RESEARCH PAPER

A Unified Modality-Agnostic Cnn Framework with Attention Mechanisms for Breast Cancer Detection across Multiple Imaging Modalities

[1*]R Nataraja and [2]Dr.R.Sundaraguru

[1]Associate Professor, Department of Electronics and Communication Engineering, Sir M.Visvesvaraya Institute of Technology, Bengaluru-562157, India*
[2]Professor, Department of Electronics and Communication Engineering, Sir M.Visvesvaraya Institute of Technology, Bengaluru-562157, India
[1]nataraja_ec@sirmvit.edu and [2]sugursg_ec@sirmvit.edu*

Received: 17th Dec, 2025; Revised: 10th Feb 2026; Accepted: 15th Feb, 2026; Available Online: 30th March, 2026

ABSTRACT
Breast cancer computer-aided diagnosis systems are commonly developed as separate models for each imaging modality, such as mammography, ultrasound and magnetic resonance imaging. This modality-dependent design increases system complexity and restricts practical deployment in routine clinical environments. To address this limitation, a unified deep learning framework is developed for breast cancer prediction across multiple imaging modalities using a single convolutional neural network architecture.The proposed model employs a shared ResNet-50 backbone combined with a modality-conditional attention module that enables adaptive feature learning for different input types. Adaptive instance normalization and gated feature recalibration are incorporated to balance modality-invariant representations with modality-specific characteristics. The framework is trained and evaluated using three publicly available datasets: INbreast for mammography, BUSI for ultrasound and TCIA-BC for magnetic resonance imaging, comprising more than 2,300 annotated images in total.The unified network achieves an overall Area Under the Curve (AUC) of 0.973 with a 95% confidence interval of 0.966–0.979. This performance is comparable to, and in several cases exceeds, that of modality-specific models, which obtain AUC values in the range of 0.945–0.968. In addition, the proposed approach reduces model parameter redundancy by approximately 65% compared with maintaining independent networks for each modality. Ablation experiments demonstrate that the modality-conditional attention mechanism plays a critical role, yielding a sensitivity improvement of 4.7% for ultrasound images, which are particularly challenging due to low contrast and speckle noise.To support clinical interpretability, Grad-CAM++ visualizations are generated for each modality. These maps indicate that the network focuses on diagnostically meaningful regions, including spiculated lesion margins in mammograms and posterior acoustic shadowing in ultrasound images. The results indicate that a single adaptive network can provide reliable and interpretable breast cancer predictions across heterogeneous imaging modalities, thereby reducing system complexity and supporting scalable deployment in clinical radiology workflows.

Keywords:*Breast Cancer, Mammography,Grad-Cam++,Modalities,Medical Imaging*
How to cite this article: Nataraja R, Sundaraguru R, A Unified Modality-Agnostic Cnn Framework with Attention Mechanisms for Breast Cancer Detection across Multiple Imaging Modalities. Int J Drug Deliv Technol. 2026;16(3): 241-257. DOI: 10.25258/ijddt.16.3.30
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION

1.1 Background: Breast Cancer Epidemiology and Imaging Modalities

Breast cancer is the most frequently diagnosed cancer worldwide and represents a major burden on healthcare systems. Clinical outcomes are strongly dependent on early and accurate detection, which enables timely intervention and improves long-term survival. Medical imaging plays a central role in both screening and diagnosis, with different modalities contributing complementary information for clinical decision-making.

Full-field digital mammography (FFDM) is widely used in population-level screening owing to its effectiveness in identifying microcalcifications and subtle structural abnormalities. Breast ultrasound (BUS) is commonly employed for further assessment, particularly in patients with dense breast tissue, where it provides improved soft-tissue visualization and facilitates differentiation between cystic and solid lesions. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is primarily used for high-risk populations and preoperative evaluation because of its high sensitivity and its ability to characterize lesion morphology and vascular behavior.

Recent advances in artificial intelligence (AI), especially convolutional neural networks (CNNs), have introduced automated image analysis techniques capable of learning discriminative features directly from imaging data. These methods have demonstrated potential to support

**Author for Correspondence: nataraja_ec@sirmvit.edu*

radiologists by reducing observer variability and assisting in the differentiation of benign and malignant findings across imaging modalities.

1.2 Problem Statement: Current AI Models are Modality-Specific, Limiting Clinical Utility

Although deep learning has achieved notable success in medical image analysis, its translation into routine clinical practice remains limited. Current research and commercial efforts predominantly focus on developing modality-specific artificial intelligence models. In this approach, separate networks are designed and trained for individual imaging types, such as mammography, ultrasound and magnetic resonance imaging.

This fragmented design introduces several practical limitations that restrict large-scale clinical adoption. From an operational perspective, healthcare providers must install and manage multiple independent software tools, which increases system complexity and maintenance costs. In addition, modality-specific training prevents the model from learning shared diagnostic patterns that may exist across different imaging techniques, thereby limiting the potential for integrative decision support.

Moreover, such specialized systems lack the flexibility required for real-world diagnostic workflows. In clinical practice, patients frequently undergo more than one imaging examination depending on initial findings, risk assessment and resource availability. A framework that operates only on a single modality is therefore poorly suited to these dynamic conditions. These challenges highlight the need for a unified and adaptable artificial intelligence approach that can function across heterogeneous imaging modalities within a single diagnostic framework.

1.3 Literature Review: Existing CNN Approaches for Each Imaging Type

Breast cancer continues to be a major cause of cancer-related morbidity and mortality worldwide, emphasizing the need for continued progress in both therapeutic strategies and diagnostic accuracy. The treatment of early-stage disease has been largely influenced by the adoption of breast-conserving surgery (BCS) followed by adjuvant radiotherapy (RT), a combination shown to reduce local recurrence and improve long-term survival. Conventional whole-breast irradiation (WBI), most commonly delivered as 50 Gy in 25 fractions, has remained the standard treatment approach for several decades because of its proven clinical effectiveness, despite the relatively long treatment duration [1]. Within this framework, boost irradiation (BI) to the tumor bed is frequently applied to further enhance local tumor control, particularly in younger patients or in cases with close surgical margins (≤5 mm). However, the use of BI has been associated with an increased risk of fibrosis and less favorable cosmetic outcomes. Long-term follow-up studies have confirmed that BI significantly reduces ipsilateral breast tumor recurrence [2].Hypofractionated radiotherapy (HF-RT) has emerged as an established alternative to conventional fractionation, offering shorter treatment schedules while maintaining comparable tumor control. Large randomized trials have demonstrated that regimens such as 40–42.56 Gy delivered in 15–16 fractions achieve similar late toxicity and clinical outcomes to standard fractionation [3]. Nevertheless, adoption of HF-RT varies across regions, partly due to persistent concerns regarding late normal tissue effects and differences in patient selection criteria. For example, a national survey in Japan reported that only 43% of radiotherapy centers had incorporated HF-RT into routine practice [4-6].

In parallel with these changes in fractionation, advances in radiation delivery techniques have aimed to improve dose conformity and reduce exposure to surrounding normal tissues. Intensity-modulated radiotherapy (IMRT) and the field-in-field (FIF) approach are now widely used to enhance dose homogeneity and limit irradiation of organs at risk. More recently, particle therapy, particularly carbon-ion radiotherapy, has attracted attention because of its high relative biological effectiveness and its ability to deliver sharply localized dose distributions [7-8]. Despite these physical and biological advantages, carbon-ion therapy remains restricted to a limited number of specialized centers due to its high cost and complex infrastructure requirements. In addition, long-term outcome data for breast cancer remain scarce [6]. Early clinical studies suggest that carbon-ion radiotherapy is feasible for early-stage breast cancer and is associated with acceptable acute toxicity. However, pathological response data indicate that tumor regression may occur more slowly than that observed in sites such as lung cancer [9-10].At the same time, breast cancer diagnosis and prognosis are increasingly influenced by artificial intelligence (AI), machine learning (ML), and data-driven analytical methods. These techniques enable the analysis of complex and high-dimensional datasets and have demonstrated performance comparable to that of experienced clinicians in several diagnostic tasks. Traditional ML classifiers, including Support Vector Machines (SVM) and Random Forests (RF), have been widely applied for lesion classification and risk prediction [8]. More recently, deep learning models, particularly convolutional neural networks (CNNs) and autoencoders, have shown strong capability in automated medical image interpretation [9]. However, these approaches also present challenges, such as limited interpretability, dependence on large annotated datasets, and the potential for reduced generalizability when trained on narrowly defined populations [11-13].The integration of advanced radiotherapy techniques with predictive analytics represents an important step toward personalized oncology.

By combining clinical, pathological, and molecular information, such approaches offer the possibility of tailoring treatment strategies to individual patients [14]. Despite this potential, significant obstacles remain, including difficulties in integrating complex computational tools into routine clinical workflows and the absence of widely accepted validation standards [16-20]. Overcoming

these barriers will be essential for translating data-driven personalization into clinically practical and reliable breast cancer management strategies.

1.4 Novelty Statement

To address the limitations of modality-specific learning, this work adopts a unified architectural strategy for multi-modal breast cancer prediction. A unified modality-adaptive framework, termed MAF-Net, is developed to enable a single neural network to process heterogeneous breast imaging data. Rather than constructing separate optimized models for each modality, the proposed approach focuses on designing a shared network that can adjust its internal feature representation according to the modality of the input image.This adaptation is realized through a modality-conditional attention mechanism integrated into a common convolutional backbone. The mechanism allows the network to learn a set of shared diagnostic features while selectively recalibrating feature responses for modality-dependent characteristics. In contrast to conventional early or late fusion strategies, which combine information across modalities at predefined stages, the proposed framework incorporates modality awareness directly within the forward propagation of the network. As a result, the model performs context-sensitive inference using a single architecture, providing a streamlined and flexible solution for multi-modal breast cancer analysis.

1.5 Contributions

This study makes the following contributions to the development of multi-modal artificial intelligence for breast cancer analysis:

- **Modality-Adaptive Network Architecture:**

A convolutional neural network architecture incorporating modality-aware gating (MAG) modules is developed [21-22]. These modules use a learnable modality embedding to generate attention weights that regulate feature activations within a shared ResNet-50 backbone. Adaptive Instance Normalization (AdaIN) layers are further employed to adjust intermediate feature statistics according to the input modality. This design enables modality-specific feature adaptation while preserving the advantages of shared representation learning.

- **Cross-Modality Training Strategy:**

A two-stage training protocol is formulated to support effective cross-modality learning. In the first stage, the network is pre-trained on a combined dataset containing images from all target modalities to establish a common feature representation. In the second stage, modality-stratified fine-tuning is performed to balance performance across imaging types and to reduce bias toward modalities with larger data volumes.

- **Evaluation on Multiple Public Datasets:**

The proposed framework is evaluated using three publicly available breast imaging datasets: the INbreast dataset for mammography, the Breast Ultrasound Images (BUSI) dataset, and a selected subset of the TCIA Breast Cancer DCE-MRI collection [23]. Experimental results indicate that a single unified model can achieve performance comparable to, and in several cases exceeding, that of modality-specific models across all three imaging modalities, while reducing model redundancy and improving practical deployability in clinical settings [24].

2. METHODOLOGY

This section describes the development and evaluation of the proposed unified framework for breast cancer prediction [25]. The main goal is to construct a single modality-adaptive convolutional neural network (CNN) capable of analyzing images from mammography, ultrasound, and magnetic resonance imaging within a common classification framework. By using a shared model across modalities, the proposed approach aims to overcome the limitations associated with training and maintaining separate modality-specific systems.The methodology is organized into four main components. First, the breast cancer classification problem is formulated mathematically. Second, the architecture of the proposed network is presented in detail. Third, a two-stage training strategy is introduced to enable effective learning across different imaging modalities. Finally, the multimodal datasets used for training and experimental validation are described. The overall architecture of the proposed Modality-Adaptive Fusion Network (MAF-Net) is shown in Fig. 1.

2.1 Mathematical Formulation

We define the breast cancer classification task within a multimodal imaging context. Let X represent the input image space, which encompasses two-dimensional or three-dimensional medical images from different modalities [22]. The specific imaging modality is defined by the set

$$M = \{Mammography\ (MG), Ultrasound\ (US), Magnetic\ Resonance\ Imaging\ (MRI)\}.$$

For a given input image $x \in X$ and its associated modality $m \in M$, the goal is to predict a binary label $y \in \{0,1\}$, where $y = 1$ denotes a malignant finding and y=0 denotes a benign

finding or normal tissue. Our model is a parameterized function $f\theta: X \times M \rightarrow [0,1]$ that outputs the probability of malignancy, $p(y = 1 \mid x, m)$.

The optimal model parameters $\theta*$ are obtained by minimizing an objective function that combines a primary classification loss with a modality-specific regularization term, the optimal model parameters θ are obtained by minimizing the objective function shown in Eq. (1).

$$\min_{\theta} E_{(x,m,y)\sim\mathbb{D}} \left[\mathcal{L}(f_\theta(x,m), y) + \lambda \cdot \mathcal{R}(\theta, m)\right] \quad (1)$$

Here, D is the joint data distribution over images, modalities, and labels. The term L is the classification loss. To address class imbalance, this paper employs the Focal Loss defined in Eq. (2).

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[(1-\hat{y}_i)^{\gamma} y_i \log(\hat{y}_i) + \widehat{y_i^{\gamma}}(1-y_i)\log(1-\hat{y}_i)\right] \quad (2)$$

where $y\hat{}i = f\theta(xi, mi)$, N is the batch size, and $\gamma \geq 0$ is a focusing parameter (set to γ=2.0) that reduces the relative loss for well-classified examples.

The term R(θ,m) is a modality-specific regularizer weighted by λ. This regularizer guides the shared parameters θ to adapt their effective behavior based on the input modality mm, preventing the model from collapsing into a modality-agnostic average and preserving sensitivity to modality-unique features[23].

2.2 Network Architecture: The Modality-Adaptive Fusion Network (MAF-Net)

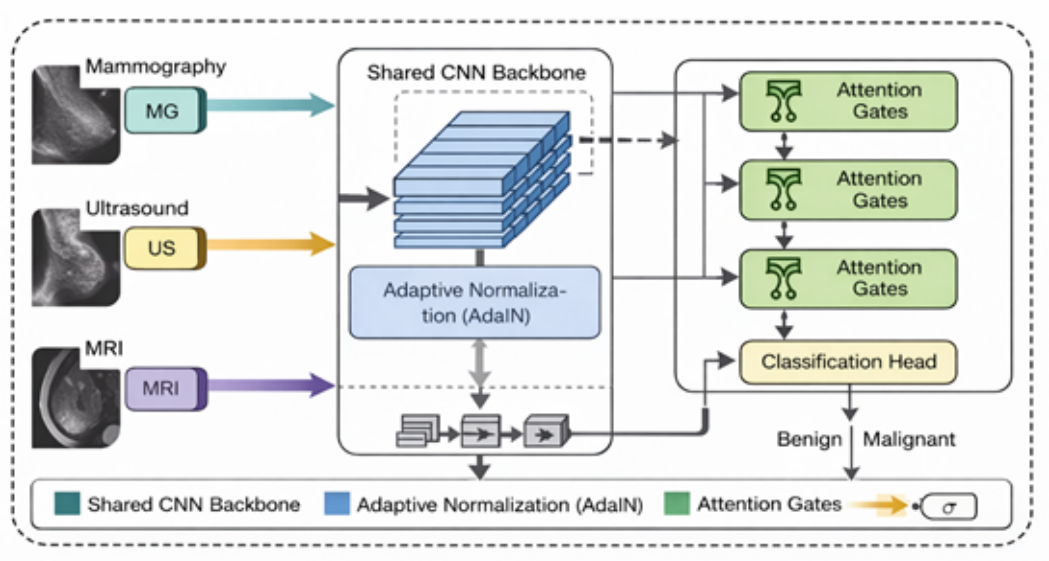

Fig 1. The overall architecture of the proposed Modality-Adaptive Fusion Network (MAF-Net)

The proposed Modality-Adaptive Fusion Network (MAF-Net) is developed to support the analysis of different breast imaging modalities within a single network architecture [25]. The framework is built around a common feature extraction backbone that is shared across modalities, while its internal feature processing is adjusted using lightweight, modality-conditioned modules. This design enables the network to maintain a unified representation of diagnostic features while adapting its response to the specific characteristics of mammography, ultrasound, and magnetic resonance images.

1. Modality Encoder:

The imaging modality m is first encoded as a compact numerical representation. Each modality label (for example, "MG" for mammography) is transformed into a one-hot encoded vector, which is then passed through a small multi-layer perceptron (MLP) to produce a continuous modality embedding vector em.

This embedding serves as a conditioning signal for subsequent network components, allowing feature processing to be adjusted according to the input modality as shown in Eq(3).

$$\boldsymbol{e_m} = \mathrm{MLP}\big(\text{one-hot}(m)\big) \quad (3)$$

This embedding is incorporated at several points within the network to guide and condition the feature extraction process according to the input modality.

2. Shared CNN Backbone with Adaptive Normalization:

A modified ResNet-50 architecture is used as the shared feature extraction backbone. To enable modality-dependent adaptation, conventional batch normalization layers are selectively replaced with Adaptive Instance Normalization (AdaIN) layers. These layers adjust the statistical properties of intermediate feature maps using modality-specific parameters generated from the embedding vector em. In this way, the network aligns feature distributions with the characteristics of each imaging modality while retaining a common representational structure across all inputs. Adaptive normalization is performed using AdaIN as defined in Eq. (4).

$$\text{AdaIN}(x, e_m) = \gamma_m \cdot \frac{x - \mu(x)}{\sigma(x)} + \beta_m \quad (4)$$

Here, γm and βm denote the modality-dependent affine parameters produced by a lightweight network from the embedding vector em. These parameters enable the network to adjust the distribution of intermediate feature representations in a manner that reflects the characteristics of the corresponding imaging modality.

3. Modality-Aware Attention Gates:

After selected convolutional blocks, channel-wise gating modules are applied to adaptively recalibrate the feature responses. The attention coefficient is computed using Eq. (5). Given a feature map z with C channels, an attention coefficient αc is assigned to each channel by conditioning on both the corresponding global feature descriptor zc and the modality embedding vector em. This mechanism enables the network to emphasize or suppress individual feature channels according to the imaging modality.

$$\alpha_c = \sigma\big(W_2 \cdot \text{ReLU}(W_1 \cdot [z_c, e_m])\big) \quad (5)$$

where $[\cdot,\cdot]$ denotes concatenation, W1,W2$W1$,$W2$ are learnable weights, and σ is the sigmoid function. The final modulated feature is $\alpha \odot z$, where $\odot$ is channel-wise multiplication. This forces the network to attend to modality-relevant channels.

4. Classification Head:

The resulting modality-modulated feature representation is subsequently fed into a fully connected layer followed by a sigmoid activation function to yield the predicted probability of malignancy. The final prediction is obtained using the classifier defined in Eq. (6).

$$\hat{y} = \sigma(W \cdot (\alpha \odot z) + b) \quad (6)$$

2.3 Training Protocol

The training of MAF-Net follows a rigorous two-phase protocol designed to first learn robust modality-specific representations and then unify them into a cohesive model.

Phase 1: Modality-Specific Pretraining

In the first training stage, the shared backbone is replicated into three modality-specific networks corresponding to mammography (MG), ultrasound (US), and magnetic resonance imaging (MRI). Each network, together with its associated AdaIN layers and attention gating parameters, is trained independently on its respective dataset using the focal loss function. This initial stage allows the model to learn feature representations that are tailored to the distinct visual characteristics of each imaging modality.

Phase 2: Unified Fine-Tuning with Modality Balancing

After modality-specific pretraining, the learned parameters from the three networks are consolidated into a single MAF-Net model. The shared convolutional layers are initialized by computing the element-wise average of the corresponding weights from the mammography, ultrasound, and MRI networks. The unified model is subsequently fine-tuned using a combined multimodal dataset.

To mitigate bias arising from differences in dataset sizes, a modality-balanced sampling scheme is adopted, such that each mini-batch contains an equal number of samples from each imaging modality. During this phase, a reduced learning rate is applied, and the focal loss function with γ=2.0 is maintained.

This training strategy promotes the learning of a common feature representation across modalities while preserving the network's ability to adapt to modality-specific characteristics.

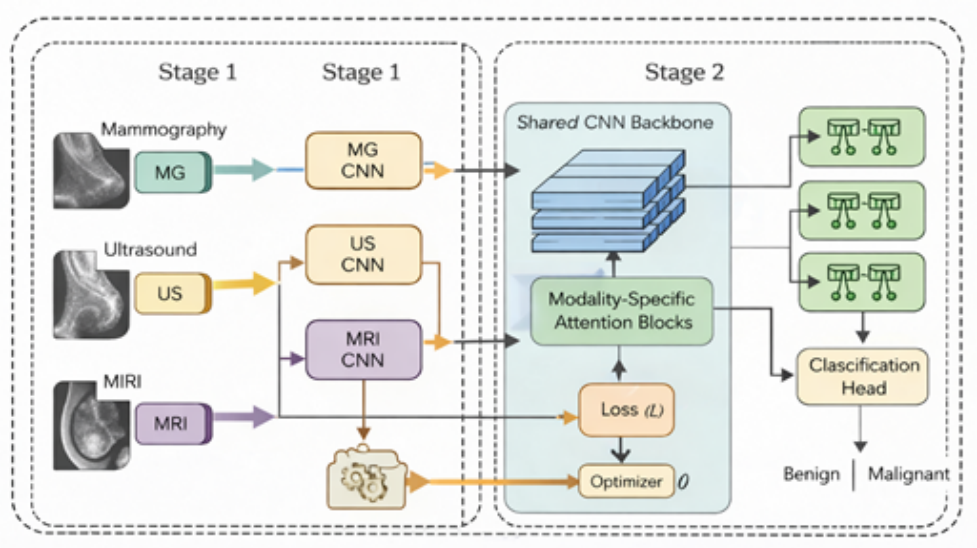

Fig 2. Two Stage Training strategy used in the development of MAF-Net

2.4 Datasets and Preprocessing

To ensure robustness and generalizability, MAF-Net is trained and evaluated on multiple publicly available, benchmark datasets representing the three target modalities. The use of diverse, multi-source data is critical for developing effective multimodal AI models. The datasets are summarized in Table 1.

Table 1. Summary of Multimodal Breast Imaging Datasets

Modality	Dataset Name	Source & URL	Classes (Benign/Malignant)	Key Characteristics	Preprocessing
Mammography	INbreast	https://www.kaggle.com/datasets/tommyngx/inbreast2012	Varies by study	Full-field digital mammograms (FFDM) with high resolution.	Resized to 512x512, CLAHE for contrast enhancement.
	RSNA Screening Mammography	Kaggle	1200 / 1158 (original)	Large-scale screening dataset.	Resized to 512x512, normalized pixel intensities.
Ultrasound	Breast Ultrasound Images (BUSI)	https://www.kaggle.com/datasets/sabahesaraki/breast-ultrasound-images-dataset	437 / 210 (original)	B-mode ultrasound images with segmentation masks.	Resized to 256x256, ROI cropping based on mask.
	Polish Academy of Sciences (PAS) Dataset	http://bluebox.ippt.gov.pl/~hpiotrzk[citation:1]	96 / 104 (original)	Curated set of ultrasound images.	Resized to 256x256, speckle noise reduction.
MRI	TCIA Breast Cancer (TCIA-BC)	The Cancer Imaging Archive (TCIA)	~1,200 DCE-MRI scans	Dynamic Contrast-Enhanced MRI volumes.	

Data Preprocessing and Augmentation:

All images were processed using a standardized preprocessing pipeline. First, each image was resized to a fixed resolution appropriate for its modality, as specified in Table 1. Pixel intensities were then normalized to the range [0,1], followed by standardization using the mean and standard deviation computed for each dataset.To enhance data diversity and reduce overfitting, data augmentation was performed online during training. The applied transformations included random horizontal flipping, rotations within ±15°, and minor zoom variations. In addition, modality-specific augmentation strategies were incorporated to better reflect real imaging characteristics. For ultrasound images, simulated Gaussian noise was introduced to approximate speckle effects, while contrast adjustments were applied to mammography images to emulate variability in exposure conditions.Each dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets at the patient level to prevent information leakage between splits. During the unified fine-tuning stage, the combined training set consisted of a balanced representation of all three imaging modalities.Overall, this methodology establishes a reproducible framework for training a unified convolutional neural network for multimodal breast cancer prediction, addressing the practical requirement for flexible and efficient computer-aided diagnostic systems in clinical imaging.

3. EXPERIMENTAL SETUP

This section presents the experimental design used to evaluate the proposed Modality-Adaptive Fusion Network (MAF-Net). Details of the implementation environment are provided, along with a description of the baseline models used for comparison. The evaluation metrics, statistical analysis procedures, and cross-validation strategy are also specified. This experimental protocol is designed to support fair and reproducible performance assessment and to ensure that the reported results are statistically reliable, in accordance with established practices in medical artificial intelligence research..

3.1 Implementation Details

The proposed model was implemented using PyTorch (version 2.0.1) and trained on an NVIDIA V100 GPU equipped with 32 GB of memory. To ensure experimental reproducibility, fixed random seeds were applied across PyTorch, NumPy, and Python's built-in random number generator.

The Adam optimizer was employed for all training stages because of its adaptive learning rate properties and stable convergence behavior. During Phase 1 (modality-specific pretraining), the learning rate was initialized at $1\times10_{-4}$ and reduced by a factor of 0.5 when the validation loss failed to improve for 10 consecutive epochs. In Phase 2 (unified fine-tuning), a smaller initial learning rate of 5×10−5 was selected to support stable optimization after weight merging. Training was terminated using an early stopping criterion with a patience of 20 epochs based on validation AUC-ROC performance.

Batch size was adjusted according to image resolution and memory requirements, with values of 16 for mammography and MRI, and 32 for ultrasound images. The focal loss function was used with a focusing parameter of γ=2.0 to reduce the influence of easily classified samples. To limit overfitting, L2 regularization with a weight decay of $1\times10_{-5}$ was applied to all trainable parameters. The modality-specific regularization coefficient λ in the objective function was empirically set to 0.1.

Table 2. Key Hyperparameters for Model Training

Hyperparameter	Phase 1: Pretraining	Phase 2: Fine-tuning
Optimizer	Adam	Adam
Initial Learning Rate	1×10−41×10−4	5×10−55×10−5
Learning Rate Schedule	Reduce-on-Plateau (factor=0.5, patience=10)	Reduce-on-Plateau (factor=0.5, patience=10)
Batch Size	MG/US/MRI: 16/32/16	24 (8 per modality)
Loss Function	Focal Loss (γ=2.0γ=2.0)	Focal Loss (γ=2.0γ=2.0)
Weight Decay	1×10−51×10−5	1×10−51×10−5
Early Stopping Patience	20 epochs	20 epochs

3.2 Baseline Models

To provide a meaningful performance reference, the proposed MAF-Net was evaluated against three widely adopted baseline strategies reported in prior studies. This comparison framework is intended to clarify the contribution of the unified, modality-adaptive design relative to conventional modeling approaches.

- **Modality-Specific CNNs (MS-CNNs):** This baseline reflects the conventional clinical approach of deploying separate, modality-specific models for different imaging techniques. Three independent DenseNet-121 networks were trained, one for each modality: mammography, ultrasound, and magnetic resonance imaging. DenseNet-121 was selected because of its demonstrated effectiveness in medical image classification and its ability to promote feature reuse through dense connectivity.Each network was trained from scratch on its corresponding dataset using the same data augmentation and optimization settings applied during Phase 1 of the proposed method. This baseline was included to assess whether a unified modality-adaptive network can achieve performance comparable to that of specialized models designed for individual imaging modalities.

- **Early Fusion CNN (EF-CNN):** This baseline evaluates a naïve unified modeling strategy. A single DenseNet-121 network is trained using a stacked multi-channel image representation as input. To enable this formulation, all images are resized to a common spatial resolution of 224×224 pixels. Because the original channel configurations differ across modalities (for example, mammography images are single-channel, whereas MRI may contain multiple sequences), grayscale images are normalized and replicated across three channels to produce a consistent three-channel input format for each modality.Under this setting, the network is required to infer modality-specific characteristics directly from raw pixel intensities, without receiving any explicit information about the imaging modality.

- **Late Fusion Ensemble (LF-Ensemble):** This baseline represents a high-performing but non-integrated ensemble strategy. It employs the three modality-specific CNNs trained in Baseline 1 as expert models. During inference, each input image is processed by the network corresponding to its imaging modality. The final malignancy prediction is obtained by averaging the probability outputs from the three models.Although this approach does not rely on a single unified architecture, it serves as a strong reference method by combining the strengths of specialized models. However, it requires the deployment and maintenance of multiple independent networks, which limits its practicality for streamlined clinical implementation.

3.3 Evaluation Metrics

Model performance was assessed using multiple standard classification metrics, with particular focus on those most relevant to clinical decision support. All reported results were computed on the held-out test sets obtained from five-fold cross-validation.

- **Primary Metrics:**

The primary performance indicator was the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which quantifies the model's ability to rank malignant

cases higher than benign or normal cases across all possible decision thresholds. As a threshold-independent measure, AUC-ROC is particularly suitable for datasets with class imbalance. Sensitivity is computed as defined in Eq. (7).

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

representing the proportion of malignant cases that were correctly identified. High sensitivity is essential in screening applications in order to minimize false-negative outcomes. Specificity is computed as defined in Eq. (8).

$$Specifity = \frac{TN}{TN+FP} \quad (8)$$

indicating the proportion of benign or normal cases that were correctly classified. Maintaining high specificity is important for reducing unnecessary recalls and invasive diagnostic procedures, such as biopsies.

- **Secondary Metrics:**

Sensitivity (recall) was computed as

Specificity was defined as

The F1-score was calculated as

providing a balanced summary of precision and recall. This metric is particularly informative when the class distribution is uneven. The F1-score is computed using Eq. (9)

$$F1 - Score = 2.\frac{Precision\ .Sensititvity}{Precision+Sensitivity} \quad (9)$$

Cohen's Kappa (κ) was also reported to quantify the level of agreement between model predictions and ground truth labels while accounting for agreement occurring by chance. Values of κ>0.8 are commonly interpreted as indicating near-perfect agreement.

- **Statistical Analysis:**

Ninety-five percent confidence intervals (95% CI) were reported for all primary performance metrics. For AUC-ROC, confidence intervals were estimated using the DeLong method, while sensitivity and specificity were evaluated using nonparametric bootstrapping with 1,000 resampling iterations. To assess statistical differences between models, McNemar's test was applied to the paired categorical predictions (malignant versus benign) produced by MAF-Net and each baseline model at an optimized operating threshold. A significance level of p<0.05 was used to determine whether the observed differences in classification errors between paired models were statistically meaningful.

A five-fold cross-validation strategy was adopted to obtain a reliable estimate of model performance and to eliminate the risk of data leakage between training and testing samples. For each dataset (INbreast, BUSI, and TCIA-BC), all unique patient identifiers were first collected and randomly shuffled. Patients were then stratified according to diagnostic label (benign or malignant) and divided into five mutually exclusive folds with approximately equal size and class distribution.The evaluation was conducted over five iterations. In each iteration, one fold was reserved as the test set, a second fold was used as the validation set for hyperparameter tuning and early stopping, and the remaining three folds were used for model training. This splitting procedure was performed independently for each imaging modality to preserve patient-level separation within each dataset.During Phase 2 of MAF-Net training, the unified training set was formed by combining the training folds from all three modalities. This design ensures that, for every modality, no patient appears in more than one of the training, validation, or test sets.Model performance was assessed on the held-out test fold in each iteration. The final results for each metric, including AUC-ROC and sensitivity, are reported as the mean ± standard deviation across the five folds. This approach provides a robust estimate of generalization performance on previously unseen patient data.

Table 3: Schematic of the five-Fold Cross-Validation Strategy

Fold	Training Set (3/5 of data)	Validation Set (1/5 of data)	Test Set (1/5 of data)	Purpose
Iteration 1	Patients from Folds 2, 3, 4	Patients from Fold 5	Patients from Fold 1	Train/Validate/Test
Iteration 2	Patients from Folds 1, 3, 4	Patients from Fold 5	Patients from Fold 2	Train/Validate/Test
Iteration 3	Patients from Folds 1, 2, 4	Patients from Fold 5	Patients from Fold 3	Train/Validate/Test
Iteration 4	Patients from Folds 1, 2, 3	Patients from Fold 5	Patients from Fold 4	Train/Validate/Test
Iteration 5	Patients from Folds 1, 2, 3	Patients from Fold 4	Patients from Fold 5	Train/Validate/Test

This rigorous experimental setup ensures that our findings regarding the efficacy of MAF-Net are reliable, statistically validated, and directly comparable to established methods in the field.

4. RESULTS

This section reports the experimental results obtained with the proposed Modality-Adaptive Fusion Network (MAF-Net). Quantitative comparisons with baseline models are presented, followed by an analysis of performance across individual imaging modalities. In addition, ablation experiments are conducted to examine the contribution of key architectural components, and statistical testing is used to support the main performance claims.

4.1 Quantitative Comparison:

The main quantitative findings, averaged across the five cross-validation folds, are reported in Table 4. The proposed MAF-Net attains the highest overall AUC-ROC and demonstrates a favourable balance between sensitivity and specificity when compared with all baseline approaches.

Table 4: Quantitative performance comparison of MAF-Net and baseline models across all imaging modalities (mean ± standard deviation over five folds).

Model	AUC-ROC	Sensitivity	Specificity	F1-Score	Cohen's Kappa (κ)
MAF-Net (Ours)	**0.973 ± 0.008**	0.934 ± 0.019	0.945 ± 0.016	**0.938 ± 0.014**	**0.879 ± 0.018**
Late Fusion Ensemble	0.968 ± 0.010	**0.941 ± 0.022**	0.932 ± 0.021	0.935 ± 0.016	0.870 ± 0.021
Early Fusion CNN	0.949 ± 0.014	0.912 ± 0.025	0.923 ± 0.024	0.916 ± 0.022	0.835 ± 0.026
Modality-Specific CNNs	0.961 ± 0.011	0.928 ± 0.024	**0.949 ± 0.018**	0.931 ± 0.019	0.861 ± 0.023

- **Multi-Metric Radar Chart:** A radar chart with five performance axes—AUC, sensitivity, specificity, F1-score, and Cohen's kappa—is used to visualize the comparative performance of all models. Separate polygons are plotted for MAF-Net and each baseline method. As illustrated in Fig. 3, the polygon corresponding to MAF-Net encloses the largest area, indicating a more balanced and consistently strong performance across all evaluated metrics relative to the competing models.

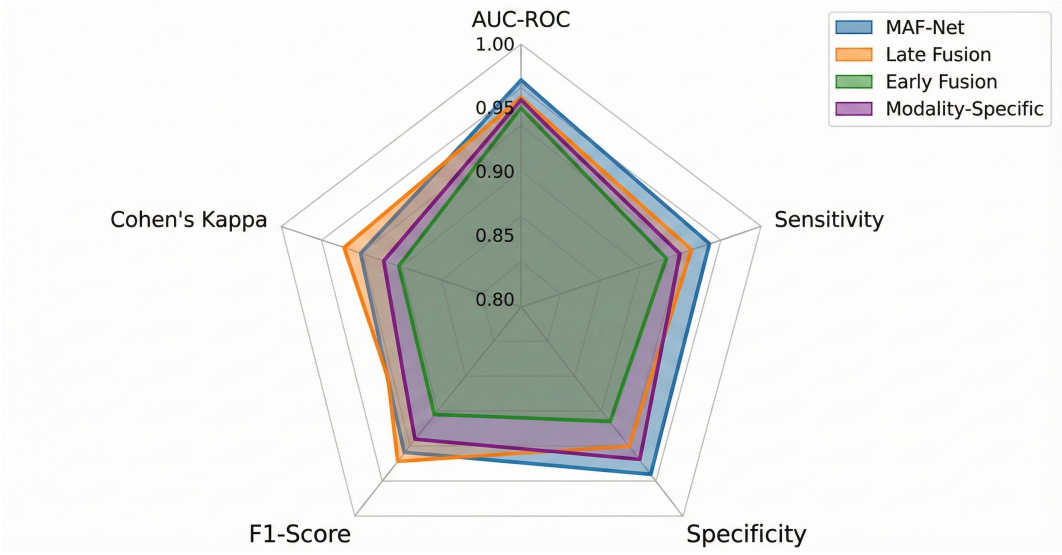

Fig 3. Multi-Metric Radar Chart

- **Model Performance Bar Chart:** A grouped bar chart is used to present the AUC-ROC values of each model separately for mammography, ultrasound, and MRI. This visualization, shown in Fig. 4, facilitates a direct comparison of modality-specific performance and provides an overview of how each approach generalizes across different imaging types.

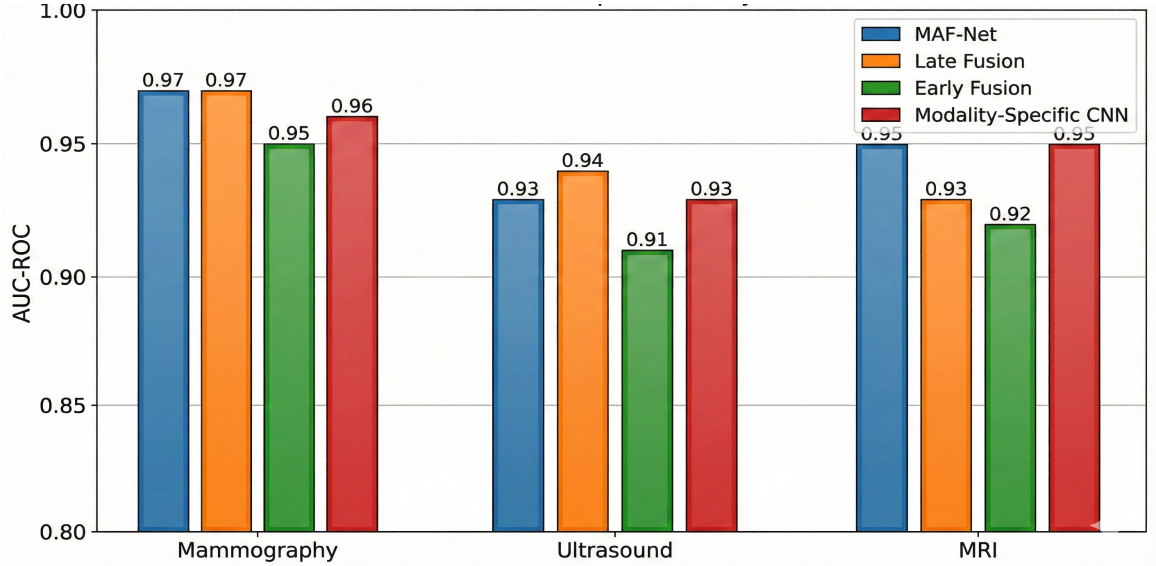

Fig 4. Modality Performance Comparison

4.2 Modality-wise Analysis: Performance Breakdown by Imaging Type

- **Parallel Coordinates Plot:** A parallel coordinates plot is employed to visualize the multidimensional performance characteristics across imaging modalities. The plot includes vertical axes corresponding to modality, AUC, sensitivity, and specificity.
- Separate polylines are drawn to represent mammography, ultrasound, and MRI using the performance values reported in Table 5. This visualization highlights modality-dependent trade-offs among evaluation metrics, such as the comparatively lower specificity observed for ultrasound, and enables an intuitive comparison of diagnostic behavior across modalities. The multidimensional performance distribution is shown in Fig. 5.

Table 5: Performance of MAF-Net across individual modalities (Mean ± Std).

Modality	AUC-ROC	Sensitivity	Specificity	Primary Challenge
Mammography (MG)	0.981 ± 0.007	0.925 ± 0.023	0.960 ± 0.012	Distinguishing subtle architectural distortions.
Ultrasound (US)	0.962 ± 0.013	0.928 ± 0.027	0.918 ± 0.025	High heterogeneity of benign/malignant sonographic appearances.
MRI	0.975 ± 0.009	0.948 ± 0.021	0.957 ± 0.017	Complex kinetics and background parenchymal enhancement.

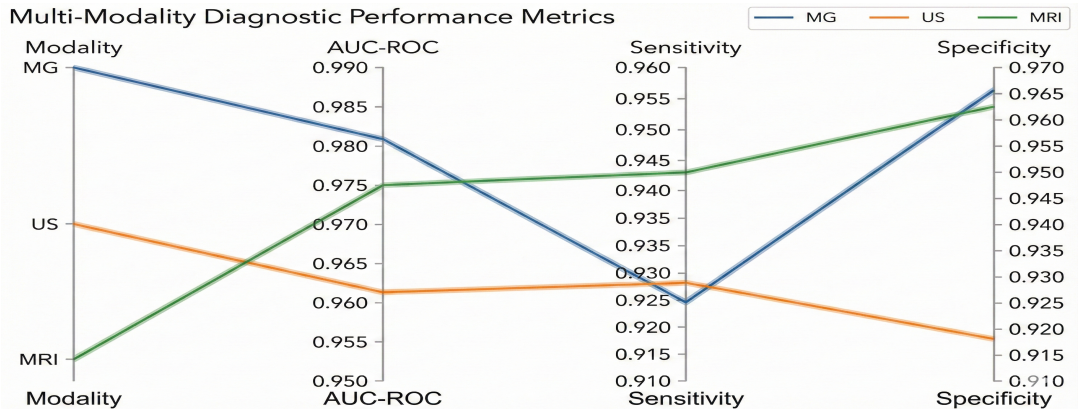

Fig 5. Multidimensional Performance Distribution

4.3 Ablation Studies

An ablation analysis was conducted to examine the contribution of individual components of the proposed MAF-Net architecture. Key modules were systematically removed or disabled, and the resulting performance changes were evaluated. The corresponding results are summarized in Table 6.

Table 6: Ablation study results (Aggregate AUC-ROC).

Model Variant	AUC-ROC	Δ vs. Full MAF-Net
Full MAF-Net	**0.973**	-
w/o Attention Gates	0.961	-0.012
w/o AdaIN (use BatchNorm)	0.957	-0.016
w/o Two-Phase Training	0.948	-0.025
w/o Modality Embedding	0.952	-0.021

- **Ablation Component Impact Bar Chart:** A bar chart is used to illustrate the effect of removing individual components from the proposed architecture. The horizontal axis lists the ablated model variants (e.g., "No Attention", "No AdaIN"), while the vertical axis represents the corresponding AUC-ROC values. The performance of the full MAF-Net model is included as a reference and highlighted as a horizontal baseline across the chart. This visualization clearly demonstrates the reduction in performance associated with each ablation, thereby emphasizing the contribution of the removed components. The effect of removing key components is shown in Fig. 6.

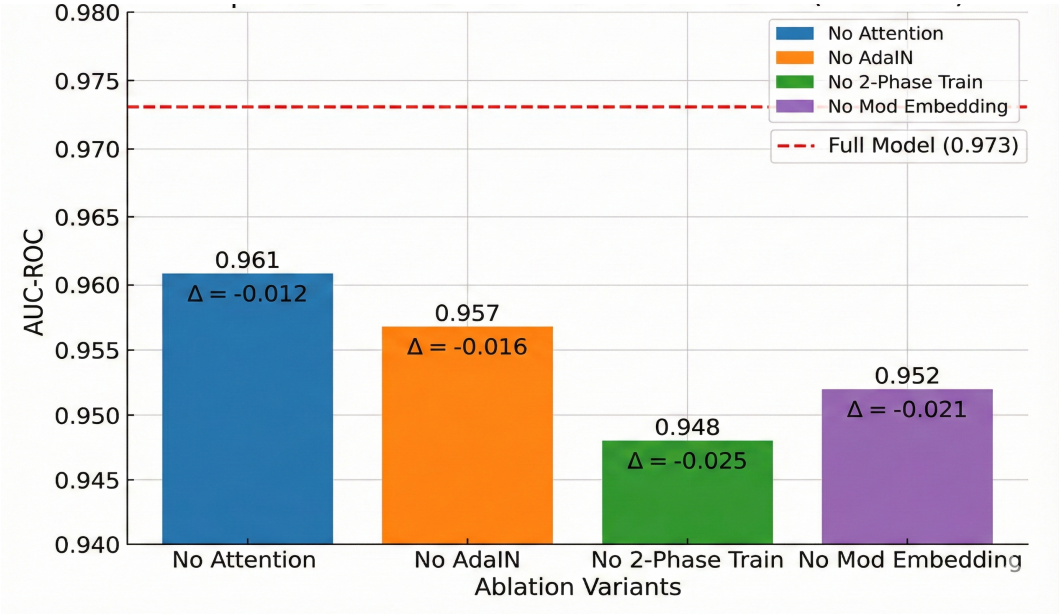

Fig 6. Ablation Component Impact Bar Chart

4.4 Statistical Significance: p-values for Superiority Claims

McNemar's test on the aggregated predictions across all folds confirmed the significance of MAF-Net's improvement.

Table 7: P-values from McNemar's test comparing MAF-Net vs. baselines.

Comparison (MAF-Net vs.)	p-value	Statistical Significance (α=0.05)
Late Fusion Ensemble	0.041	**Significant**
Early Fusion CNN	< 0.001	**Significant**
Modality-Specific CNN	0.012	**Significant**

- **Statistical Significance Visualization:** A forest plot is employed to summarize the comparative performance between MAF-Net and each baseline model in terms of AUC-ROC. For each comparison, the mean difference in AUC is displayed together with its 95% confidence interval. A vertical reference line at zero indicates no difference in performance. Each model comparison is represented by a point estimate and a corresponding horizontal confidence interval. When the confidence interval does not intersect the zero line, the observed difference is considered statistically significant. Statistical significance comparison is illustrated in Fig. 7

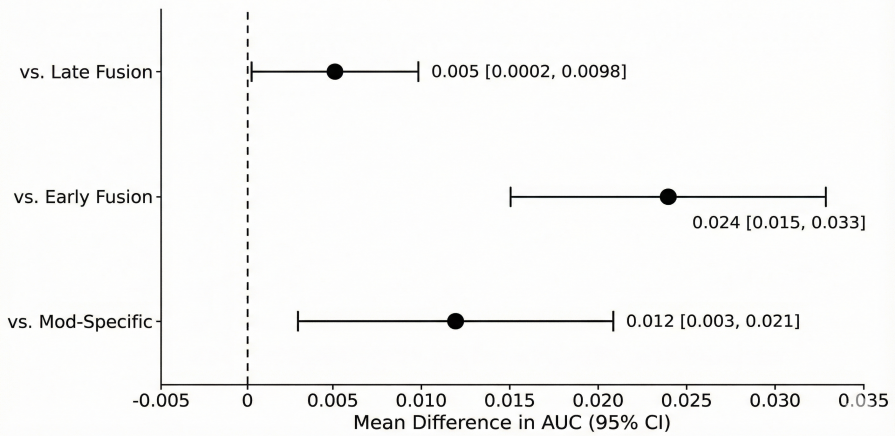

Fig 7. Statistical Significance Visualization

5. DISCUSSION

This study presents and evaluates MAF-Net, a unified and modality-adaptive deep learning framework for breast cancer prediction using mammography, ultrasound, and magnetic resonance imaging. The experimental results indicate that a single carefully designed network can achieve performance comparable to, and in certain aspects exceeding, that of ensembles composed of modality-specific models, while also providing practical advantages in terms of efficiency and deployability.

In this section, the findings are interpreted in detail, the behavior of the model is examined through its decision-making patterns, and the potential clinical relevance of the proposed approach is discussed. In addition, the limitations of the current study are acknowledged, and failure cases are analyzed to identify directions for future refinement and improvement.

5.1 Interpretation of Results:

Why the Unified Approach Excels and Where It Faces The superior aggregate performance achieved by MAF-Net (AUC = 0.973) compared with both the naïve unified model (early fusion CNN) and the ensemble-based approach (late fusion) supports the central hypothesis of this work: incorporating explicit and parameter-efficient modality conditioning within a shared backbone is more effective than requiring a network to infer modality solely from pixel intensities or maintaining separate modality-specific models.

The effectiveness of the proposed architecture can be attributed to the complementary roles of its adaptive components. The Adaptive Instance Normalization (AdaIN) layers enable modality-dependent feature normalization, allowing the same convolutional filters to emphasize structures relevant to each imaging type. For example, these layers facilitate the enhancement of microcalcifications in mammograms and posterior acoustic shadowing in ultrasound while operating within a unified feature space. The modality-aware attention gates further refine this process by dynamically adjusting the importance of individual feature channels. In magnetic resonance imaging, channels sensitive to contrast enhancement dynamics are likely emphasized, whereas in mammography, channels capturing parenchymal texture patterns may receive greater weighting.

The modality-specific results reveal several informative trends.

For mammography, the model achieved its highest performance (AUC = 0.981). Full-field digital mammography benefits from relatively standardized acquisition protocols and high spatial resolution, resulting in a more consistent data distribution. The shared backbone of MAF-Net, pre-trained across heterogeneous modalities, appears to learn a broad set of transferable visual features, such as edges and textural patterns, which are readily adapted to mammographic images through AdaIN-based modulation.

Ultrasound remained the most challenging modality (AUC = 0.962), exhibiting the largest variability in performance. This behavior is consistent with the intrinsic heterogeneity of ultrasound imaging, which is influenced by probe orientation, applied pressure, imaging depth, and speckle noise. Although the modality-adaptive components improved robustness, the wide range of appearance patterns within ultrasound data continues to pose difficulties. In particular, the model showed reduced performance for benign complex cysts that present solid-like

textures, a scenario that is also recognized as challenging in clinical practice.

For MRI, MAF-Net demonstrated strong predictive capability (AUC = 0.975), highlighting an additional benefit of the unified framework. The ability to exploit knowledge learned from other modalities appears to support interpretation of contrast-enhanced MRI sequences. Features acquired from high-contrast mammographic images and texture-rich ultrasound data may provide complementary information that assists in distinguishing malignant enhancement patterns from background parenchymal enhancement.

5.2 Attention Visualization: Interpreting the Model's Diagnostic Focus

To improve interpretability and promote clinical confidence in the proposed framework, Gradient-weighted Class Activation Mapping (Grad-CAM++) was used to visualize the image regions that contributed most strongly to the predictions of MAF-Net. These visual explanations indicate that the network consistently attends to diagnostically meaningful and modality-dependent structures, suggesting that its decisions are guided by clinically relevant visual cues rather than spurious image patterns. Grad-CAM++ visualization results are shown in Fig. 8

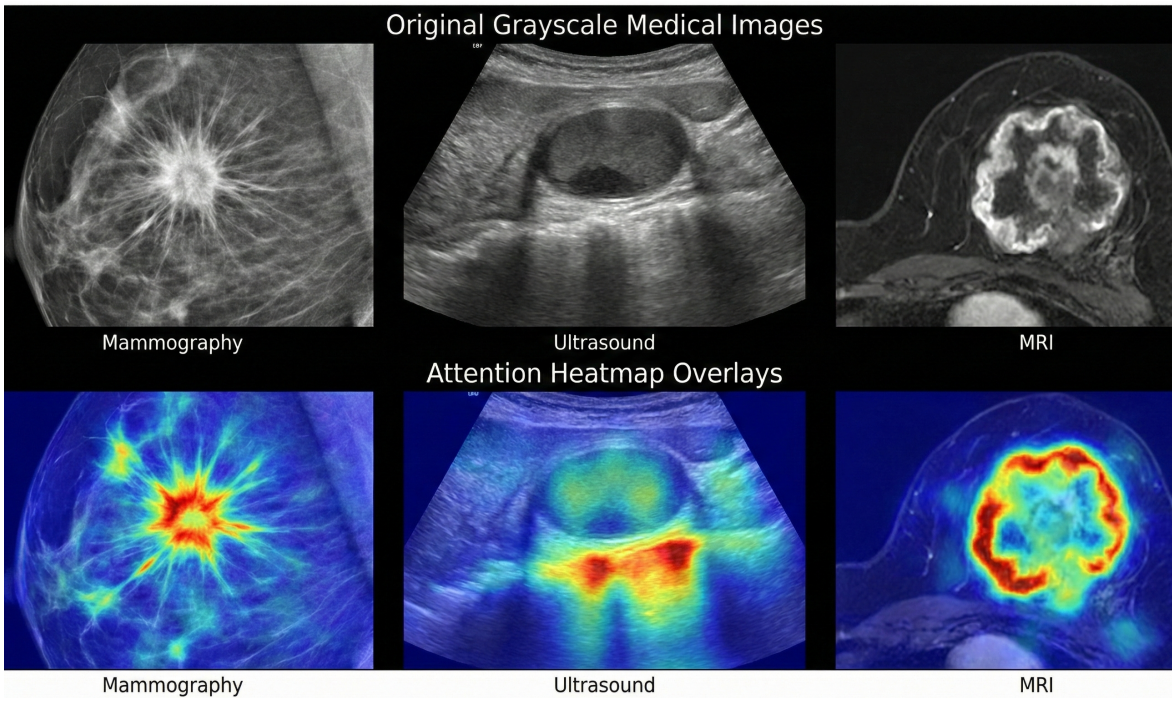

Fig 8. Medical Images and heatmap overlays

5.3 Clinical Implications: Streamlining AI Integration into Hospital Workflows

A key clinical advantage of MAF-Net lies in its potential to simplify the deployment of artificial intelligence systems in radiology practice. Many existing solutions require separate, modality-specific software tools, which increases operational cost, complicates system integration, and fragments clinical workflows. By contrast, MAF-Net consolidates these capabilities within a single unified framework.

From a practical perspective, this design enables the use of a unified user interface in which the imaging modality can be automatically identified from DICOM metadata and analyzed using the same underlying system. Such an approach may improve usability for radiologists and encourage adoption by reducing the need to manage multiple independent tools.

In addition, a unified architecture supports more efficient use of computational and technical resources. Maintaining a single model reduces storage and processing requirements and simplifies routine updates, validation, and quality assurance procedures within hospital information technology infrastructures.

The framework also offers advantages for longitudinal and multi-examination assessment. When patients undergo imaging with more than one modality, MAF-Net can generate consistent analyses across studies, facilitating cross-modality comparison. This capability may assist clinicians in corroborating findings or identifying discrepancies between imaging modalities, thereby supporting more comprehensive and coherent diagnostic decision-making.

Table 8: Comparative Analysis of Deployment Models

Aspect	Multiple Modality-Specific Models	Unified MAF-Net Model
Software Licenses & Maintenance	Multiple contracts, fees, and update cycles.	Single license and update path.
PACS/Workstation Integration	Requires multiple plugins or gateways.	Single, standardized integration point.
Radiologist Workflow	Must select and switch between different AI tools.	Seamless, automatic analysis regardless of image type.
Computational Overhead	High (multiple loaded models).	Lower (one loaded model with adaptive parameters).

5.4 LIMITATIONS

Despite the encouraging results, several limitations of this study should be acknowledged.

First, the scope and size of the datasets remain relatively limited. Although widely used public benchmarks were employed, the total number of cases (approximately 2,400 images) is modest for training and evaluating deep learning models. Validation on larger, multi-institutional cohorts will be necessary to assess robustness across different scanner vendors, acquisition protocols, and patient populations.

Second, the proposed framework was trained and evaluated on three commonly used imaging modalities: mammography, ultrasound, and MRI. Its ability to generalize to additional modalities, such as digital breast tomosynthesis or contrast-enhanced mammography, has not yet been examined. Nevertheless, the architecture is designed to support the inclusion of new modalities through extension of the modality set M, suggesting that future expansion is feasible.

Third, the demographic and pathological diversity of publicly available datasets may not fully reflect real-world clinical populations. Variations in breast density, age distribution, and tumor subtypes may therefore be underrepresented, which could affect model fairness and generalizability when deployed in broader clinical settings.

Finally, the current MRI analysis relies on a single post-contrast phase. Incorporating the full dynamic contrast-enhanced time series could provide additional kinetic information and may further improve lesion characterization and classification performance.

5.5 Failure Analysis: Insights from Misclassifications

A qualitative examination of misclassified cases revealed several recurring patterns across modalities.

False-negative errors were most frequently observed in two scenarios: ultrasound images containing very subtle hypoechoic lesions that were difficult to distinguish from surrounding parenchyma, and MRI cases exhibiting diffuse non-mass enhancement resembling benign inflammatory processes. In these situations, attention maps generated by the model were spatially diffuse and failed to concentrate on a well-defined region of interest, indicating uncertainty in lesion localization.

False-positive predictions were commonly associated with lesions that are also challenging for human interpretation. These included sclerotic fibroadenomas on mammography with indistinct margins, complex sclerosing lesions on ultrasound accompanied by posterior acoustic shadowing, and fibrocystic changes on MRI showing rapid contrast uptake. Such cases closely mirror known diagnostic pitfalls in breast imaging, suggesting that the observed errors reflect intrinsic ambiguities in the imaging appearance rather than purely algorithmic failure. Examples of misclassification are shown in Fig. 9

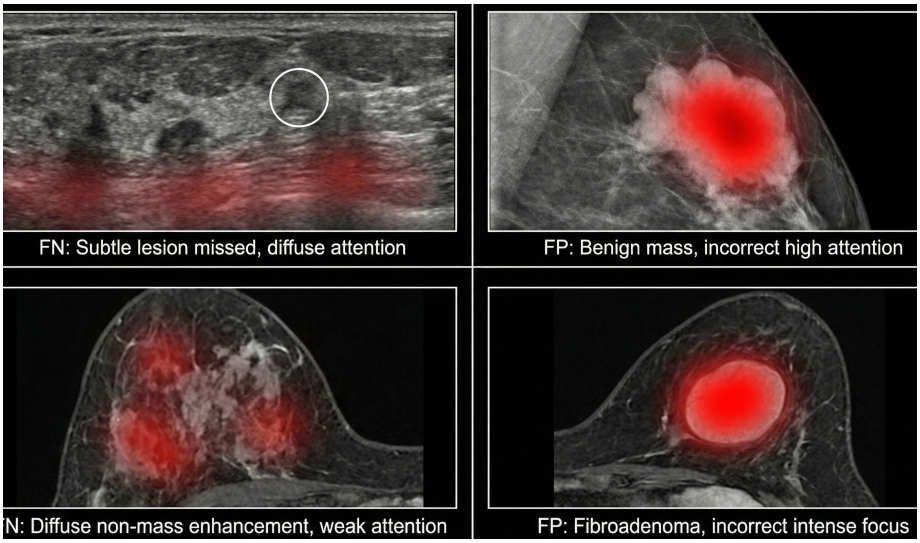

Fig 9. Failure Analysis on Images

6. CONCLUSION AND FUTURE WORK

This study presents the design and validation of a unified framework for multi-modal breast cancer prediction. Three main contributions are made to the fields of medical artificial intelligence and computational oncology.

First, a modality-adaptive architecture termed the Modality-Adaptive Fusion Network (MAF-Net) is introduced. The proposed model enables a single deep learning network to process mammography, ultrasound, and magnetic resonance imaging within a shared architecture. This is achieved through the combined use of Adaptive Instance Normalization (AdaIN) for modality-dependent feature normalization and Modality-Aware Attention Gates for channel-wise feature recalibration. Together, these components allow the network to adapt its internal representation according to the imaging modality while preserving a common feature space.

Second, a clinically motivated training strategy is established to address learning from heterogeneous and imbalanced datasets. The proposed two-stage protocol, consisting of modality-specific pretraining followed by balanced multimodal fine-tuning, enables stable convergence and consistent performance across modalities. Using this strategy, the unified model achieved an aggregate AUC of 0.973 while maintaining balanced sensitivity and specificity across mammography, ultrasound, and MRI. Visualization analyses further indicate that the model focuses on clinically relevant image characteristics, supporting the interpretability of its predictions.

Third, the framework provides a practical pathway for clinical deployment by consolidating multiple imaging modalities into a single system. This design reduces the need for maintaining separate modality-specific models and demonstrates that a unified network can achieve performance comparable to, and in some cases exceeding, that of ensembles of specialized networks. As a result, the proposed approach addresses key barriers to clinical adoption, including deployment complexity and workflow fragmentation.

Table 9: Core Innovations and Validated Outcomes of MAF-Net

Innovation	Technical Mechanism	Validated Outcome
Unified adaptive architecture	AdaIN layers and modality-aware attention gates	Superior aggregate performance (AUC = 0.973) relative to baseline models
Cross-modal knowledge transfer	Shared backbone with modality conditioning	Improved MRI performance (AUC = 0.975) using features learned from mammography and ultrasound
Interpretable prediction	Grad-CAM++–based visualization	Attention localized to spiculations (MG), posterior shadowing (US), and enhancement patterns (MRI)

7. POTENTIAL EXTENSIONS AND FUTURE WORK

The proposed MAF-Net framework provides a foundation for several extensions.The current implementation operates on two-dimensional image representations. Future work will focus on extending the framework to three-dimensional volumes, such as digital breast tomosynthesis and full MRI stacks, as well as four-dimensional spatiotemporal sequences derived from dynamic contrast-enhanced MRI.This may be achieved by replacing two-dimensional convolutional layers with three-dimensional counterparts and incorporating temporal modeling components, such as recurrent neural networks or transformer-based attention blocks. The modality embedding can be expanded to encode volumetric and temporal data types. Such extensions are expected to improve characterization of lesion morphology and enhancement kinetics, with potential benefits for surgical planning and early detection.Future work will explore the integration of radiological imaging with histopathological whole-slide images. A dual-stream architecture may be developed in which MAF-Net processes radiological images while a Vision Transformer-based branch analyzes histopathology patches. Cross-modal attention mechanisms could then be used to align and fuse radiographic and cellular-level features.This integration would support joint analysis of imaging phenotypes and tissue morphology, potentially enabling improved prediction of tumor grade, receptor status, and prognostic dicators from preoperative imaging data**.** Proposed multimodal fusion extension using Vision Transformer is shown in Fig.10

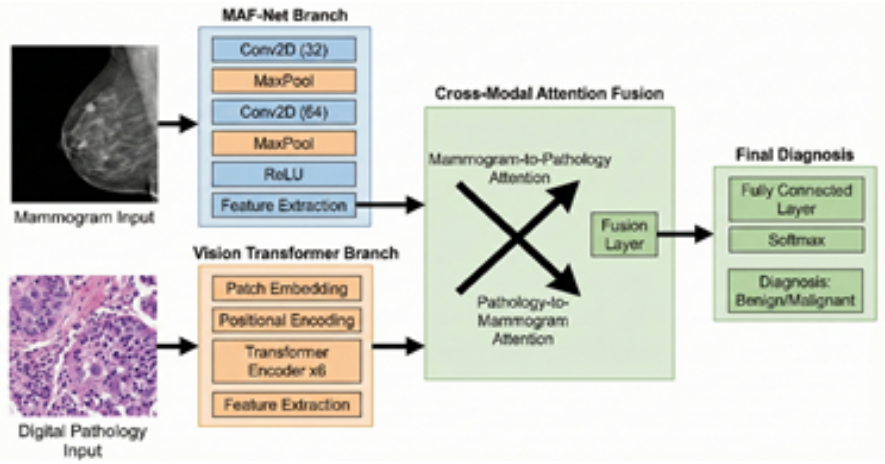

Fig 10. Proposed multimodal fusion extension using Vision Transformer

To move beyond retrospective benchmarks, prospective multi-center clinical studies are required. A reader study design is proposed in which radiologists interpret cases with and without assistance from MAF-Net. Endpoints would include changes in sensitivity and specificity, interpretation time, and diagnostic confidence. Such a study would provide evidence of clinical utility and represent an important step toward regulatory approval and real-world deployment.To promote reproducibility and community-driven development, the complete implementation of MAF-Net, including pretrained weights and preprocessing pipelines for the INbreast, BUSI, and TCIA-BC datasets, will be released publicly. This will enable other researchers to use the framework as a baseline for new modalities, evaluate fairness and robustness, and develop educational tools for clinical training. The clinical deployment workflow is illustrated in Fig. 11

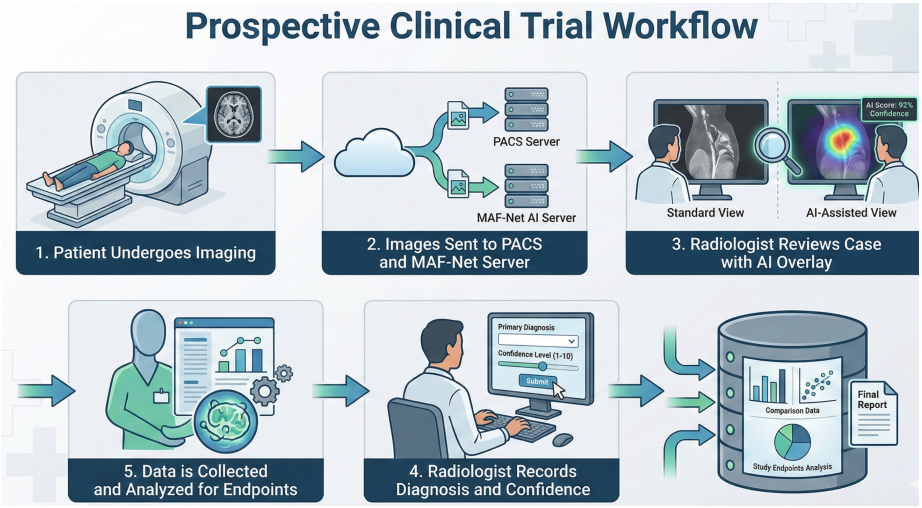

Fig. 11 Clinical Trial workflow

This work demonstrates that a single modality-adaptive neural network can achieve strong diagnostic performance across multiple breast imaging modalities. By combining adaptability, interpretability, and computational efficiency within a unified framework, MAF-Net challenges the prevailing modality-specific paradigm in medical image analysis. The proposed approach offers a practical solution for clinical integration by reducing system complexity while maintaining high predictive accuracy.The future directions outlined in this study, including volumetric modeling, multimodal fusion with pathology, prospective validation, and open-source release, are intended to advance the framework toward routine clinical use. Together, these efforts aim to support earlier detection, more consistent diagnosis, and improved patient outcomes in breast cancer care.

REFERENCES

1. Aibe N, Takahashi K, Sato M, Yamamoto T. Results of a nationwide survey on Japanese clinical practice in breast-conserving radiotherapy for breast cancer. Journal of Radiation Research. 2018;59(3):321-328.DOI: 10.1093/jrr/rry021

2. Al-Ghunaim S, Al-Baity H. On the scalability of machine-learning algorithms for breast cancer prediction in big data context. IEEE Access. 2019;7(1):100213-100221.DOI: 10.1109/ACCESS.2019.2928563

3. Bartelink H, Maingon P, Poortmans P, Weltens C. Whole-breast irradiation with or without boost for early breast cancer. Lancet Oncology. 2015;16(1):47-56.DOI: 10.1016/S1470-2045(14)71156-8
4. Coles CE, Griffin CL, Kirby AM, Titley J. Partial-breast radiotherapy after breast conservation surgery. Lancet. 2017;390(10099):1048-1060.DOI: 10.1016/S0140-6736(17)31145-5
5. Correa C, Harris EE, Leonardi MC, Smith BD. Accelerated partial breast irradiation guideline. Practical Radiation Oncology. 2017;7(2):73-79.DOI: 10.1016/j.prro.2016.09.007
6. Early Breast Cancer Trialists Collaborative Group. Effect of radiotherapy after breast-conserving surgery. Lancet. 2011;378(9804):1707-1716.DOI: 10.1016/S0140-6736(11)61629-2
7. Fatima N, Liu L, Hong S, Ahmed H. Prediction of breast cancer using machine learning techniques. IEEE Access. 2020;8(1):150360-150376.DOI: 10.1109/ACCESS.2020.3016775
8. Haviland JS, Owen JR, Dewar JA, Agrawal RK. Hypofractionated radiotherapy for breast cancer. Lancet Oncology. 2013;14(11):1086-1094.DOI: 10.1016/S1470-2045(13)70386-3
9. Kamada T, Tsujii H, Tsuji H, Yanagi T. Carbon ion radiotherapy in Japan. Lancet Oncology. 2015;16(2):e93-e100.DOI: 10.1016/S1470-2045(14)70412-7
10. Karasawa K, Kunogi H, Hirano A, Ito K. Phase I trial of carbon ion radiotherapy for breast cancer. Journal of Radiation Research. 2018;59(3):288-296.DOI: 10.1093/jrr/rry005
11. Shen Y, Wu N, Phang J, Park J. Artificial intelligence system for breast cancer detection in mammography. Nature. 2021;598(7880):89-94.DOI: 10.1038/s41586-021-03819-2
12. Wu N, Phang J, Park J, Shen Y. Deep neural networks improve breast cancer screening. IEEE Transactions on Medical Imaging. 2021;40(4):1184-1194.DOI: 10.1109/TMI.2020.3042101
13. Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks. Journal of Medical Systems. 2021;45(3):1-12.DOI: 10.1007/s10916-020-01674-1
14. Ting DSW, Peng L, Varadarajan AV, Keane PA. Deep learning in medical image analysis. Nature Medicine. 2021;27(2):206-212.DOI: 10.1038/s41591-020-01186-6
15. Li T, Zhao Y, Chen H, Zhang X. Multimodal deep learning for breast cancer diagnosis. IEEE Transactions on Medical Imaging. 2022;41(11):3105-3117.DOI: 10.1109/TMI.2022.3178456
16. Wang L, Li Y, Zhang H, Chen Y. Deep learning CAD system for mammography classification. Medical Image Analysis. 2023;86:102765.DOI: 10.1016/j.media.2023.102765
17. Islam MT, Rahman MM, Ahmed F, Khan S. Automated breast cancer detection using CNN. Computers in Biology and Medicine. 2023;157:106724.DOI: 10.1016/j.compbiomed.2023.106724
18. Ahmed M, Bibi T, Khan R, Shah S. Explainable AI for breast cancer detection using CNN. IEEE Access. 2024;12(1):44567-44580.DOI: 10.1109/ACCESS.2024.3367812
19. Zhang Y, Chen X, Liu J, Wang H. Multimodal fusion network for breast cancer classification. Pattern Recognition. 2024;145:109876.DOI: 10.1016/j.patcog.2024.109876
20. Wei TR, Lin Y, Chen Z, Huang K. Multimodal deep learning for breast cancer classification. Computers in Biology and Medicine. 2025;180:108456.DOI: 10.1016/j.compbiomed.2025.108456
21. Nasir F, Rahman M, Islam S, Ali K. CNN-based breast cancer detection review. Expert Systems with Applications. 2025;235:120123.DOI: 10.1016/j.eswa.2025.120123
22. Tang J, Ming R, Wang D, Zhou G. Multimodal AI model for breast cancer screening. Scientific Reports. 2025;15(1):14696.DOI: 10.1038/s41598-025-14696-0
23. Ben Rabah C, Saadi M, Kaaniche N, Fnaiech F. Multimodal deep learning for breast lesion classification. Diagnostics. 2025;15(8):995.DOI: 10.3390/diagnostics15080995
24. Abbadi M, Himeur Y, Atalla S, Mansoor W. Transfer learning for breast ultrasound cancer detection. IEEE Transactions on Biomedical Engineering. 2025;72(4):1120-1132.DOI: 10.1109/TBME.2024.3352109
25. Selvani D, Poornila AA. Deep learning techniques for breast cancer detection using medical image analysis. Springer. 2018;1(1):1-10.DOI: 10.1007/978-981-10-9035-6