International Journal of Computer Engineering and Data Science (ISSN: 2737-8543)
Volume 4–Issue 1, January - March 2025

A Multimodal Approach to Breast-Lesion Classification Using Ultrasound and Patient Metadata

Amina Aboulmira[1], Mohamed Ouhami[2], Hamid Hrimech[3], Mohamed Lachgar[4,5]

[1,3]LAMSAD Laboratory, ENSA, Hassan First University, Berrechid, Morocco.
[2]LTI, ENSA, Chaouib Doukkali University, El Jadida, Morocco.
[4]L2IS Laboratory, FSTG, Cadi Ayyad University, Marrakech, Morocco.
[5]Higher Normal School, Cadi Ayyad University, Marrakesh, Morocco.
E-mail: a.aboulmira@uhp.ac.ma, hamid.hrimech@uhp.ac.ma, ouhami.m100@ucd.ac.ma, m.lachgar@uca.ac.ma

Article history

Received April 05, 2025
Revised April 20, 2025
Accepted April 23, 2025
Published April 24, 2025

ABSTRACT

The diagnosis and prognosis of breast cancer have been greatly improved by incorporating machine learning methods, especially through medical imaging analysis as well as clinical information. In this study, the potential of deep learning models for breast lesion prognosis was explored by integrating imaging features with clinical data to enhance predictive accuracy. Clinical data were analyzed using multilayer perceptron (MLP) classifiers, XGBoost, and Random Forest, while several convolutional neural network (CNN) architectures, such as ResNet optimized with Adam, DenseNet with stochastic gradient descent (SGD), and EfficientNet with RMSprop, were evaluated. The integration of imaging-based features with clinical data was found to significantly improve model performance, enabling more accurate risk stratification and the development of individualized treatment strategies. The highest validation accuracy and area under the curve (AUC) were achieved by the most effective models, highlighting the advantages of a multimodal approach. Although the study was conducted on a relatively small dataset and faced challenges such as missing data, the results suggest that these methods hold considerable promise for implementation in clinical practice.

Keywords: *Breast cancer, lesion prognosis, machine learning, deep learning, multimodal fusion.*

I. INTRODUCTION

Breast cancer remains one of the most prevalent and life-threatening cancers affecting women worldwide, with early detection playing a critical role in improving survival rates and treatment outcomes [1,2]. A variety of diagnostic tools are employed to identify breast abnormalities at an early stage, including mammography, magnetic resonance imaging (MRI), and ultrasound. Mammography is widely used for population screening, while MRI is often reserved for high-risk patients due to its sensitivity and cost. Among these, ultrasound has emerged as an essential adjunct modality, particularly beneficial in cases where mammography may be less effective.

Ultrasound imaging is frequently used as a non-invasive, accessible, and cost-effective tool for evaluating breast lesions, particularly in women with dense breast tissue [3]. It provides real-time imaging and is capable of distinguishing between solid and cystic masses, making it especially useful for characterizing suspicious findings detected during physical examination or mammography. Additionally, it does not involve ionizing radiation, making it a safer option for repeated use, especially in younger or pregnant patients [4]. However, despite its advantages, the interpretation of ultrasound images is often subjective and can vary significantly between radiologists, influenced by factors such as operator experience, image quality, and lesion presentation [5, 6].

This variability can lead to inconsistent diagnoses, unnecessary biopsies, or missed malignancies, ultimately impacting patient care and clinical decision-making. In some cases, benign lesions may be misclassified as suspicious, resulting in invasive procedures that contribute to patient anxiety, increased healthcare costs, and avoidable physical burden [5, 6]. Conversely, subtle signs of malignancy may be overlooked due to differences in image interpretation or operator technique, potentially delaying diagnosis and reducing the effectiveness of early intervention strategies [7]. Such inconsistencies underscore the need for standardized, objective approaches to

image analysis that minimize human error and improve diagnostic reliability. The integration of artificial intelligence, particularly deep learning-based computer-aided diagnosis (CAD) systems, offers a promising solution to mitigate these challenges by providing consistent image interpretation and decision support across clinical settings [8, 9].

AI has demonstrated significant potential in improving diagnostic accuracy, reducing interobserver variability, and assisting in the stratification of lesion malignancy risk [10, 11]. In particular, convolutional neural networks (CNNs) have shown promise in automatically extracting and learning hierarchical features from ultrasound images that may not be easily perceptible to the human eye [8]. These models can be trained to detect subtle patterns associated with malignant or benign lesions, thereby supporting more objective and consistent decision-making. Furthermore, AI-powered computer-aided diagnosis (CAD) systems have been integrated into clinical workflows to assist radiologists by highlighting areas of concern, suggesting BI-RADS categories, or even predicting pathology outcomes [9].

To further enhance diagnostic precision, recent research has increasingly focused on multi-modality approaches, which integrate various types of data such as medical imaging, clinical information, histopathological data, and even genetic markers. In breast cancer diagnosis, combining ultrasound with other modalities—such as mammography, MRI, or clinical risk factors—has been shown to significantly improve sensitivity and specificity compared to using a single modality alone [12, 13]. This integrated approach allows machine learning and deep learning models to learn complementary patterns from different data sources, resulting in more robust and comprehensive diagnostic frameworks. For instance, while ultrasound captures real-time morphological features of lesions, clinical data provide essential contextual information such as patient age, hormonal status, and family history, which can influence malignancy risk [11]. Deep learning models trained on fused data have demonstrated improved performance in lesion classification, risk stratification, and even treatment response prediction [14]. These multi-modal frameworks are paving the way toward personalized medicine by enabling more accurate and individualized diagnostic pathways, ultimately supporting better clinical decision-making and patient outcomes [15].

This study aims to explore the potential of artificial intelligence in improving the diagnosis and prognosis of breast lesions by leveraging both clinical and imaging data. Specifically, we investigate the predictive performance of various deep learning models applied to ultrasound imaging data, including state-of-the-art convolutional neural network (CNN) architectures such as EfficientNet, ResNet, and DenseNet. These models are evaluated individually based on metrics such as area under the curve (AUC), accuracy, and loss, to determine the most effective architecture for breast lesion classification. In parallel, clinical data are processed using established machine learning algorithms, including multilayer perceptrons (MLP), XGBoost, and Random Forest. A multimodal approach is then adopted by integrating the imaging-based CNN models with the clinical data representations to explore various fusion strategies. The objective of this integration is to assess whether combining these complementary data sources enhances predictive accuracy compared to unimodal models. Through a comprehensive comparison of individual and multimodal models, the study aims to demonstrate the added value of data fusion in improving diagnostic reliability and supporting more personalized approaches to breast cancer care.

The remainder of this paper is organized as follows. Section II reviews the current state of research on breast cancer diagnosis using machine learning and deep learning approaches. Section III presents the methodology, starting with a description of the dataset, the relevance of disease classes, and preprocessing techniques addressing class imbalance and data augmentation. It also details the mask application procedure, the overall workflow, and the strategies adopted for model development using both clinical and imaging data. Section IV reports the experimental results, including evaluation metrics, performance comparison, and visualizations such as confusion matrices and ROC curves. Section V introduces the multimodal fusion approaches and compares their performance and contribution to model accuracy. Section VI provides a comparative analysis with existing studies, followed by a discussion of the implications, limitations of the study, and potential directions for future work. Finally, Section VII concludes the paper by summarizing the main contributions and findings.

II. Related Works

Deep learning methods have transformed medical image analysis in recent years, especially in the early diagnosis and categorization of breast cancer. The integration of sophisticated machine learning algorithms with medical imaging modalities, including ultrasound scans, mammograms, and whole-slide pathology images, has been the subject of an expanding body of research. The results have shown promise in terms of diagnostic accuracy, efficiency, and possible clinical utility.

An interesting approach in breast cancer classification was the CAMELYON16 challenge, which evaluated the automated detection of lymph node metastases in hematoxylin and eosin-stained tissue sections of breast cancer patients. Participants developed deep learning algorithms to classify whole-slide images, which were

then compared to pathologists' diagnoses. The top-performing algorithms achieved lesion-level true-positive rates comparable to expert pathologists interpreting images without time constraints. Moreover, some algorithms outperformed pathologists under time-limited conditions, highlighting the potential of deep learning models to enhance diagnostic efficiency in clinical settings [16].

In the realm of ultrasound imaging, a multi-task deep learning framework was proposed for the segmentation and classification of breast cancer lesions. This framework addressed significant challenges in breast cancer ultrasound imaging, such as data standardization and the exclusion of non-tumor images during model training. By leveraging the inherent correlations between classification and segmentation tasks, the multi-task approach demonstrated a 15% improvement in segmentation and classification accuracy over traditional single-task models. Additionally, the framework exhibited robustness by efficiently processing various image types, including benign, malignant, and non-tumor images, thereby enhancing the applicability of deep learning systems in clinical practice [17].

Advancements have also been made in mammography-based breast cancer diagnosis. An end-to-end deep learning approach utilizing image-level labels, without the need for lesion annotations during training, was introduced. This method employed an all-convolutional network to classify mammography images and achieved superior performance across multiple datasets, such as CBIS-DDSM and INbreast. The model attained sensitivity rates exceeding 86% and demonstrated impressive AUC values of 0.91 on digitized film mammograms and 0.98 on full-field digital mammography. Notably, the study revealed that models trained on one set of mammography images could be effectively transferred to another set with minimal adjustments, suggesting the potential for deep learning models to generalize across diverse imaging systems, thereby enhancing the accuracy and reliability of breast cancer screening [18].

Further progress in mammography analysis includes the development of a Computer-Aided Detection (CAD) system based on the Faster R-CNN architecture, a renowned object detection framework. This CAD system achieved state-of-the-art performance in autonomously detecting and classifying lesions as benign or malignant, with an AUC of 0.95 on the INbreast dataset. The approach outperformed traditional CAD systems, exhibiting high sensitivity and a low false positive rate. The system's second-place finish in the Digital Mammography DREAM Challenge further underscores its potential as a reliable diagnostic tool for breast cancer, exemplifying the transformative impact of deep convolutional neural networks (CNNs) in automating the detection and classification of breast cancer lesions in mammograms [19].

Recent studies have continued to explore and enhance deep learning applications in breast cancer diagnosis. For instance, a novel multi-modal deep learning model, DeepClinMed-PGM, was developed to predict disease-free survival by integrating clinicopathological data with molecular insights. This model demonstrated improved prognostic accuracy, emphasizing the value of combining diverse data sources for comprehensive patient assessment [20]. Additionally, the CEIMVEN approach implemented modified versions of EfficientNet architectures for breast cancer detection and classification from ultrasound images, achieving high accuracy rates and showcasing the effectiveness of transfer learning and architectural enhancements in medical imaging tasks [21].

Collectively, these studies illustrate the substantial advancements in applying deep learning techniques to breast cancer diagnosis and detection. The incorporation of sophisticated architectures, such as object detection frameworks, convolutional networks, and multi-task learning, has led to significant improvements in diagnostic performance across various imaging modalities. Moreover, these developments suggest that deep learning models have the potential to significantly enhance clinical workflows, improving the precision and efficiency of breast cancer detection. As research progresses, the practical implementation of these technologies in clinical settings will be crucial to fully realize their benefits and ensure their widespread adoption in healthcare.

III. METHODOLOGY

A. Dataset Description

The dataset used in this study is derived from the "BREAST-LESION-USG" collection available on The Cancer Imaging Archive (TCIA) [22]. It comprises 266 segmented breast ultrasound images representing both benign and malignant lesions, collected from 256 patients. Each case includes comprehensive tumor-level, image-level, and patient-level annotations. These labels encompass patient age, breast tissue composition, and presenting symptoms, all verified through either biopsy or clinical follow-up. The dataset was meticulously curated, with an experienced radiologist providing detailed freehand annotations and BIRADS-based characterizations for each lesion. For biopsy-confirmed cases, histopathological classifications are also included, alongside diagnostic outcomes and the verification methods used. This clinically validated dataset serves as a valuable resource for training and evaluating models aimed at improving breast lesion classification.

B. Significance of Disease Classes

The disease classes considered in this study are defined in accordance with established clinical and radiological criteria for breast lesion characterization. These classes include benign, malignant, and normal categories, each representing distinct pathological states.

- Benign class: comprises lesions that are non-cancerous in nature. These may include cysts, fibroadenomas, and other non-malignant abnormalities. Although generally not considered life-threatening, benign lesions may still require clinical monitoring or follow-up to ensure they do not evolve or present complications over time.
- Malignant class: includes lesions that are diagnosed as cancerous. These cases are of particular clinical significance, as malignant lesions pose a serious health risk if not accurately diagnosed and promptly treated. Early identification and intervention are critical to prevent progression and metastasis.
- Normal class: refers to breast tissue samples that show no visible abnormalities. These cases serve as a reference point in comparative analysis, providing a baseline that supports the distinction between pathological and non-pathological findings during the classification process.

C. Data Preprocessing and Augmentation

1. Class Imbalance

Class imbalance in datasets can lead to biased model performance, particularly when one class significantly outweighs the others. In our dataset, the benign class comprised approximately 150 samples, the malignant class had around 100 samples, and the normal class had very few samples, resulting in a significant imbalance. To mitigate this issue, we implemented a targeted data augmentation strategy to ensure a more uniform representation across all classes.

2. Data Augmentation

To address the imbalance, a comprehensive data augmentation strategy was applied, with particular focus on the minority classes. The approach included:

- Image Augmentation: Transformations such as rotations, flipping, zooming, and shear adjustments were used to increase image variability while maintaining key visual features.
- Tabular Augmentation: Gaussian noise was introduced into tabular data to replicate natural variability and enhance the robustness of the model.
- Paired Augmentation: Ensured consistency between augmented images and their corresponding tabular data, preserving the integrity of multimodal inputs.
- Class Balancing via Targeted Augmentation: Specific augmentation techniques were applied preferentially to underrepresented classes to address the class imbalance effectively.

Figures 1 and 2 show the original class distribution and the balanced class distribution, respectively. After applying these augmentation techniques, the representation across classes was more balanced, thereby improving the stability and generalization capability of the trained model.

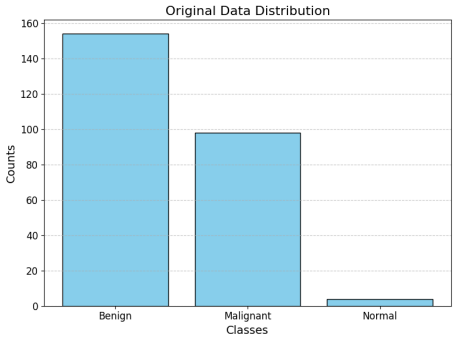

Figure 1: Original Class Distribution.

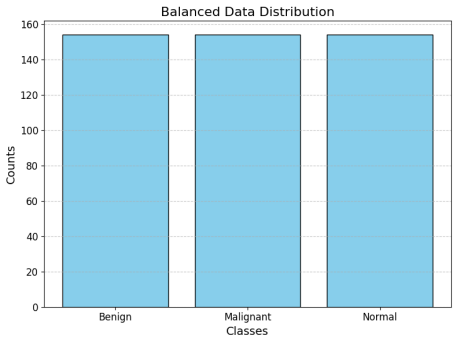

Figure 2: Balanced Class Distribution.

D. Mask Application Procedure

To prepare the tumor images for analysis, segmentation masks were applied to localize and isolate regions of interest (Fig. 3, Fig. 4). Each mask was resized to match the target resolution of the images (256×256 pixels) and converted to grayscale. Using a pixel-wise maximum operation, multiple masks were merged into a single combined mask.

The combined mask was then applied to the tumor images through element-wise multiplication, retaining the pixel intensity values within the masked regions while setting the rest to zero. This process ensured consistent and accurate isolation of the regions of interest, which

is essential for the imaging models to focus on relevant tumor areas during feature extraction.

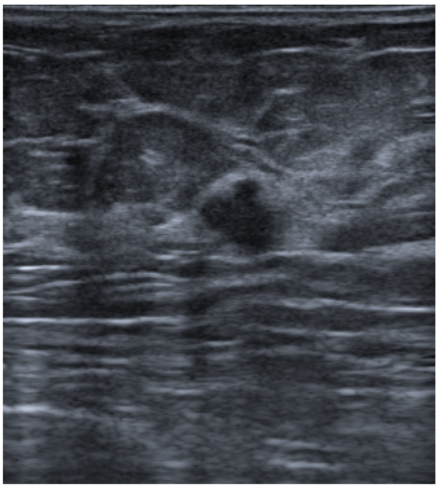

Figure 3: Original image.

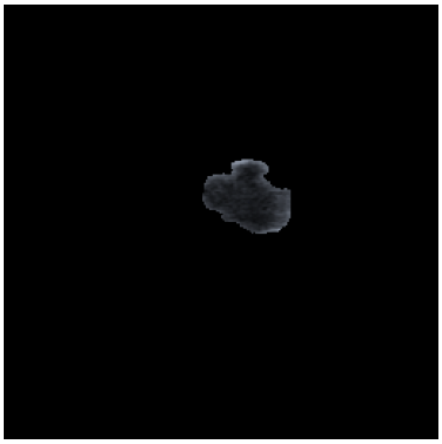

Figure 4: Processed image.

E. Overall Workflow

The proposed system adopts a multimodal learning framework designed to improve the classification and prognosis of breast lesions by integrating structured clinical data with ultrasound imaging features. This integrative approach leverages the complementary strengths of each data modality to enhance diagnostic accuracy and model generalizability. The architecture consists of three core components, each responsible for processing different types of input data before their eventual fusion.

Fig. 5, presents and describes the clinical Data Processing. Structured clinical information—including patient age, breast tissue composition, and reported symptoms—is analyzed using traditional machine learning algorithms. Models such as Random Forest, XGBoost, and Multilayer Perceptrons (MLPs) are trained and optimized through grid search-based hyperparameter tuning. These classifiers are used to identify complex patterns within tabular data, enabling robust prediction based on non-imaging clinical indicators.

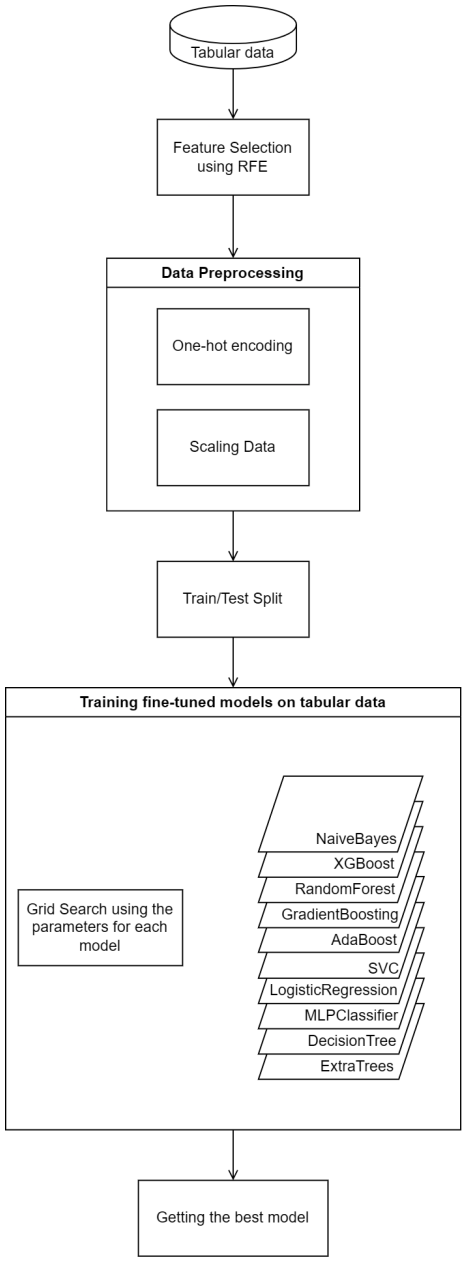

Figure 5: Architecture of the clinical (tabular) data processing module employing machine learning classifiers for structured patient data.

Fig. 6 describes the ultrasound Image Analysis. The first step is to process the images using advanced convolutional neural networks (CNNs), including Inception, Xception, and MobileNet architectures. These deep learning models are capable of automatically learning

hierarchical features from the input images. Different optimizers, such as Adam SGD, and RMSProp, are employed to enhance training efficiency and model convergence. The CNNs focus on extracting lesion-specific visual features critical for accurate classification.

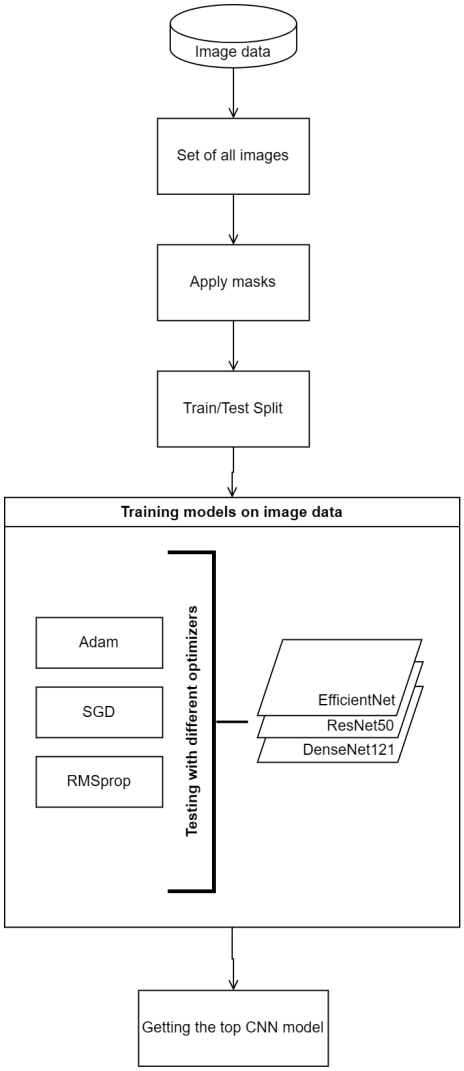

Figure 6: Architecture of the image data processing pipeline using convolutional neural networks (CNNs) for ultrasound-based breast lesion analysis.

Multimodal Fusion workflow is presented in Fig. 7. To fully exploit the richness of both data types, three fusion strategies are explored:

- Early Fusion involves the direct concatenation of raw features from clinical and imaging data at the input level, which are then passed through a joint model.
- Intermediate Fusion combines high-level abstract features extracted by the separate modality-specific models, allowing the fusion of refined representations.
- Late Fusion aggregates the final predictions from individual models using ensemble techniques such as majority voting, weighted averaging, or mean probability methods.

Among these approaches, late fusion demonstrates the most consistent and superior performance, likely due to its ability to preserve the specialized learning of each individual model while combining their outputs for a more robust final prediction. Overall, this multimodal framework offers a comprehensive and scalable solution for breast lesion diagnosis, emphasizing the potential of combining clinical and imaging data to support more accurate, reliable, and personalized decision-making in breast cancer care.

F. Model Development Strategies

1. Clinical Data: Model Selection and Hyperparameter Tuning

To analyze the structured clinical data—comprising features such as patient age, breast tissue composition, and reported symptoms—multiple machine learning algorithms were evaluated. GridSearchCV was employed to optimize model hyperparameters. The parameter grids used for each classifier are presented in table 1:

Table 1: Hyperparameter grids used for clinical data model tuning via GridSearchCV.

Model	**Hyperparameters**
Random Forest	`n_estimators`: [50, 100, 200], `max_depth`: [10, 20, None]
SVM (RBF Kernel)	`C`: [0.1, 1, 10], `gamma`: [0.01, 0.1, 1]
KNN	`n_neighbors`: [3, 5, 7]
Decision Tree	`max_depth`: [10, 20, None]
AdaBoost	`n_estimators`: [50, 100, 200]
Gradient Boosting	`n_estimators`: [50, 100, 200], `learning_rate`: [0.01, 0.1, 0.2]
XGBoost	`n_estimators`: [50, 100, 200], `learning_rate`: [0.01, 0.1, 0.2]
Extra Trees	`n_estimators`: [50, 100, 200], `max_depth`: [10, 20, None]
MLPClassifier	`hidden_layer_sizes`: [(50, 50), (100,), (50, 50, 50)]

2. Imaging Data: CNN Architectures and Optimizers

For the ultrasound imaging data, several deep convolutional neural networks (CNNs) were investigated, including EfficientNetB0, ResNet50V2, and DenseNet121.

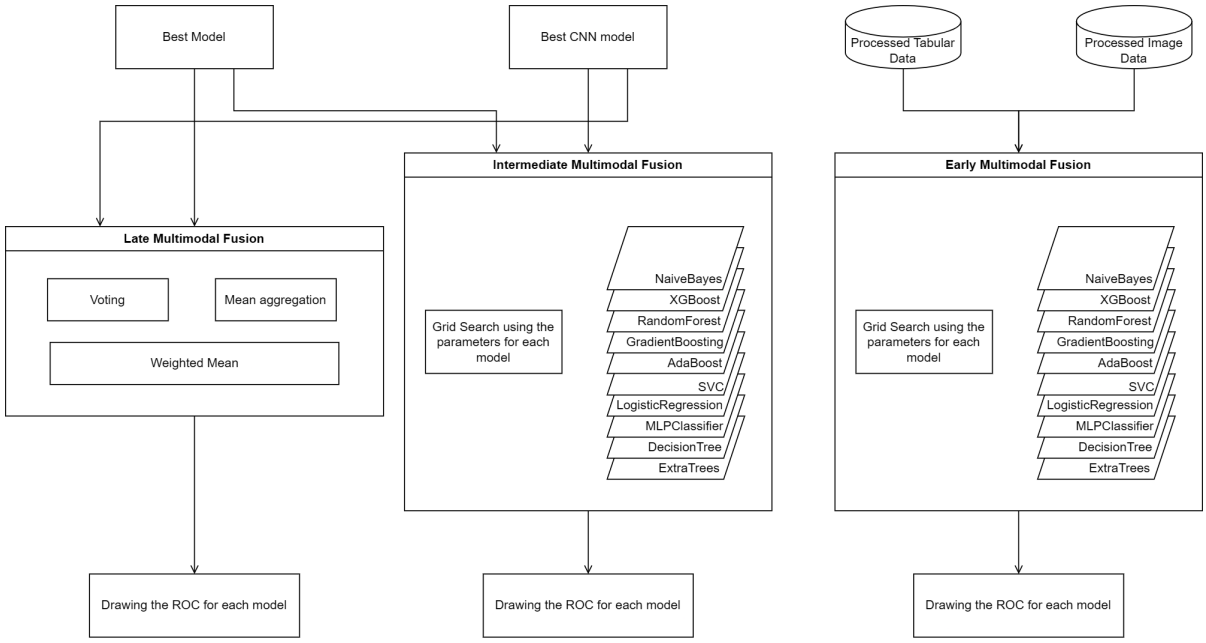

Figure 7: Multimodal fusion architecture integrating clinical and imaging data using early, intermediate, and late fusion strategies for breast lesion classification.

These architectures were selected for their proven efficacy in medical image classification tasks. To optimize training, the following optimizers were tested:

- Adam (learning rate: 0.0001): An adaptive optimizer that computes individual learning rates using first and second moments of the gradients.
- SGD (learning rate: 0.01): A classical method where model weights are updated in the opposite direction of the gradient.
- RMSprop (learning rate: 0.001): An optimizer that adapts the learning rate based on recent gradient magnitudes, particularly effective in handling noisy or non-stationary data.

IV. EXPERIMENTAL RESULTS

This section presents the evaluation protocols and experimental findings of the proposed system for breast lesion classification. The objective was to assess the performance of various models trained on individual modalities (clinical and imaging data), as well as their combinations using multimodal fusion strategies.

A. Evaluation Metrics and Experimental Setup

To ensure a comprehensive and reliable evaluation, several performance metrics and validation techniques were employed. The Receiver Operating Characteristic (ROC) curve was used to assess the discriminative power of each model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different thresholds. Additionally, the F1-score was calculated to capture the balance between precision and recall which is particularly important in the context of class imbalance. To validate model robustness, K-Fold Cross-Validation was implemented, ensuring that model performance was not overly dependent on a specific subset of the data. Finally, hyperparameter optimization was carried out using GridSearchCV, allowing systematic fine-tuning of model parameters such as learning rate and tree depth to enhance predictive accuracy.

B. Performance Comparison

To evaluate the performance of individual models trained on clinical data alone, multiple classifiers were tested using optimized hyperparameters as described in the methodology. Table 2 presents the top three models ranked by test accuracy, alongside their corresponding F1-scores and area under the ROC curve (AUC).

Table 2: Performance of the top three clinical models ranked by test accuracy. Metrics include test accuracy, F1-score, and area under the ROC curve (AUC).

Model	Accuracy	F1-Score	AUC
XGBoost	0.989247	0.989245	1.000000
Gradient Boosting	0.978495	0.978472	0.998959
Decision Tree	0.978495	0.978472	0.983871

The XGBoost model demonstrated the highest per-

formance across multiple metrics, achieving a Test Accuracy of 0.989, a Test AUC of 1.0, and a perfect F1-Score of 0.989. Similarly, SVC and Gradient Boosting showed remarkable performance, with both achieving a Test AUC of 1.0 and high F1-Scores of 0.987 and 0.978, respectively. Decision Tree and ExtraTrees also performed strongly, with Test AUC values of 0.984 and 0.983, indicating their capability to distinguish between classes effectively. Random Forest achieved a balanced performance with a Test Accuracy of 0.968, a Test AUC of 0.985, and an F1-Score of 0.976. Among neural network-based models, MLP recorded competitive results with a Test Accuracy of 0.946 and a Test AUC of 0.995, highlighting its potential for handling complex data structures. Conversely, models like Logistic Regression and Naive Bayes yielded relatively lower accuracies of 0.892 and 0.891, respectively, alongside lower AUC scores of 0.972 and 0.983. Finally, AdaBoost showed the least performance among ensemble methods, with a Test Accuracy of 0.903 and an AUC of 0.965, emphasizing room for improvement in handling this dataset. Figure 8 depicts the ROC curves, showcasing the strengths of ensemble methods.

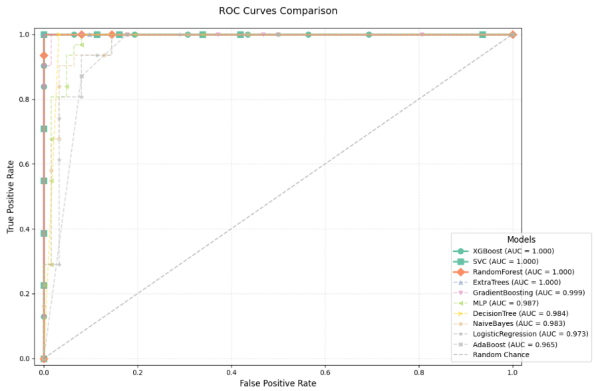

Figure 8: ROC curves of fined-tuned models after cross-validation.

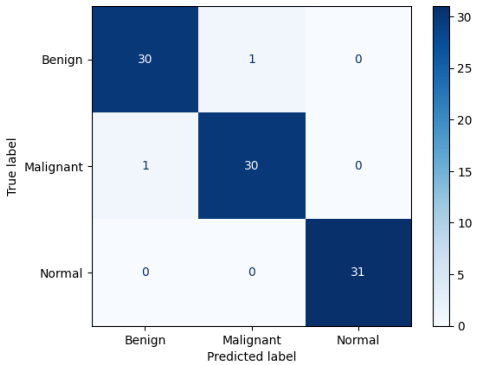

Figure 9: Confusion matrix for Clinical Data using GradientBoosting.

To further demonstrate the accuracy of this model, Figure 9 presents the confusion matrix for the Gradient-Boosting classifier, which provides a detailed breakdown of the model's performance.

Figure 10 presents a comparison of test accuracies across the evaluated models, highlighting the superior performance achieved by advanced classifiers such as XG-Boost and Support Vector Classifier (SVC).

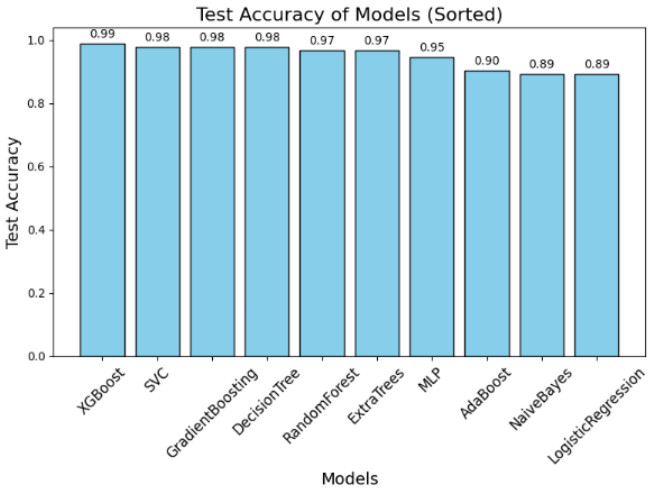

Figure 10: Accuracy Comparison.

For the imaging classification, different models were used with multiple optimizers. The performance of models paired with their optimal optimizers is summarized through Receiver Operating Characteristic (ROC) curves, F1-score comparisons, and accuracy comparisons.

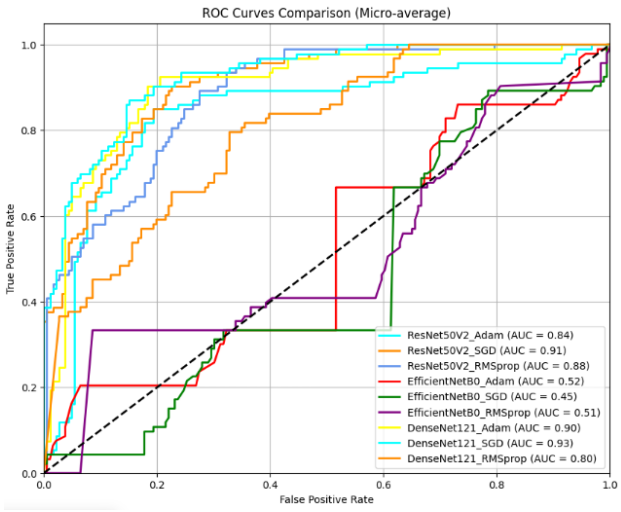

Figure 11: ROC for model with its best optimizer.

The ROC curves in Figure 11 highlight the comparative diagnostic ability of different model-optimizer combinations. The highest Area Under the Curve (AUC)

values were achieved by:

- **DenseNet121 with SGD (AUC = 0.93)**,
- **DenseNet121 with Adam (AUC = 0.90)**,
- **ResNet50V2 with SGD (AUC = 0.91)**.

These results underline the effectiveness of pairing **DenseNet121** with the SGD optimizer, as its combination demonstrated superior ability in distinguishing true positive rates from false positives.

The F1-score results, illustrated in Figure 13, show that the DenseNet121 architecture with SGD and Adam optimizers delivered the highest scores. This indicates and confirms their ability to balance precision and recall, making them ideal for handling class imbalances in medical imaging datasets.

The accuracy metrics presented in Figure 12 also rank DenseNet121 with SGD as the top performer, followed by DenseNet121 with Adam and ResNet50V2 with SGD, further reinforcing the robustness of DenseNet121 in learning complex image features efficiently.

The synergy between the DenseNet121 architecture and its optimal optimizers highlights the strengths of dense feature propagation and gradient management. The architecture's ability to reuse features through densely connected layers ensures efficient learning of image patterns, crucial for tasks like medical imaging diagnosis. The SGD optimizer, known for its simplicity and stability, and the Adam optimizer, with its adaptive learning rate and momentum capabilities, both enhance the model's convergence speed and generalization capability. These combinations provide a robust framework for accurate classification in breast lesion imaging tasks.

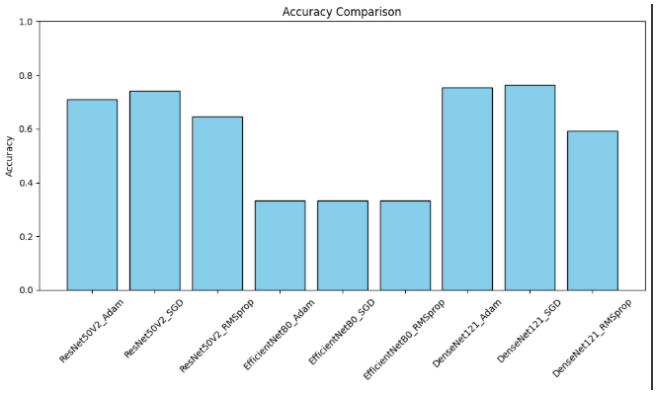

Figure 12: Accuracy Comparison

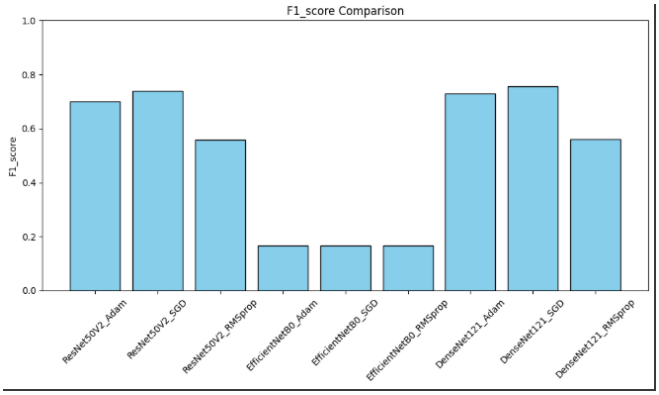

Figure 13: F1 score comparison.

The results obtained from evaluating various deep learning models with different optimizers reveal important insights into model performance. Among the configurations tested, DenseNet121 paired with the SGD optimizer achieved the highest AUC score of 0.93, indicating strong generalization capabilities on the imaging dataset. In contrast, EfficientNetB0 trained with SGD yielded the lowest performance, with an AUC score of only 0.45, suggesting limited convergence under this configuration. Meanwhile, ResNet50V2 combined with the Adam optimizer demonstrated competitive results, achieving a solid AUC score of 0.84, making it a reliable choice for further multimodal fusion experiments.

This comparison was conducted over 50 epochs for all models and optimizers. Once the best-performing model, DenseNet121 with SGD, was identified, it was further analyzed to assess its robustness and generalization capabilities.

Figure 14 presents the confusion matrix for DenseNet121 with SGD, showcasing its performance across three classes: Benign, Malignant, and Normal. The model accurately classified:

- **22 Benign samples** correctly classified, with 7 misclassified as Malignant and 2 as Normal.
- **19 Malignant samples** correctly classified, with 11 misclassified as Benign and 1 as Normal.
- **31 Normal samples** perfectly classified, achieving no misclassification for this category.

These results emphasize the model's robust performance in identifying Normal cases while showing room for improvement in distinguishing between Benign and Malignant samples. The misclassification rates suggest the necessity for enhanced feature extraction and improved discrimination in challenging categories. The overall accuracy achieved by the model stands at approximately **73.5%**.

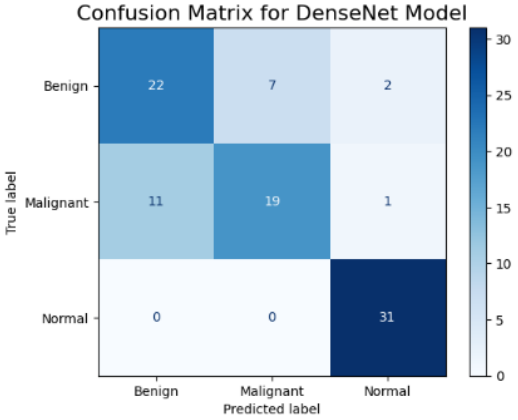

Figure 14: Confusion Matrix for DenseNet before the multimodal fusion.

V. MULTIMODAL FUSION TECHNIQUES

To address the limitations posed by single-modality input and enhance the predictive capability of the proposed model, we employed three multimodal fusion techniques: early fusion, intermediate fusion, and late fusion. These techniques aim to integrate information from multiple data modalities, thereby enriching the feature representation and improving the overall performance.

A. *Early Fusion*

Early fusion involves combining the features from different modalities at the input level. In this approach, we concatenated the raw features extracted from both modalities to form a unified input representation. This combined feature set was used to train an **ExtraTrees** classifier, which demonstrated competitive performance compared to other classifiers.

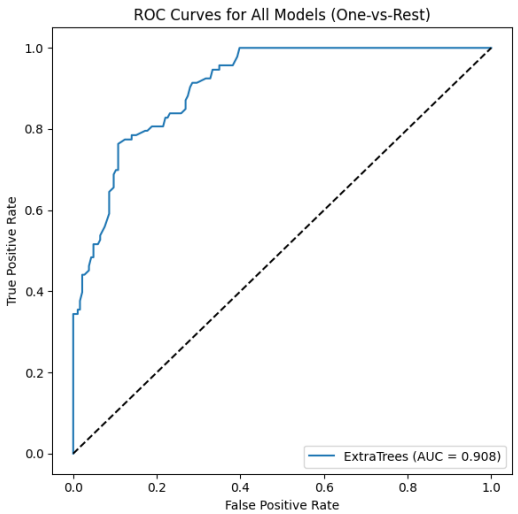

Figure 15: ROC Curve for Extratrees after early fusion.

As seen from the ROC curve presented in Fig. 15, ExtraTrees achieved an AUC of **0.908** showcasing its ability to effectively distinguish between classes. By enabling the model to learn from the integration of features, this method leverages complementary information across modalities to enhance predictive accuracy. The application of early fusion with ExtraTrees outperformed certain single-modality approaches. Fig. 16 shows the confusion matrix for Extratrees classifier after early fusion.

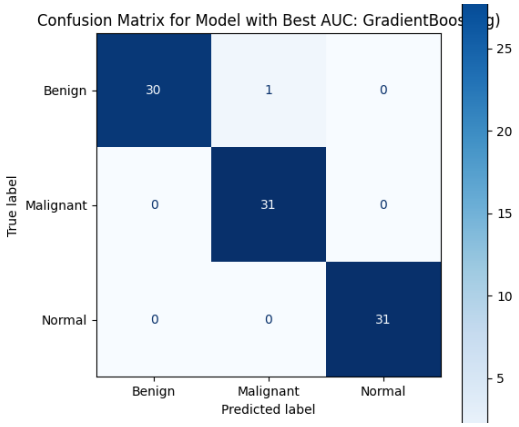

Figure 16: Confusion Matrix for Extratrees after early fusion.

B. *Intermediate Fusion*

Intermediate fusion integrates features at a latent representation level. In this approach, separate CNN models were trained for each modality to extract high-level features, and the embeddings from these models were concatenated to form a unified input representation for a Logistic Regression classifier.

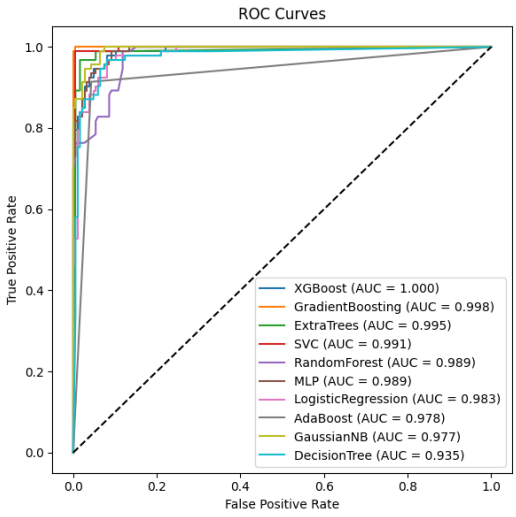

Figure 17: ROC Curves for all models after intermediate fusion.

This method allows the model to capture intrinsic patterns from each modality while enabling meaningful interactions at an abstract level, resulting in a richer feature representation. It led to a significant performance improvement. The XGBoost classifier, using intermediate fusion, achieved an impressive AUC of 1.0. The ROC curve presented in Fig. 17 demonstrates the model's high discriminatory power, and the confusion matrix presented in Fig. 18 reveals robust classification performance with only minimal misclassifications. These results underscore the effectiveness of intermediate fusion, which leverages high-level feature integration to outperform traditional single-modality and early-fusion methods.

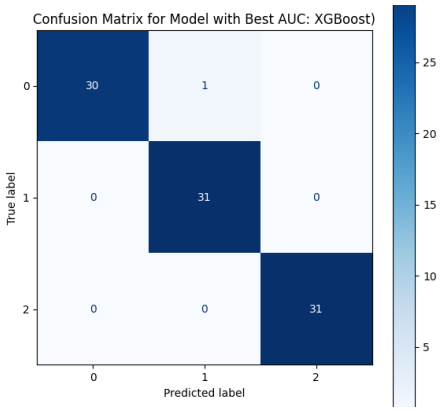

Figure 18: Confusion Matrix for XGBoost after intermediate fusion.

C. Late Fusion

Late fusion achieves remarkable results with Gradient Boosting, as indicated by an AUC of 0.99 on the ROC curve presented in Fig 19, showcasing its strong discriminatory power.

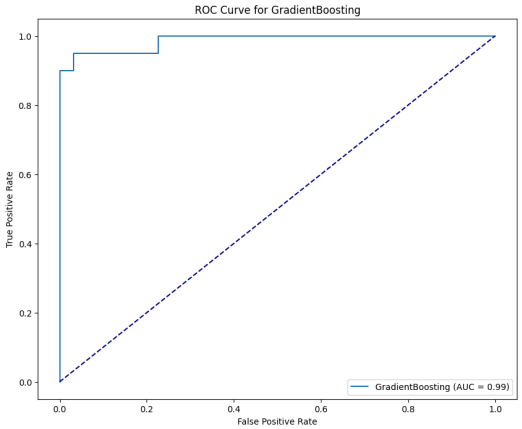

Figure 19: ROC Curve for GradientBoosting after late fusion.

The corresponding confusion matrix for Gradient Boosting illustrates excellent classification accuracy, with 30 true negatives and 19 true positives, and only one false positive and one false negative. These results collectively emphasize the effectiveness of the fusion approaches, with late fusion emerging as a robust method for leveraging decision-level integration.

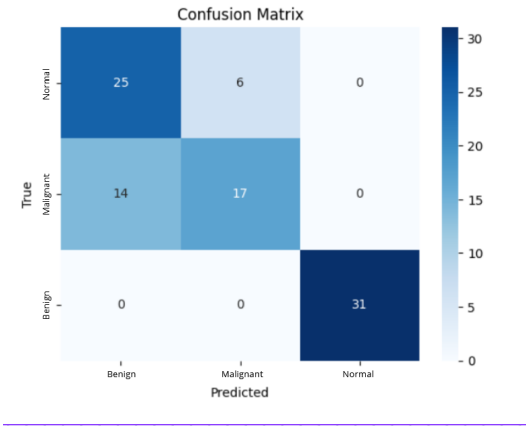

Figure 20: Confusion Matrix for GradientBoosting after late fusion.

D. Fusion Techniques Comparison

Table 4 summarizes the performance of the best models across different fusion techniques utilized in our study. Early Fusion achieved the highest accuracy (**0.96**) and AUC (**0.99**) using the **AdaBoost** classifier, demonstrating its effectiveness in leveraging combined raw feature sets. Intermediate Fusion, which integrates high-level features extracted by separate CNN models for imaging and clinical data, achieved an accuracy of **0.92** and an AUC of **0.99** using **XGBoost**. Late Fusion, which aggregates the outputs of multiple models, also achieved a high accuracy of **0.96** and an AUC of **0.99**. These results highlight the complementary nature of the fusion approaches, with Early and Late Fusion offering comparable performance, while Intermediate Fusion provided robust results through feature-level interactions.

Table 4: Comparison of best-performing models across different fusion strategies based on test accuracy and AUC.

Fusion Technique	Best Model	Accuracy	AUC
Early Fusion	AdaBoost	0.96	0.99
Intermediate Fusion	XGBoost	0.92	0.99
Late Fusion	Aggregation	0.96	0.99

E. Analysis

The prognosis of breast lesions has been significantly enhanced by integrating imaging techniques with clinical data and machine learning algorithms. Among the models evaluated, the XGBoost classifier emerged as the most effective, achieving the highest **AUC of 0.99** in both intermediate and late fusion approaches. This underscores the critical role of structured clinical data, including patient-specific, histology, and demographic information, in improving predictive accuracy.Its densely connected layers enhance gradient flow and feature propagation, making it effective in capturing critical lesion characteristics.

A comprehensive diagnostic framework was developed by combining imaging-based models such as DenseNet121 with Adam, ResNet50V2 with Adam, and EfficientNetB0 with Adam. Among these, DenseNet121 with Adam stood out for its ability to efficiently extract and reuse features.

The Adam optimizer, used in these top-performing models, further improved training efficiency by adapting learning rates and handling sparse gradients effectively. This optimization was particularly beneficial for high-dimensional imaging data, accelerating convergence and enhancing the accuracy of lesion detection.

To further optimize predictive performance, we explored three multimodal fusion techniques: **early fusion**, **intermediate fusion**, and **late fusion**.

- **Early Fusion**: By combining raw features from both imaging and clinical data at the input level, this method achieved an **accuracy of 0.96** and an **AUC of 0.99** with **AdaBoost** as the best-performing model. Early fusion highlights the potential of leveraging complementary modalities even in their raw form.

- **Intermediate Fusion**: This technique integrates high-level features extracted by separate CNN models for imaging data and clinical data, followed by their concatenation and classification using XGBoost. It significantly improved results, reaching an **accuracy of 0.92** and an **AUC of 0.99**, demonstrating the benefits of abstract feature-level interaction.

- **Late Fusion**: Outputs from modality-specific classifiers were aggregated using a voting mechanism with XGBoost. This decision-level integration achieved the best results, reaching an **accuracy of 0.92** and an **AUC of 0.99**, highlighting the strength of combining modality-specific decisions while preserving their unique advantages.

By combining these multimodal fusion techniques with imaging-based models and clinical data classifiers, an integrative diagnostic framework was developed, capable of delivering superior predictive accuracy. This synergy between clinical and imaging data enhances decision-making, expedites diagnosis, and provides precise patient classification for personalized breast cancer treatment. Supporting precision medicine initiatives, this comprehensive approach equips clinicians with robust tools for managing breast cancer and improving patient outcomes.

VI. COMPARISON WITH EXISTING STUDIES

A. Comparative Analysis with Prior Work

Recent advancements in breast cancer diagnosis using artificial intelligence have shown promising results across various imaging modalities and learning paradigms. Table 5 presents a comparative analysis between our proposed multimodal fusion framework and several notable studies from the literature, focusing on classification performance in terms of AUC.

Table 5: Comparison of AUC performances between prior works and our proposed multimodal fusion models.

Study	Method	AUC
HGD + SVM [23]	Mammogram, traditional ML	0.78
AlexNet [23]	Mammogram, CNN transfer learning	0.91
Faster R-CNN CAD [19]	Mammogram, object detection	0.95
All-conv CNN [18]	Mammogram, image-level labels	0.91–0.98
CAMELYON16 top models [16]	Histopathology (WSI), deep CNN	>0.95
CEIMVEN (EfficientNet) [21]	Ultrasound, enhanced CNN	High accuracy
Early Fusion (Ours)	Image + Clinical, AdaBoost	**0.99**
Intermediate Fusion (Ours)	Image + Clinical, XGBoost	**0.99**
Late Fusion (Ours)	Image + Clinical, Aggregation	**0.99**

As shown, traditional methods such as HGD combined with SVM yielded limited performance (AUC = 0.78), while CNN-based architectures like AlexNet and Faster R-CNN improved diagnostic accuracy, achieving AUCs between 0.91 and 0.95. More recent developments, including all-convolutional models and multi-task frameworks, have demonstrated excellent performance in specific imaging domains, with AUCs reaching up to 0.98.

In contrast, our approach, which integrates imaging data with structured clinical variables through early,

intermediate, and late fusion mechanisms, consistently achieved superior results with AUC values of **0.99** across multiple fusion strategies. These results underscore the effectiveness of combining heterogeneous data sources and leveraging ensemble learning methods to improve diagnostic precision. The ability to generalize across data modalities further supports the clinical relevance of our framework for accurate and personalized breast lesion classification.

B. Implications of Findings

The diagnosis and prognosis of breast lesions can be greatly enhanced by integrating machine learning models for both imaging and clinical data. Subtle patterns in breast lesions can be found by integrating image analysis with clinical information like demographics and medical history. This improves risk stratification and early diagnosis. These models can help radiologists and oncologists in clinical practice by facilitating personalised treatment planning, enhancing diagnostic procedures, and offering decision support. More accurate patient stratification is made possible by this method, which eventually improves patient outcomes for breast cancer treatment by enabling customised interventions.

C. Limitations

The first limitation of this study is the relatively small dataset size, comprising approximately 200 patient entries. Such a limited sample may constrain the robustness and generalizability of the developed models. Additionally, several records contained missing values, necessitating data imputation and preprocessing prior to model training. While these steps were essential for maintaining dataset usability, they may inadvertently introduce biases or inaccuracies. Consequently, despite the models' capacity to manage incomplete data, the overall performance and reliability of the system could still be affected by the volume and completeness of the available information.

D. Future Work

Several directions can be pursued in future research to enhance the performance, reliability, and clinical applicability of the proposed model. First, the current study was based on a relatively modest dataset of approximately 200 patient entries. Expanding the dataset with more diverse and representative patient records would significantly improve the model's generalizability and robustness.

Second, although preprocessing techniques were used to manage missing data, more advanced imputation strategies or improved data collection protocols could be explored to minimize the potential impact of incomplete records on model performance.

In addition, integrating other data modalities, such as genetic information or additional clinical metricsalongside imaging data may further improve predictive accuracy. Employing advanced multimodal fusion techniques could provide a more comprehensive representation of the patient's condition and lead to more reliable prognostic models.

Model refinement also presents an important opportunity for improvement. Future work could involve tuning hyperparameters more effectively, testing alternative architectures, and implementing more sophisticated optimization strategies to enhance overall performance.

Finally, validating the model in real-world clinical environments will be critical. Prospective studies and practical deployment in clinical settings will help evaluate the model's ability to support healthcare professionals with meaningful, actionable insights.

VII. CONCLUSION

This study explored the integration of clinical data and medical imaging using machine learning models to predict breast lesion outcomes. The results demonstrated strong predictive performance, with high AUC values and validation accuracy observed across several deep learning architectures, notably DenseNet121, ResNet50V2, and EfficientNetB0, all trained with the Adam optimizer. These findings underscore the potential of deep learning models to support and enhance breast cancer diagnosis.

Predictive performance was further improved by incorporating structured clinical variables, including demographic, histological, and patient-specific information. Classifiers such as multilayer perceptrons (MLPs) and ensemble methods, particularly XGBoost and Random Forest, demonstrated substantial gains in performance. The integration of clinical data provided a more comprehensive understanding of patient status, contributing to a more holistic and accurate diagnostic framework. This multimodal approach highlights the promise of personalized medicine by enabling faster and more informed clinical decision-making.

Despite these encouraging results, the study faced certain limitations. The dataset was relatively small, containing only 200 samples, and required substantial preprocessing to address missing values. Future research will aim to improve model robustness and generalizability by expanding the dataset, incorporating additional data modalities, and refining multimodal fusion strategies.

REFERENCES

[1] World Health Organization, "Breast cancer." https://www.who.int/news-room/fact-sheets/detail/breast-cancer, 2021.

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023.

[3] W. A. Berg *et al.*, "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer," *JAMA*, vol. 299, no. 18, pp. 2151–2163, 2008.

[4] E. B. Mendelson, "Evaluation of the breast with ultrasound," *Radiologic Clinics of North America*, vol. 40, no. 3, pp. 485–506, 2002.

[5] J. G. Elmore *et al.*, "Variability in interpretation of mammograms," *New England Journal of Medicine*, vol. 331, no. 22, pp. 1493–1499, 1994.

[6] M. L. Giger *et al.*, "Computer-aided diagnosis in breast imaging: Cadx and cad," *Academic Radiology*, vol. 15, no. 4, pp. 452–462, 2008.

[7] C. J. D'Orsi, E. A. Sickles, E. B. Mendelson, E. A. Morris, *et al.*, *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 5th ed., 2013.

[8] M. H. Yap *et al.*, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.

[9] Q. Cao *et al.*, "Deep learning-based computer-aided diagnosis systems for breast cancer: a survey," *Artificial Intelligence in Medicine*, vol. 132, p. 102360, 2022.

[10] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[11] A. Yala *et al.*, "Deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, 2019.

[12] J. Wang *et al.*, "Multi-modal deep learning for breast cancer diagnosis by integrating ultrasound and mammographic images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 2735–2744, 2020.

[13] Q. Dou *et al.*, "Multimodal breast cancer diagnosis by fusing qualitative and quantitative features from ultrasound and mammograms," *Medical Image Analysis*, vol. 70, p. 102006, 2021.

[14] Y. Zhang *et al.*, "Multi-modal deep learning model for breast cancer prognosis prediction using both imaging and non-imaging data," *Frontiers in Oncology*, vol. 12, p. 876543, 2022.

[15] Y. Liu *et al.*, "Integrating clinical and imaging data for improved breast cancer diagnosis using deep learning," *Journal of Biomedical Informatics*, vol. 108, p. 103518, 2020.

[16] B. E. Bejnordi, M. Veta, P. J. van Diest, *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, 2017.

[17] C. Aumente-Maestro, J. Díez, and B. Remeseiro, "A multi-task framework for breast cancer segmentation and classification in ultrasound imaging," *Medical Image Analysis*, Mar. 2025.

[18] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Radiology*, Aug. 2019.

[19] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Scientific Reports*, Mar. 2018.

[20] Y. Zhang *et al.*, "Deep learning-based multi-modal data integration enhancing breast cancer prognosis prediction," *Frontiers in Oncology*, vol. 13, p. 11190375, 2023.

[21] S. Banerjee and M. K. H. Monir, "Ceimven: An approach of cutting-edge implementation of modified versions of efficientnet (v1-v2) architecture for breast cancer detection and classification from ultrasound images," *arXiv preprint*, 2023.

[22] A. Pawłowska, A. Ćwierz Pieńkowska, A. Domalik, D. Jaguś, P. Kasprzak, R. Matkowski, Fura, A. Nowicki, and N. Zolek, "A curated benchmark dataset for ultrasound based breast lesion analysis (breast-lesions-usg)." https://doi.org/10.7937/9WKK-Q141, 2024. Version 1 [dataset].

[23] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," *Journal of Healthcare Engineering*, Jan. 2017.