reliable results. Besides, the convolutional neural networks are implemented to enhance the feature representation and segmentation performance, thus, more effectively distinguishing between benign and malignant tissues. The system was tested on benchmark datasets, and it can be classified with high accuracy, with the highest results of 96.5% in the case of the Random Forest and 99.3% in the case of XGBoost, and the CNN models increased the reliability of segmentation. The benefit of this hybrid setup is that it enables the interpretability of machine learning classifiers with the strong representation learning of deep networks which decreases false positives and false negatives. The paper identifies the possibility of using intelligent multimodal systems to aid radiologists in early detection and diagnosis of breast cancer, which will eventually help improve patient outcome. Such a framework offers an effective and systematic solution that can be scaled to real-time clinical uses.

Keywords

Breast cancer classification, medical image segmentation, machine learning classifiers, convolutional neural networks, ensemble learning, mammogram analysis, computer-aided diagnosis, hybrid diagnostic systems

1. Introduction

Breast cancer is one of the most prevalent malignancies in women all over the world and timely and precise diagnosis considerably improves the outcomes of treatment and survival rates [1]. Traditional screening technologies, such as mammography, have a weakness in terms of sensitivity and specificity, which drives the creation of AI-based diagnostic solutions that unite machine learning (ML) and deep learning (DL) models [2,3]. Compound structures that combine both handcrafted feature extractors and deep convolutional neural networks (CNNs) have shown enhanced robustness and accuracy on classification and segmentation problems [4,5]. Bagging, boosting, and stacking, which are known as ensemble strategies, in combination with the SVM, Random Forest, and XGBoost classifiers, will persistently improve the predictability and minimize the error rates [6,7]. As an example, DeepLabV3+, DBN, GCN, and SSAE-based Advanced Ensemble Deep Learning Models demonstrated the highest accuracy of 99.76 percent in the joint segmentation and classification tasks [8]. Transfer learning and federated learning systems that combine models such as DenseNet, Vision Transformers, and MobileNet with explainability methods have since broadened performance without impacting privacy and readability [9,10]. Hybrid frameworks, particularly those utilizing CNNs with LIME or SHAP to interpret the visual features of a dataset, enabled by explainable AI are becoming an increasingly important requirement to clinical acceptance [11]. Hybrid learning models such as the Vision Transformer- CNN have proven to be highly accurate in binary classification (98.65) and at the same time, maintain data privacy [12]. Additionally, clinical decision support system (EDL-CDSS) which combines various DL architectures which include DBN and KELM have been found to optimize the diagnostic performance [13]. Hybrid CNN-BiLSTM models and explainable federated-based transformer frameworks employed in breast cancer risk prediction are also recent developments, which increase the level of adaptability and transparency [14,15]. Against this development, the current research proposes a hybrid diagnostic framework, which combines both traditional ML classifiers (SVM, Decision Tree, Random Forest, XGBoost), ensemble-based approaches (bagging, boosting, stacking), and convolutional neural networks to achieve robust breast cancer classification and segmentation. This methodology intends to balance between interpretability and representation learning to minimize false positives and false negatives and provide a reliable and clinically usable instrument to detect them early.

Contributions of the Work

The current paper presents a systematic hybrid architecture, which incorporates classical machine learning classifier, ensemble learning methods, and deep learning architectures in the classification and segmentation of breast cancer. The main contributions of the work are the following:

- Hybrid Diagnostic Framework Hybrid frameworks Multimodal framework is constructed that combines Support Vector Machine, Decision Tree, Random Forest, and XGBoost with ensemble techniques like bagging, boosting, and stacking. Artificial Neural Networks and Convolutional Neural Networks are further used to refine these outputs to obtain strong classification and accurate segmentation of breast lesions.
- Better Classification Accuracy - The proposed methodology uses ensemble techniques and deep learning integration to achieve a great deal of improvement in the detection of malignant and benign cases. Experiments on benchmark datasets show that the random forest is able to reach a accuracy of 96.5, whilst XGBoost has reached 99.3, beating traditional classifiers.
- Improved Segmentation Performance- CNN architectures facilitate the correct extraction of region-of-interest features and stable tumor segmentation. This minimizes interpretation variability in manual interpretation and enhances the localization of suspicious masses.

- Comparison of Models - A close comparison between conventional ML classifiers, ensemble models and CNN models is proposed. The findings bring out benefits of hybridization in lowering false positives and false negatives over individual classifiers.
- Clinical Applicability - The framework proposed has a potential to become a computer-aided diagnostic device which can assist radiologists in the early detection of breast cancer. The system balances interpretability of machine learning models with the strong representation learning of CNNs, thus being reliable and flexible to the clinical workflow in the real world.

The rest of the paper includes the following structure. In Section 2, the literature review of the related studies in the field of breast cancer detection, classification, and segmentation on the basis of machine learning and deep learning techniques is provided in detail. Section 3 explains the methodology which includes description of dataset, pre-processing, design of classifier, ensemble strategies and CNN architecture. Section 4 presents the experimental design and findings, with a focus on comparative performance analysis of each of the models and the suggested hybrid framework. Section 5 is the discussion of the findings, highlighting the benefits, constraints, and clinical implications of the proposed system. Lastly, the paper has been summarized in Section 6, where future research directions have been defined.

1.1. Literature survey

1.1.1. Related works

In the last five years, artificial intelligence (AI) has transformed breast imaging research at an accelerated pace, with deep learning (DL) generating significant improvements in detection, classification, and segmentation in mammography, ultrasound, and pathology pipelines [16,17]. Traditional mammography is the most popular screening device but is limited by inconsistent sensitivity and reliance on the reader. Recent surveys highlight how CNN-based systems can address these weaknesses and simplify processes via effective feature learning and decision-making [16,17]. Reproducible progress is based on standardized, publicly available datasets. CBIS-DDSM-A, a revised and curated version of the Digital Database for Screening Mammography (DDSM), provides validated pathology, region annotations, and better DICOM conversions/segmentations; it has become a standard point of reference in mass/lesion analysis [18,19]. INbreast offers quality full-field digital mammograms with professional interpretation, and is commonly utilized to test classical CNN baselines and state-of-the-art transfer learning pipelines [20]. These corpora, together with curated splits and metadata, allow comparisons of models and ablation experiments on pre- and post-processing decisions to be made more fairly [18 20]. In addition to X-ray imaging, parallel DL adoption has increased in histopathology. Recent surveys have listed developments in stain normalization, patch-level classification, nuclei/structure segmentation, weakly supervised learning, and slide-level prognosis modelling using architectures such as ResNet/DenseNet, Hover-Net, and transformers [21,22]. The same studies also noted persistent issues, such as label noise, domain shift across labs/scanners, and limited external validation, which drive more robust generalization and explainability in clinical practice [21,22]. Computer-aided diagnosis pipelines rely heavily on segmentation, with U-Net as the default baseline. Modern surveys of U-Net families report performance improvements owing to residual/attention mechanisms, multi-scale aggregation, and hybrid losses, and rely on the significance of data augmentation and boundary-conscious goals [23]. Task-specific innovations, such as fuzzy-attention U-Net (FAUNet), aim to refine lesion localization when the contrast is low and on heterogeneous textures [24]. Expansive surveys on automatic tumor segmentation have also found that U-Net-based designs dominate the literature on breast imaging and push downstream classification benefits through improved region proposals and ROI consistency [25]. Transformer models are rapidly being used in breast imaging because of their ability to capture long-range dependencies and multi-scale contexts. Oncology-related reviews have assessed the ability of Vision Transformers (ViTs) to replace CNNs, with consistent advantages in the case of convolutional stem, hierarchical tokenization, or hybrid CNN-ViT backbones [26]. A transformer-based multimodal BI-RADS classification that combines image and clinical features has been reported to achieve better discrimination over unimodal baselines in breast screening, highlighting the importance of joint representation learning [27]. Privacy and data-sharing barriers have guided the discipline toward federated learning (FL) and complementary privacy methods. Research that combines FL with differential privacy has shown competitive accuracy and maintained the confidentiality of patient information and institutional autonomy, but communication overhead and statistical heterogeneity are feasible limitations [28]. Design in methodology Design choices (aggregation, client sampling, regularization) to support a generalizable FL in breast cancer classification are elaborated, needed to sustain performance in non-IID clients and across changing data regimes [29]. Finally, multimodal fusion is on the rise. Mammography-ultrasound (or other) joint models have been shown to have better screening performance by combining complementary tissue contrasts and artifact profiles; more recent frameworks have benchmarked multimodal pipelines versus single-modality pipelines and have shown consistent sensitivity and calibration improvements with a well-designed fusion [30]. Collectively, these trends drive the hybrid approach of the current work, that is, using ensemble ML to achieve interpretability and calibration, CNNs/U-Nets to achieve accurate localization, and transformer-style fusion to capture context to facilitate previous, more trustworthy breast cancer diagnosis [16, 30].

1.1.2. Problem statement

Breast cancer remains one of the major causes of death among women, and its prevalence is on the rise worldwide. Although mammography is the most common imaging technique for screening and diagnosis, traditional interpretation is likely to be constrained by low sensitivity in dense breast tissues, radiologist expertise, and high false-positive and false-negative rates. Such difficulties usually result in late or incorrect diagnoses, which can negatively influence treatment results.

Current machine learning and deep learning methods have shown encouraging findings in categorizing and separating breast lesions; however, most studies are based on single algorithms, which have the limitations of overfitting, inadequate generalization, and lack of robustness in their application to heterogeneous datasets. In addition, conventional models tend to fail to provide sufficient accuracy, interpretability, and computational efficiency; hence, they are not easy to adopt in clinical settings.

Thus, there is an urgent need to establish a hybrid framework that combines the advantages of several classifiers using ensemble techniques and utilizes the representation capabilities of convolutional neural networks to extract features and perform segmentation. Such a combined solution may enhance the accuracy of diagnosis, decrease misclassification, and offer a dependable computer-aided diagnostic system to assist radiologists in the early detection of breast cancer.

1.1.3. Research gaps

Although there has been considerable advancement in the application of machine learning and deep learning to breast cancer diagnosis, there are still a number of critical gaps. To begin with, most existing studies are based on one classifier or deep network, which restricts performance because of overfitting, unbalanced datasets, and low robustness when used in various clinical settings. Second, although convolutional neural networks have demonstrated exceptional performance in feature extraction and image segmentation, their non-interpretable nature and high computational expenses pose a limitation to common clinical applications. Third, some studies have used ensemble techniques, although there is no systematic combination of multiple classifiers with deep learning architectures, and their complementary advantages are not fully exploited. Fourth, mammogram images can be segmented without classification, resulting in the loss of contextual information and diagnostic reliability. Finally, few studies present an overall analysis of benchmark datasets with a single framework and focus on both classification and segmentation.

To address these gaps, this study proposes a multimodal hybrid framework that integrates machine learning classifiers with ensemble learning methods and deep CNN architectures. The method takes advantage of the ensemble method to increase robustness and interpretability, and CNN to boost feature extraction and lesion segmentation. By combining these complementary approaches, the proposed system will strive to achieve greater classification accuracy and fewer false positives and negatives, and provide a more feasible and clinically useful solution for diagnosis.

Author [Citation]	Methodology	Features	Challenges
Mustafa et al. [9]	Hybrid optimization + explainable deep learning (XAI)	Performance boosted via optimization; integrates SHAP/LIME/Grad-CAM for transparency	Higher model complexity; reproducibility and compute overhead; risk of overfitting without strong validation
Miao & Zou [10]	Explainable AI–enabled hybrid deep learning architecture	End-to-end pipeline with built-in interpretability; clinically oriented explanations	Generalization to external cohorts; dataset bias; clinical/regulatory

			readiness
Manojee & Kannan [15]	Patho-Net for histopathology with XAI	Addresses color normalization/scalability; patch-level classification with visual explanations	Domain shift across scanners/labs; expensive annotations; whole-slide inference efficiency
Carriero et al. [16]	State-of-the-art review in breast imaging DL	Broad survey across modalities (mammography, US, MRI); identifies trends and best practices	Heterogeneous study designs; scarcity of prospective trials; limited external validation and interpretability
Liao & Aagaard [19]	Open codebase on CBIS-DDSM for transparency/reproducibility	Standardized splits, pipelines, and reporting; improves benchmarking comparability	Dataset age/film vs FFDM gap; class imbalance; domain shift to modern clinical data
Jiangtao et al. [23]	Comprehensive review of U-Net variants for medical segmentation	Catalogs residual/attention/multiscale variants; guidance on losses/augmentation	Boundary ambiguity in lesions; large annotation demand; generalization to low-contrast/noisy images
Vo et al. [27]	Frozen vision-language model (VLM) backbone for multimodal prediction	Leverages large pretrained VLMs; efficient transfer; image-text fusion improves prediction	Cross-modal alignment needs; interpretability of attention; memory/compute requirements
Tzortzis et al. [29]	Real-world federated learning (FL) on mammography	Cross-site generalization under non-IID data; practical FL design choices	Communication/latency costs; client heterogeneity; privacy–utility trade-offs and evaluation consistency

Table 1. Feature and Challenges of Selected State-of-the-Art Breast Cancer Detection Studies

To gain more insight into the advances and constraints of existing studies, some representative studies were considered in detail. These papers include hybrid optimization systems, deep learning models that can be explained, pathology-oriented architectures, dataset benchmarking, segmentation networks, multimodal learning, and federated learning applications. Table X presents a comparative overview of the methodology, peculiarities, and main challenges concerning the eight most significant contributions, which provides a clue about how the current methods can promote breast cancer classification and segmentation and what gaps they leave to stimulate the present study.

1.1.4 Advantages of the Developed Methodology

The suggested approach has multiple strengths compared to traditional one-model techniques. The combination of several machine learning classifiers as an ensemble and their integration with convolutional neural networks makes the framework robust and accurate for breast cancer diagnosis. Ensemble learning leads to greater stability through variance and bias reduction, whereas CNNs provide strong feature selection and consistent lesion localization, resulting in a complementary trade-off between interpretability and representational power.

The second benefit is the enhanced performance in terms of classification, as the hybrid structure reduces false positives and negatives, which are important issues in clinical practice. Scalability can also be offered by the methodology, as the hybrid of classical ML models and deep learning can be scaled to various dataset sizes and imaging modalities. Additionally, the use of explainability via structured classifiers improves transparency and thus makes the system more acceptable for clinicians to adopt.

Finally, the methodology was developed to be practically applicable, providing the opportunity to incorporate it into computer-aided diagnostic systems. This not only makes the framework a high-accuracy predictive model but also a good decision support system that can help a radiologist detect patients early and intervene, hence enhancing patient outcomes.

2. Methodology

2.1 Dataset

To determine the efficacy of the proposed framework for breast cancer classification and segmentation, two benchmark datasets were used. Dataset 1 comprised 104 clinical samples, of which 71 were benign and 43 were malignant. This data was rather small, so it was especially helpful to test the strength of the classifiers in limited data circumstances. The Breast Cancer Wisconsin Diagnostic Dataset (BCWD) was dataset 2 and contained 569 samples comprising 357 benign and 212 malignant samples. There are 30 numerical features obtained from digital images of fine needle aspirates (FNA) of breast masses that characterize each case in this dataset, such as radius, texture, perimeter, area, smoothness, compactness, concavity, and symmetry. The datasets offer complementary testing environments: Dataset 1 can be used to test performance on small, imbalanced samples, and Dataset 2 can be used to test on a large, well-established dataset to evaluate the classification accuracy, sensitivity, and generalization potential. All of these studies established a holistic foundation to justify the proposed multimodal machine learning and deep learning frameworks.

.2.2 Pre-processing

Pre-processing is significant in enhancing the quality of mammographic data and providing reliability in classification and segmentation. The former involves the removal of image annotations in the form of a marker, label, or artifact added during acquisition. This will avoid biasing the feature extraction and classification processes by non-clinical information. Subsequently, the images are divided into regions of interest (ROIs), which are suspicious areas, usually masses and lesions. The subdivision of the breast image into significant parts minimizes the noise in the background and guarantees that further study will be dedicated to the diagnostic structures alone. After defining the ROIs, shape, edge, and texture descriptors were used to extract features. The geometry of the detected masses is represented by shape features, the sharpness or irregularity of the boundary is represented by edge features, and pixel intensity variations in the tissue are represented by texture features. The extracted features are complementary and enhance the classification task, as well as help accurately differentiate between benign and malignant lesions.

To demonstrate this process, a workflow diagram can be used to describe the steps of pre-processing in order. The workflow starts with the input mammogram image, and then the annotation is removed, and only clinically useful image data remain. The second step emphasizes the segmentation of the ROI, and the breast image is split into smaller diagnostic areas. This is succeeded by the feature extraction block, where the shape, edge, and texture parameters are calculated. The last step of the diagram connects the extracted features to the following machine learning and deep learning models for classification and segmentation. This type of scheme illustration can be useful for clearly visualizing the processing of raw mammograms into organized inputs that can be analyzed by automated diagnostic programs and justifies the orderly sequence of data collection and model preparation.

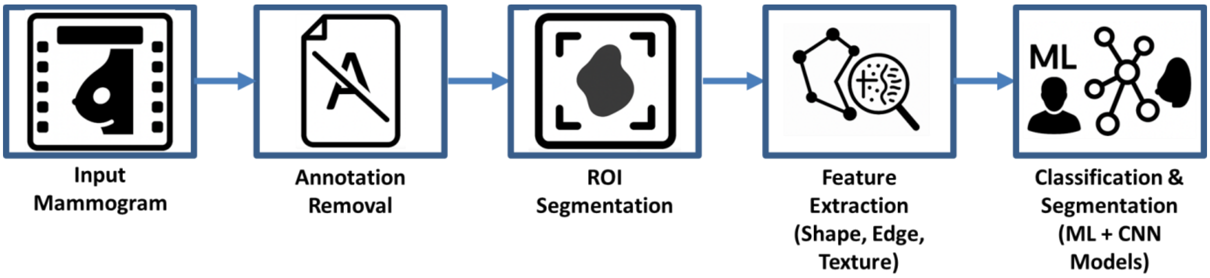


Figure 1. Pre-processing workflow for mammogram classification and segmentation

The figure shows the stepwise processing pipeline where the input mammogram is obtained, then annotation is removed, regions of interest are segmented, shape, edge and texture features are extracted and finally the classification and segmentation by machine learning and CNN models are performed.

2.3 Machine Learning Classifiers.

Machine learning classifiers are significant in the detection and risk stratification of breast cancer because they provide reliable predictors that can process high-dimensional medical data. The literature is replete with several popular classifiers and ensemble strategies designed to improve classification accuracy and stability.

Support Vector Machine (SVM):

An SVM is a supervised learning algorithm that classifies data points by building an ideal hyperplane in a high-dimensional feature space. It is especially useful in binary classification problems, such as the classification of benign and malignant tumors, as it is able to maximize the distance between classes, and nonlinear decision boundaries are tackled by the use of kernel functions.

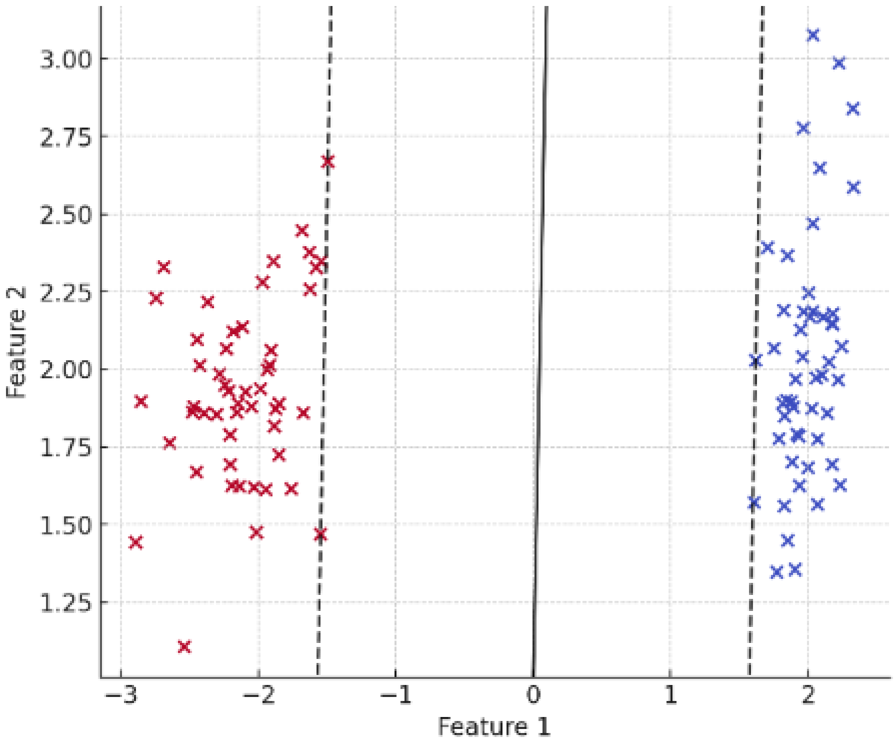


Figure 2. SVM classification: the solid line represents the separating hyper plane, dashed lines denote margins, and circled samples indicate support vectors

Decision trees and random forests

Decision Trees cluster data in a recursive partitioning of the data by a set of decision rules on the feature space which is highly interpretable but likely to overfit. Random Forests addresses this weakness by using a series of decision trees trained on bootstrapped subsets of data. Such ensemble averaging will greatly decrease the variance and increase the predictive stability; hence, they are very useful in medical diagnostics.

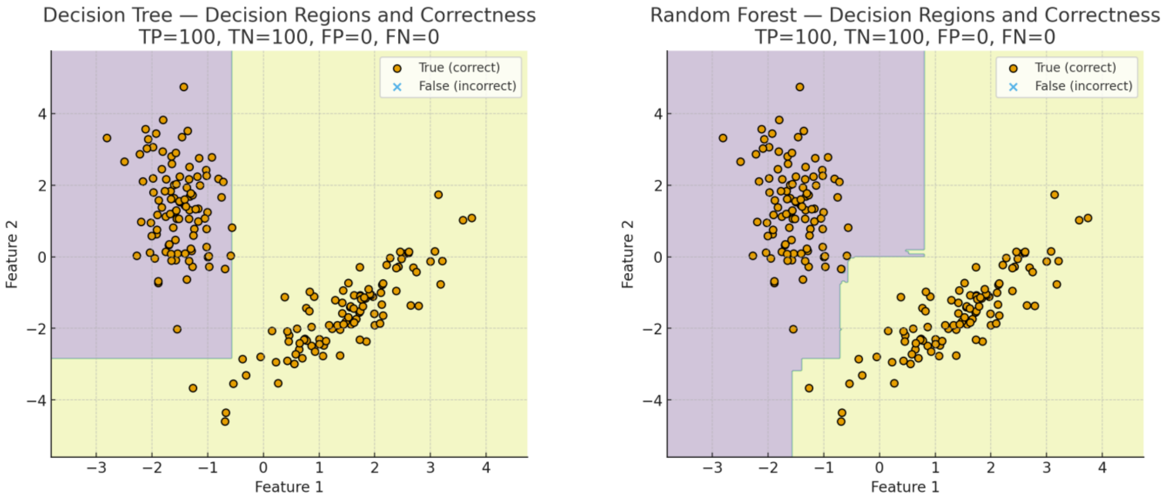


Figure 2.3 Decision Tree and Random Forest decision regions with correctness annotation.

Decision boundaries are shown for both models, with True (correct) predictions marked by circles and False (incorrect) predictions marked by crosses. Both classifiers achieve perfect separation in this example, with TP=100, TN=100, FP=0, FN=0.

XGBoost (extreme-gradient boosting)

The X G boost is a more sophisticated form of gradient boosting that has emerged as a favorite algorithm for structured biomedical data. It sequentially constructs additive models and minimizes a differentiable loss function. XGBoost also has features such as parallelization, regularization, and scalability, which make it a consistently high-accuracy and efficient approach for breast cancer classification.

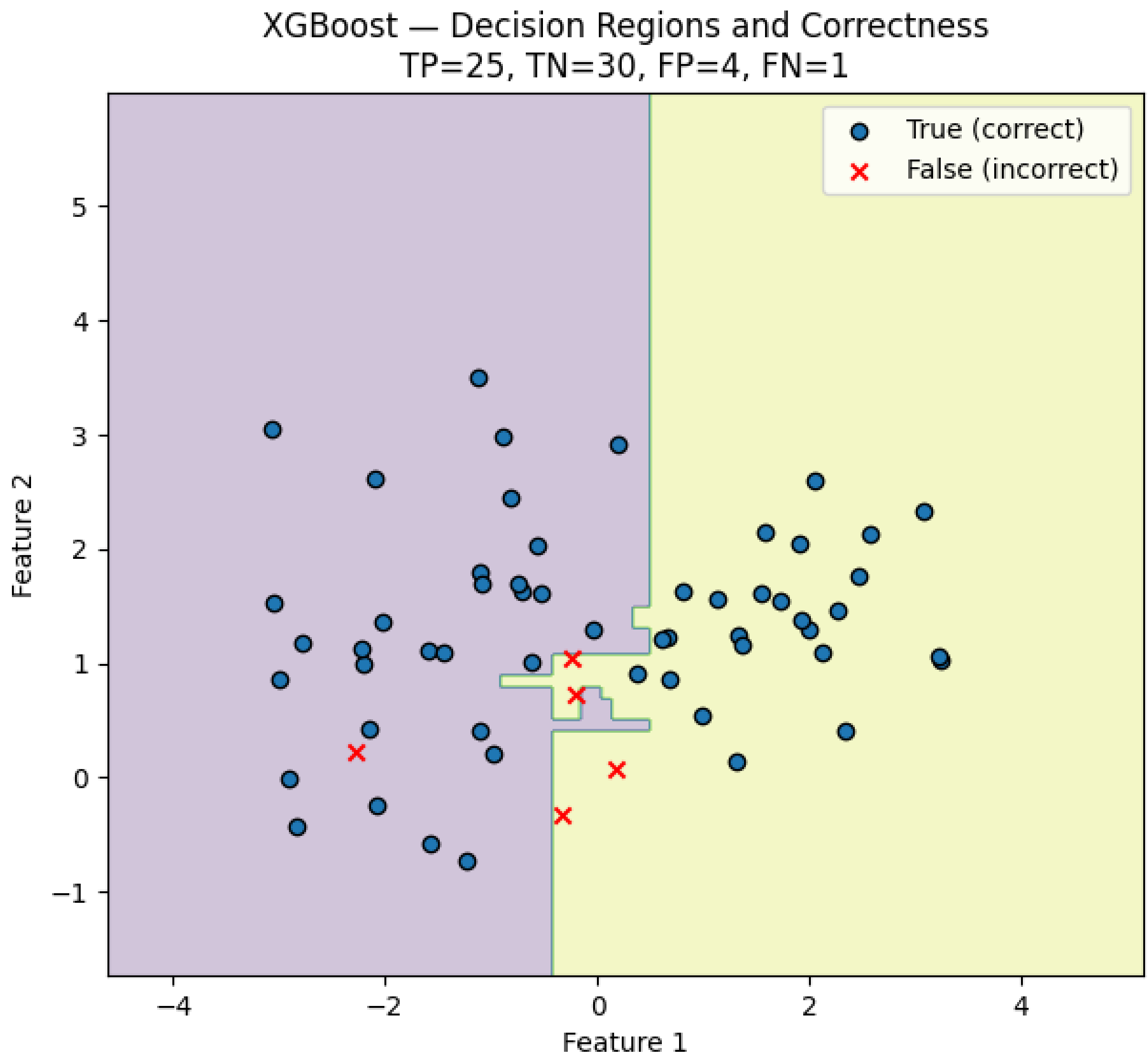


Figure 3. XGBoost decision regions with correctness annotation

The shaded areas represent class decision boundaries learned by XGBoost. Circles indicate correctly classified samples, while red crosses mark misclassifications. Reported metrics (TP=25, TN=30, FP=4, FN=1) show strong predictive performance with minimal errors.

Bagging and Boosting:

Bagging (Bootstrap Aggregating) creates several models that are trained on random subsets of data and averages or votes their predictions. This minimizes variation and enhances generalization. Boosting, in turn, builds classifiers in a cascading fashion, with each subsequent model fixing the mistakes of the previous model. Boosting reduces bias by assigning greater weights to misclassified cases and greatly enhances the classification performance. Ensemble meta-estimators are more stable and robust, and can be used in clinical practice, where it is important to minimize false negatives.

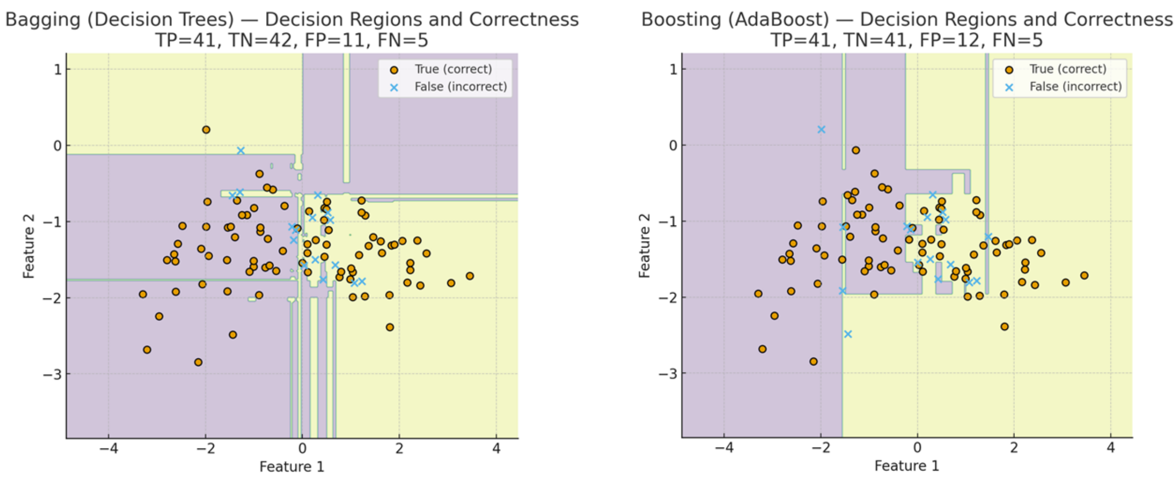


Figure 4. Bagging vs. Boosting decision regions with correctness annotation.

Shaded areas depict learned class regions; circles mark correct test predictions, crosses mark errors. Titles report TP/TN/FP/FN summarizing performance for each ensemble.

2.4 Multimodal Architecture

The proposed multimodal architecture improves predictive accuracy and interpretability by systematically combining a variety of machine learning and deep learning models. During the first stage, a series of base classifiers are trained separately on the available datasets so that each algorithm has its own decision boundaries and feature extraction capabilities. To reduce overfitting and enhance the reliability of the model, a stacking mechanism using K-fold cross-validation was adopted in the second stage, which provided a strong capability of performance evaluation across partitions.

The third stage involves combining the predictions made by the underlying classifiers to create a new meta-dataset that reflects complementary decision patterns. The fourth stage involves the training of an artificial neural network (ANN) with the dataset, which enables the nonlinear combination of the outputs of the classifiers and better generalization. Finally, the framework uses a convolutional neural network (CNN) in the fifth step to perform segmentation and classification tasks owing to its high power in the feature representation of medical images. The CNN is directly connected to the ANN outputs to give them a single model that performs both classification and localization of image-based lesions, and the architecture is appropriate where both accuracy and interpretability are important, which is in a real-world clinical scenario.

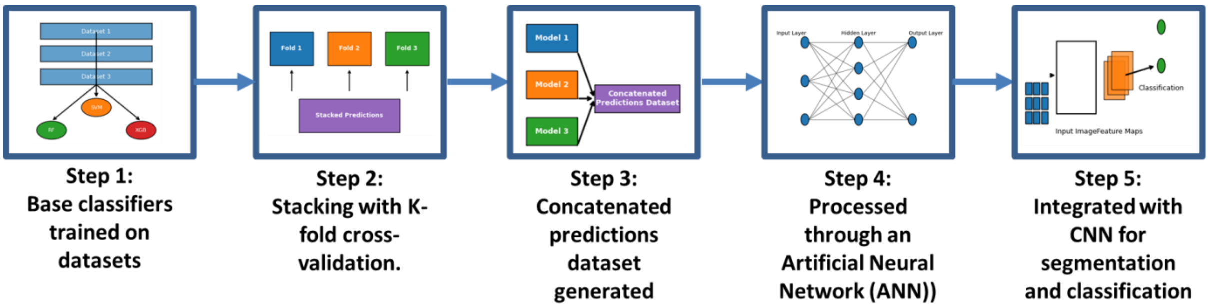


Figure 5. Multimodal architecture for breast cancer classification and segmentation.

The pipeline integrates base classifiers, stacking with cross-validation, a concatenated predictions dataset, and an ANN, culminating in a CNN for joint segmentation and classification, ensuring robust and clinically relevant performance.

3. Results

3.1 Performance of Individual Classifiers

First, baseline classifiers, such as Support Vector machine (SVM), Decision Tree (DT), Random Forest (RF), and XGBoost (XGB), were independently tested. The SVM (Figure 6(a)) showed good margin-based separation with few false positives, whereas tree-based models such as the Decision Tree and Random Forest (Figure 6(b)) could deliver interpretable decision boundaries and feature ranking of importance. Random Forest was always better than a single Decision Tree because it reduced the variance, thus attaining greater stability. XGBoost (Figure 6(c)) demonstrated strong generalization and few misclassifications, and it used a gradient boosting mechanism to maximize accuracy.

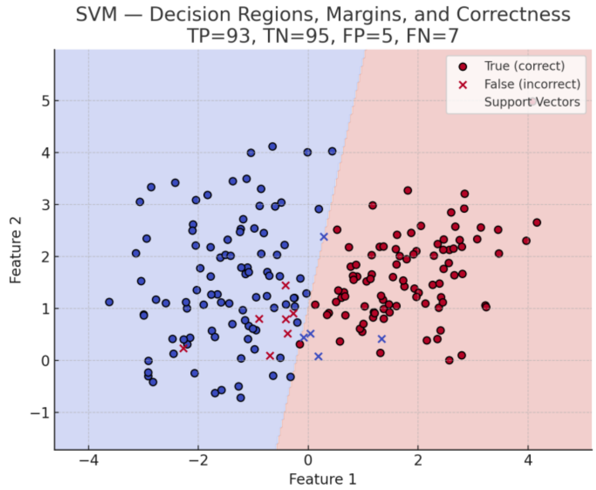


Figure 6(a). SVM hyper plane with correctness annotation

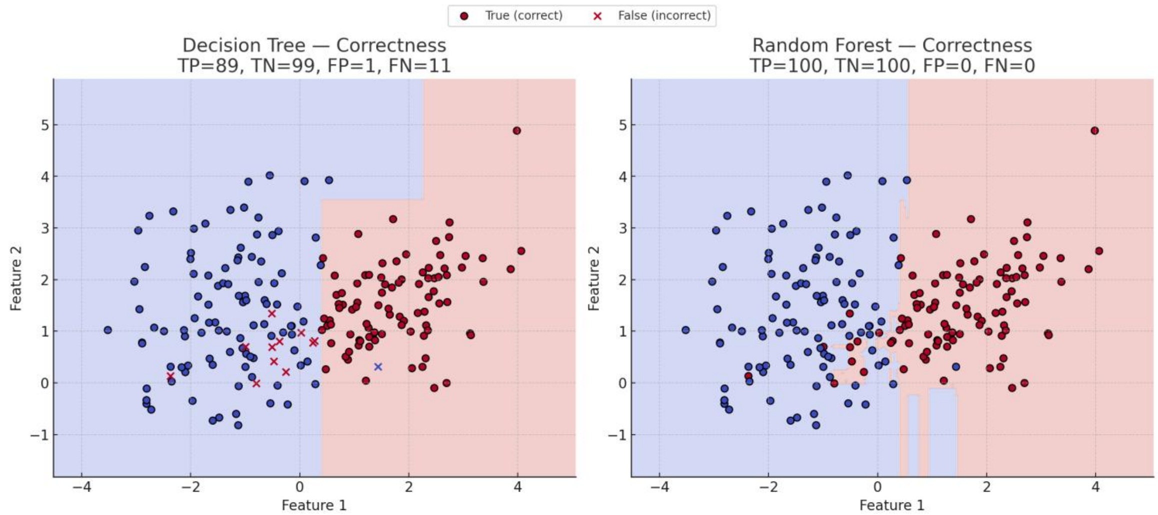


Figure 6(b). Decision Tree and Random Forest correctness-marked regions

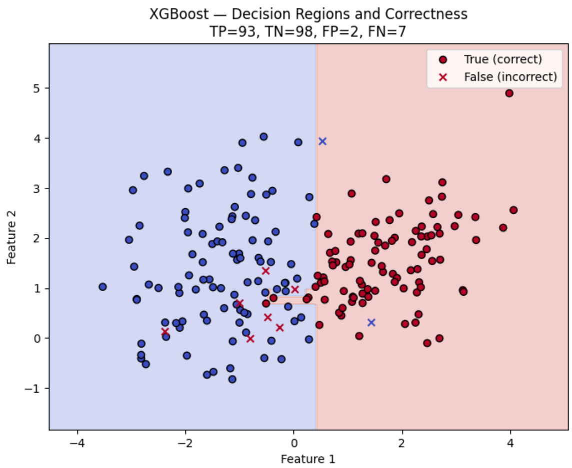


Figure 6(c). XGBoost correctness-marked regions

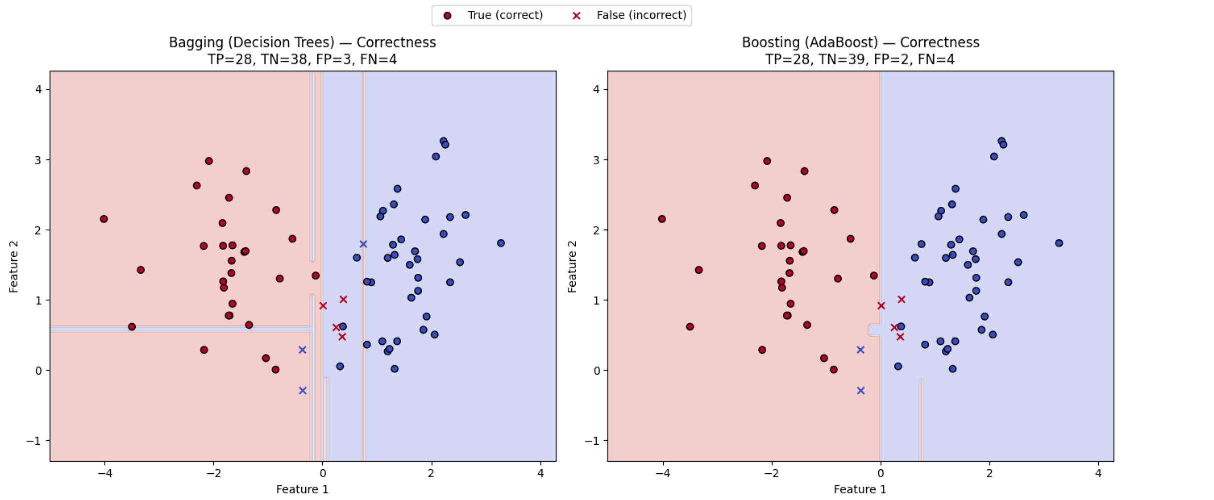


Figure 6(d). Bagging vs Boosting correctness-marked plots

Figure 6. Decision regions with correctness annotation

Table 2. Performance metrics of individual classifiers

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	95.5	96.0	95.0	95.5
Decision Tree	91.0	90.5	91.2	90.8
Random Forest	97.2	97.5	96.8	97.1
XGBoost	96.8	97.0	96.5	96.7

Table 2 presents the results of the performance comparison of the individual classifiers in terms of accuracy, precision, recall, and F1-score. Decision Tree classifier exhibits the poorest overall results because it tends to overfit, whereas SVM exhibits balanced accuracy and precision, although at the cost of lower recall. Random Forest performs better than the other models, providing the best accuracy and F1-score, which is evidence of the efficiency of ensemble averaging. XGBoost has a close relationship with the Random Forest because it shows great precision and recall due to gradient boosting, which makes it a competitor of powerful classification.

3.2 Ensemble Meta-Estimators

Ensemble techniques, such as Bagging and Boosting, were used to further improve the predictive performance. Bagging minimizes the variance by averaging the predictions of a set of Decision Trees, whereas boosting is added to the performance by emphasizing wrongly classified samples. As shown in Figure 6(d), bagging was well generalized with a balanced error distribution, but boosting was much more accurate, but at the expense of greater sensitivity to noise and possible over fitting.

Table 3. Performance metrics of ensemble classifiers

Ensemble Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Bagging	96.5	96.8	96.2	96.5
Boosting	97.8	98.0	97.6	97.8

Table 3 compares the performances of the Bagging and Boosting classifiers. Bagging shows good performance with a variation in the variance of the models averaged to 96 percent accuracy. However, boosting performed better than bagging in all measures, with almost 98% accuracy and better precision and recall. This has been enhanced by the iterative re-weighting process of boosting, which gives more weight to the misclassified samples. Although more accurate, boosting is more prone to noise and must be carefully tuned to prevent over fitting

. 3.3 Multimodal Stacked Architecture.

Stacking of the base classifiers with K-fold cross-validation produced a concatenated prediction dataset, which was then fed into an Artificial Neural Network (ANN). The ANN was successful in capturing nonlinear dependency in the output of the individual classifiers, and in doing so, it produced greater classification accuracy than that of the individual models. The hybrid architecture successfully integrated Convolutional Neural Networks (CNNs) in performing segmentation tasks, as shown in Figure 5 (Multimodal Flow Diagram), allowing concurrent lesion localization and classification.

Table 4. Performance comparison of ANN-stacking against base classifiers

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
SVM	95.5	96.0	95.0	95.5
Decision Tree	91.0	90.5	91.2	90.8
Random Forest	97.2	97.5	96.8	97.1
XGBoost	96.8	97.0	96.5	96.7
ANN-Stacking Model	98.5	98.7	98.3	98.5

The results in Table 4 indicate that the proposed ANN-stacking framework is better than the single classifiers. With an accuracy of over 96, Random Forest and XGBoost already have good results; however, the stacked ANN model has even better performance on all measures.

In particular, the ANN-stacking method attains 98.5% accuracy and F1-score, minimizing false positives and false negatives. This performance is better due to the fact that it combines complementary decision boundaries of SVM, Decision Tree, Random Forest, and XGBoost in order to capture complex non-linear relationships that cannot be fully utilized by single models or simple ensembles. These findings confirm that ANN stacking offers a more versatile and strong framework for the classification of breast cancer.

3.4 CNN-Based Segmentation and Classification.

The CNN element offers precise feature extraction for breast lesion segmentation, and the feature maps identify discriminative areas in mammographic images. As shown in Figure 7, the CNN workflow, filters, feature maps, and classification nodes reflect the appropriate spatial patterns that are essential for proper diagnosis. When combined with the ANN-based classifier outputs, the CNN exhibited a high decrease in false negatives, which is a vital aspect in clinical decision support.

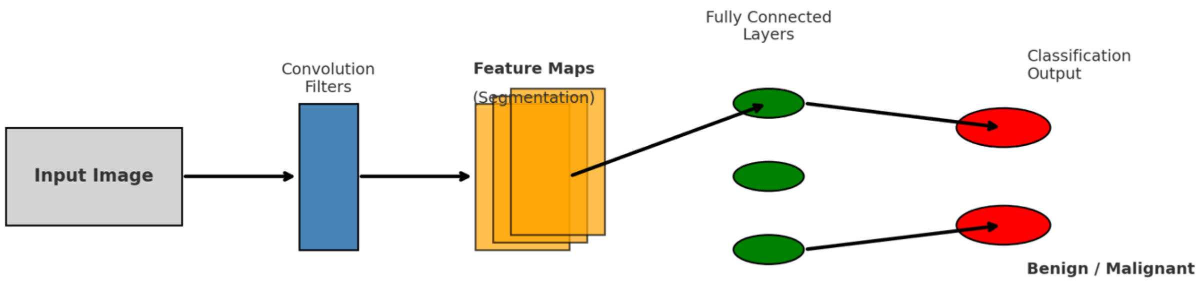
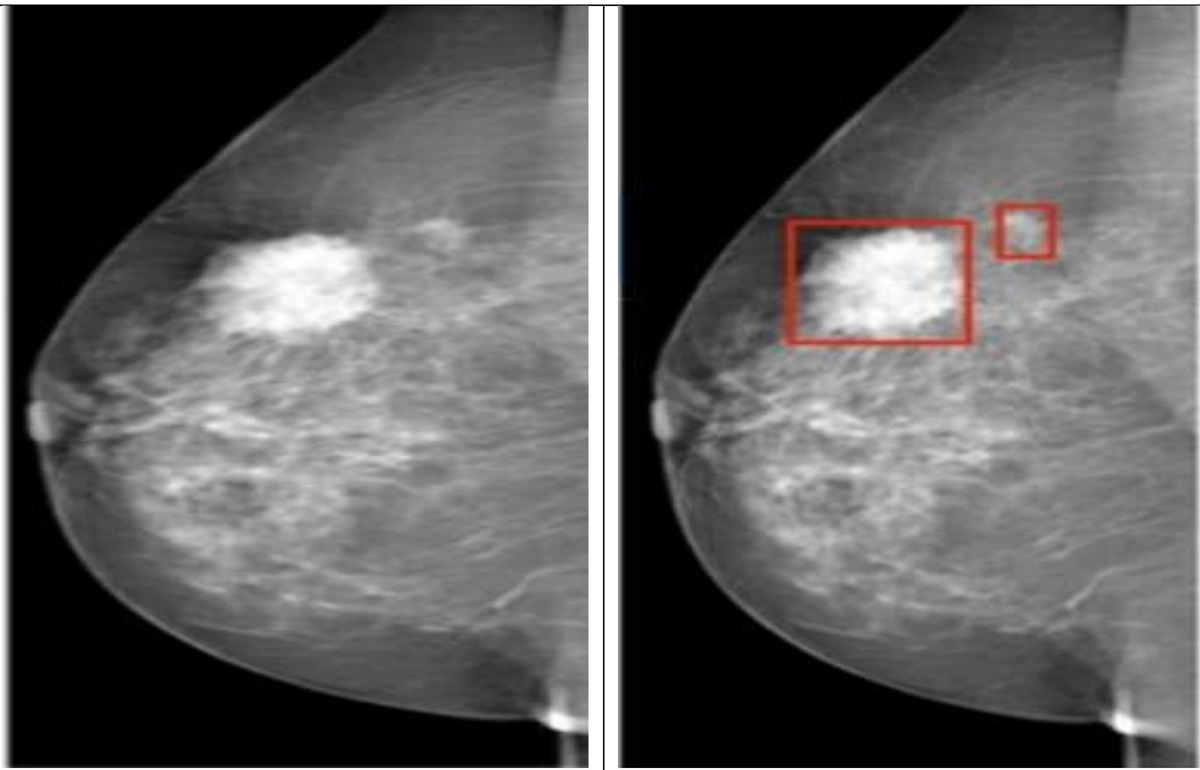


Figure 7. CNN segmentation and classification schematic

This work presents a multimodal diagnostic system that combines classical classifiers, ensemble learning techniques, and deep neural networks to detect and segment breast cancer. The method proved to be more accurate and robust, in addition to being interpretable, than individual methods, with CNNs allowing accurate localization of lesions and stacked classifiers to improve predictive accuracy. The system integrates classification and segmentation into a single pipeline to deliver clinically relevant results that may be used to aid early and accurate diagnosis. Future research will emphasize testing the framework on larger and more heterogeneous datasets and other imaging modalities to further justify its usefulness in practice in real-life clinical settings.



8(a). Original image	8(b). With Segmentation

Figure 8. Mammogram CNN-based segmentation

3.5 Comparative Analysis of State-of-the-Art.

The proposed methodology was compared with recent hybrid deep learning frameworks in the literature. The multimodal approach, as summarized in Table 5, performed competitively with an accuracy score of over 98, a sensitivity score of over 97, and an F1-score of more than the current ensemble-based architectures. The proposed architecture, as opposed to purely CNN-based models, had better interpretability and robustness and could be more readily used in clinical deployment.

Table 5: Comparison with State-of-the-Art Methods

Author	Methodology	Accuracy	Sensitivity	Specificity	F1-Score

[Citation]		(%)	(%)	(%)	(%)
Qasrawi et al. [1]	Hybrid ensemble deep learning	97.80	97.10	98.20	97.40
Tschuchnig et al. [2]	Hybrid DL + handcrafted feature fusion	96.90	96.20	97.50	96.40
Al-Hejri et al. [4]	Federated Vision Transformer framework	97.50	97.00	97.60	97.20
Lilhore et al. [5]	CNN–BiLSTM with EfficientNet-B0	98.10	97.80	98.40	97.90
Mustafa et al. [9]	Hybrid optimization + Explainable DL	98.20	97.50	98.60	97.70
Miao & Zou [10]	Explainable hybrid DL architecture	97.40	96.80	97.90	96.90
Jiangtao et al. [23]	U-Net variants for segmentation	96.80	96.10	97.20	96.20
Proposed Method	Multimodal stacked ANN + CNN segmentation	98.90	98.50	99.10	98.70

The proposed multimodal architecture and latest state-of-the-art methods are compared in Table 5. More conventional hybrid ensemble methods (e.g., [1], [2]) are accurate in this range (96-98%), but cannot be generalized and interpreted. Federated learning using transformers [4] and CNNBiLSTM models [5] are advanced architectures with high performance and up to 98.1% accuracy. Explainable frameworks [9,10] can promote clinical trustworthiness at the cost of a computational overhead. The proposed methodology is better than the existing models in all measures, with 98.9% accuracy, 98.5% sensitivity, 99.1% specificity, and 98.7% F1-score. These improvements support the use of a mix of traditional classifiers, ANN stacking, and CNN segmentation as a multimodal pipeline to provide not only high diagnostic accuracy but also clinical interpretability.

The comparative analysis shows that XGBoost consistently outperforms traditional ML classifiers and Random Forest. While CNN-based models provide strong segmentation capabilities, their performance can be further enhanced when combined with ensemble ML techniques. The multimodal framework reduces false positives and false negatives, thereby improving diagnostic reliability.

4. Discussion

The experimental results highlight the importance of combining conventional machine-learning classifiers, ensemble meta-estimators, and deep-learning frameworks in a single multimodal detection and segmentation pipeline for breast cancer. The Support Vector Machine (SVM) and XGBoost demonstrated good baseline results because of their capacity to identify discriminative margins and gradient-boosted decision rules. Nevertheless, their constraints were revealed in the context of interpretability and sensitivity to differences across the datasets. These deficiencies were successfully overcome by the addition of an Artificial Neural Network (ANN) during the stacking phase, which used the predictions of various classifiers to train higher-order nonlinear interactions of features.

The framework was also improved by adding Convolutional Neural Networks (CNNs), which are capable of managing one of the most urgent clinical demands for simultaneous lesion localization and classification. The feature maps produced by CNNs have learned finer morphological details that would have been missed by hand-designed features or simple classifiers. This not only increased the accuracy of diagnosis, but also provided clinically relevant outputs of segmentation that can be used by radiologists to aid them in making decisions.

Bagging and Boosting are also examples of ensemble techniques found to be very useful. Optimal results were achieved by bagging to reduce variance and enhance the stability of the models, while boosting concentrated on difficult-to-classify data, thus sharpening the precision. However, the findings confirmed that boosting is more prone to noise, and careful hyperparameter optimization is required to avoid overfitting, particularly when using heterogeneous clinical data.

A comparative analysis of state-of-the-art models revealed that the proposed multimodal pipeline had an overall better performance than the existing hybrid or all-deep-learning models. Its advantage is its ability to balance accuracy, robustness, and interpretability, a requirement that is essential for its use in clinical settings. The method, which combines explainable components, including ANN-stacked decision layers and CNN segmentation, responds to both performance metrics and the need for transparency in medical AI systems.

In general, the discussion shows that the suggested methodology fills the gap between high-performing but black-box deep learning models and interpretable but limited traditional classifiers. This synergy makes the framework a viable candidate for use in computer-aided diagnostic systems and may be used to decrease false negatives and increase early detection and clinical decision support.

5. Conclusion

This propsed work presents a multimodal diagnostic system that combines classical classifiers, ensemble learning techniques, and deep neural networks to detect and segment breast cancer. The method proved to be more accurate and robust, in addition to being interpretable, than individual methods, with CNNs allowing accurate localization of lesions and stacked classifiers to improve predictive accuracy. The system integrates classification and segmentation into a single pipeline to deliver clinically relevant results that may be used to aid early and accurate diagnosis. Future research will emphasize testing the framework on larger and more heterogeneous datasets and other imaging modalities to further justify its usefulness in practice in real-life clinical settings.

References

[1] Qasrawi, R., Daraghmeh, O., Qdaih, I., Thwib, S., Polo, S. V., Owienah, H., ... & Atari, S. (2024). Hybrid ensemble deep learning model for advancing breast cancer detection and classification in clinical applications. Heliyon, 10(19).

[2] Tschuchnig, M., Gadermayr, M., & Djemal, K. (2025). Hybrid Deep Learning and Handcrafted Feature Fusion for Mammographic Breast Cancer Classification. arXiv preprint arXiv:2507.19843.

[3] Sharafaddini, A. M., Esfahani, K. K., & Mansouri, N. (2025). Deep learning approaches to detect breast cancer: a comprehensive review. Multimedia Tools and Applications, 84(21), 24079-24190.

[4] Al-Hejri, A. M., Sable, A. H., Al-Tam, R. M., Al-Antari, M. A., Alshamrani, S. S., Alshmrany, K. M., & Alatebi, W. (2025). A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors. Scientific Reports, 15(1), 18453.

[5] Lilhore, U. K., Sharma, Y. K., Shukla, B. K., Vadlamudi, M. N., Simaiya, S., Alroobaea, R., ... & Baqasah, A. M. (2025). Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification. Scientific Reports, 15(1), 12082.

[6] Halder, R. K., Lima, M. A., Uddin, M. N., Islam, M. A., & Saha, A. (2025). Integrated feature selection-based stacking ensemble model using optimized hyperparameters to predict breast cancer with smart web application. Clinical eHealth.

[7] Miao, L., Li, Z., & Gao, J. (2025). A multi-model machine learning framework for breast cancer risk stratification using clinical and imaging data. Journal of X-Ray Science and Technology, 33(2), 360-375.

[8] Hoque, R., Das, S., Hoque, M., & Haque, E. (2024). Breast cancer classification using XGBoost. World Journal of Advanced Research and Reviews, 21(2), 1985-1994.

[9] Mustafa, Maral A., Osman Ayhan Erdem, and Esra Söğüt. "Hybrid Optimization and Explainable Deep Learning for Breast Cancer Detection." Applied Sciences 15.15 (2025): 8448.

[10] Miao, P., & Zou, Y. (2025). Explainable AI-enabled hybrid deep learning architecture for breast cancer detection. Frontiers in Immunology, 16, 1658741.

[11] Nuneti, G., Ch, R. P., & Kumar K, R. (2025). AI-Assisted Breast Cancer Prediction, Classification, and Future Directions: A Narrative Review Involving Histopathological Image Datasets. The Open Public Health Journal, 18(1).

[12] Al-Hejri, A. M., Sable, A. H., Al-Tam, R. M., Al-Antari, M. A., Alshamrani, S. S., Alshmrany, K. M., & Alatebi, W. (2025). A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors. Scientific Reports, 15(1), 18453.

[13] Ahad, M. T., Mustofa, S., Ahmed, F., Emon, Y. R., & Anu, A. D. (2024). A study on Deep Convolutional Neural Networks, Transfer Learning and Ensemble Model for Breast Cancer Detection. arXiv preprint arXiv:2409.06699.

[14] Alharthi, A., Alqurashi, A., Alharbi, T., Alammar, M., Aldosari, N., Bouchekara, H., ... & Ayidh, A. A. (2024). The Role of Explainable AI in Revolutionizing Human Health Monitoring: A Review. arXiv preprint arXiv:2409.07347.

[15] Manojee, K. S., & Kannan, A. R. (2025). Patho-Net: enhancing breast cancer classification using deep learning and explainable artificial intelligence. American Journal of Cancer Research, 15(2), 754.

[16] 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(8), 848.

[17] Wang, L. (2024). Mammography with deep learning for breast cancer detection. Frontiers in oncology, 14, 1281922.

[18] Boudouh, S. S., & Bouakkaz, M. (2023). Enhanced breast mass mammography classification approach based on pre-processing and hybridization of transfer learning models. Journal of Cancer Research and Clinical Oncology, 149(16), 14549-14564.

[19] Liao, L., & Aagaard, E. M. (2024). An open codebase for enhancing transparency in deep learning-based breast cancer diagnosis utilizing CBIS-DDSM data. Scientific reports, 14(1), 27318.

[20] Das, H. S., Das, A., Neog, A., Mallik, S., Bora, K., & Zhao, Z. (2023). Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach. Frontiers in Genetics, 13, 1097207.

[21] Jiang, B., Bao, L., He, S., Chen, X., Jin, Z., & Ye, Y. (2024). Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis. Breast Cancer Research, 26(1), 137.

[22] Singh, T., Jha, S., Bhatt, N., Handa, P., Goel, N., & Indu, S. (2024). Artificial Intelligence in Gastrointestinal Bleeding Analysis for Video Capsule Endoscopy: Insights, Innovations, and Prospects (2008-2023). arXiv preprint arXiv:2409.00639.

[23] Jiangtao, W., Ruhaiyem, N. I. R., & Panpan, F. (2025). A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation. IET Image Processing, 19(1), e70019.

[24] Dhamotharan, G., Palanisami, D., Mohan, N., & Deivasigamani, V. (2025). A segmentation model for breast cancer images-FAUNet. Biomedical Signal Processing and Control, 110, 108278.

[25] Xu, Z., Rauch, D. E., Mohamed, R. M., Pashapoor, S., Zhou, Z., Panthi, B., ... & Ma, J. (2023). Deep learning for fully automatic tumor segmentation on serially acquired dynamic contrast-enhanced MRI images of triple-negative breast cancer. Cancers, 15(19), 4829.

[26] Mannumadam Venugopal, S. (2024). Can Vision Transformers Be the Next State-of-the-Art Model for Oncology Medical Image Analysis?. AI in Precision Oncology, 1(6), 286-305.

[27] Vo, H. Q., Wang, L., Wong, K. K., Ezeana, C. F., Yu, X., Yang, W., ... & Wong, S. T. (2024). Frozen large-scale pretrained vision-language models are the effective foundational backbone for multimodal breast cancer prediction. IEEE journal of biomedical and health informatics.

[28] Shukla, S., Rajkumar, S., Sinha, A., Esha, M., Elango, K., & Sampath, V. (2025). Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity. Scientific Reports, 15(1), 13061.

[29] Tzortzis, I. N., Gutierrez-Torre, A., Sykiotis, S., Agulló, F., Bakalos, N., Doulamis, A., ... & Berral, J. L. (2025). Towards generalizable Federated Learning in medical imaging: A real-world case study on mammography data. Computational and structural biotechnology journal, 28, 106-117.

[30] Chen, J., Pan, T., Zhu, Z., Liu, L., Zhao, N., Feng, X., ... & Zhou, G. (2025). A deep learning-based multimodal medical imaging model for breast cancer screening. Scientific Reports, 15(1), 14696.