NaFV-Net: An Adversarial Four-view Network for Mammogram Classification

Feng Lu[*1], Yuxiang Hou[*1], Wei Li[*2], Xiangying Yang[3], Haibo Zheng[1], Wenxi Luo[1], Leqing Chen[3], Yuyang Cao[2], Xiaofei Liao[1], Yu Zhang[1], Fan Yang[✉3], Albert Zomaya[✉2], Hai Jin[✉1]

[1] National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China

[2] The Australia-China Joint Research Centre for Energy Informatics and Demand Response Technologies, Centre for Distributed and High Performance Computing, School of Computer Science, The University of Sydney, Australia

[3] Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, China

{lufeng,houyx,xiangyingyang,zhenghb,d202280886,leqingchen,xfliao,zhyu,fyang,hjin}@hust.edu.cn, {weiwilson.li,albert.zomaya}@sydney.edu.au,{ycao0726}@uni.sydney.edu.au.

Abstract

Breast cancer remains a leading cause of mortality among women, with millions of new cases diagnosed annually. Early detection through screening is crucial. Using neural networks to improve the accuracy of breast cancer screening has become increasingly important. In accordance with radiologists' practices, we propose using images from the unaffected side to create adversarial samples with critical medical implications in our adversarial learning process. By introducing beneficial perturbations, this method aims to reduce overconfidence and improve the precision and robustness of breast cancer classification. Our proposed framework is an adversarial four-view classification network (NaFV-Net) incorporating images from both affected and unaffected perspectives. By comprehensively capturing local and global information and implementing adversarial learning from four mammography views, this framework allows for the fusion of features and the integration of medical principles and radiologist evaluation techniques, thus facilitating the accurate identification and characterization of breast tissues. Extensive experiments have shown the high effectiveness of our model in accurately distinguishing between benign and malignant findings, demonstrating state-of-the-art classification performance on both internal and public datasets.

Introduction

Breast cancer remains a significant global health challenge, affecting millions of women annually. In 2020, the World Health Organization reported 2.3 million predominantly female cases diagnosed, resulting in 685,000 deaths. While mammography is widely used for breast cancer screening (Duffy et al. 2002; Gøtzsche 2002; Gotzsche 2002), it often yields high false positive rates due to the complex nature of breast cancer (Alukić et al. 2023) and breast tissue density (Ding et al. 2009). This requires additional radiological and/or ultrasound assessments for confirmation (Kopans 2015), though only a small percentage ultimately reveal malignancies (Wu 2020). Consequently, enhancing the accuracy, both precision and recall, of breast cancer diagnosis in mammograms using AI is crucial. Such improvements would reduce unnecessary resource consumption, expedite confirmation times, and alleviate patient anxiety for those not requiring further testing.

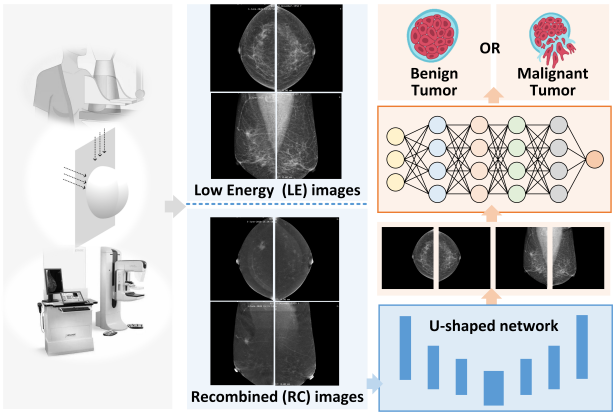

Figure 1: In mammograms, the *cranio-caudal* (CC) view is top-down, while the *medio-lateral oblique* (MLO) view is taken at a certain angle from the side. Recombined (RC) images have clear positional information of the mass, but have lost the specific texture and morphological information of the mass contained in the *Low Energy* (LE) images.

Recent trends in breast cancer mammogram diagnosis have favored multi-view feature fusion methods. This approach leverages the nature of mammography, which provides *cranio-caudal* (CC) and *medio-lateral oblique* (MLO) views of both breasts independently. As illustrated in Figure 1, the CC view offers a top-to-bottom perspective of the breast, while the MLO view provides an oblique view of the breast tissue from the side (Wu et al. 2019; Zhao, Yu, and Wang 2020). Common approaches involve using networks to extract features from both CC and MLO views, then combining these features to reach conclusive results (Ma et al. 2021; Li et al. 2020; Yang et al. 2021). However, current methods often fail to fully exploit multi-view mammogra-

[*]These authors contributed equally.

phy, focusing primarily on the CC and MLO views of the affected side while neglecting the non-affected side. Research indicates that relying solely on the CC and MLO views of the affected side yields more accurate results than including all four views or breast-level samples (Zhao, Yu, and Wang 2020). Nevertheless, radiologists routinely examine both affected and non-affected side views to form comprehensive and insightful evaluations.

We propose investigating the potential of generating adversarial samples from non-affected mammography images to introduce favorable perturbations that can enhance the accuracy of breast cancer classification. Previous research has demonstrated that adversarial learning methods can subtly differentiate between benign and malignant lesions by extracting features from synthesized adversarial samples. However, prior studies have encountered challenges regarding the reliability of traditionally generated adversarial samples and their medical implications. To address this issue, we propose using preprocessed contralateral images as adversarial samples, in conjunction with the CC and MLO views of the affected side, to create four-view mammographic samples. This approach aims to mitigate model overconfidence and risks of overfitting during model training, ultimately improving overall model performance.

To address these challenges, we propose an adversarial multi-view classification network framework that integrates affected and non-affected side images to improve breast cancer classification accuracy. The framework comprises three primary components: breast mass segmentation, generative adversarial, and multi-view classification network. The breast mass segmentation employs an asymmetric sensitivity algorithm and a cross-attention mechanism to precisely identify and segment mass areas in affected images. It evaluates breast mass by comparing asymmetric areas in affected and non-affected side images. The generative adversarial algorithm utilizes the non-affected side of the same patient and the perturbation of masses extracted from the affected side to create adversarial samples. These samples are valuable as they represent genuine false positive instances. The multi-view classification network integrates images from CC and MLO views on both sides to provide different perspectives of the same lesion area, aiding in accurate mass classification by analyzing morphological and anatomical information derived from multi-view images, similar to an experienced radiologist. In summary, the main contributions of our work are as follows:

- We propose an adversarial four-view network (NaFV-Net) integrating affected and non-affected side images that leverages medical principles and radiologist evaluation methods for accurate breast cancer classification on mammography images.
- We propose a method based on constructing adversarial examples using real images from the non-affected side, and introduce a training method combining affected and non-affected side samples based on lesion area perturbation, along with a corresponding loss function.
- The experiment results show that our approach is effective on real-world datasets. Our approach outperforms other *state-of-the-art* (SOTA) solutions for breast cancer classification, using internal and public datasets.

Background and Related Work

Background: In the context of neural network-based classification of breast cancer into benign and malignant categories, the accurate segmentation of masses is crucial. *Contrast-enhanced mammography* (CEM) technology (Beuque et al. 2023), especially CEM Recombined (RC) images, enables precise identification of mass locations and boundaries, as demonstrated in Figure 1.

Multi-view Classification for Breast Cancer: Exploring multi-view classification methods to improve the classification accuracy for identifying benign and malignant is crucial, considering that patients typically undergo imaging generating four views: CC-Left, CC-Right, MLO-Left, and MLO-Right. Recent research has utilized intricate network structures to analyze connections among mammography views (Yang et al. 2021; Van Tulder, Tong, and Marchiori 2021; Ma et al. 2021; Zhao, Yu, and Wang 2020; Hu et al. 2018). For example, deep learning models with feature connection strategies have been employed to improve breast tissue classification (Khan et al. 2019; Wu 2020; Li et al. 2020). However, a significant challenge in this approach is that the breast tissues occupy different areas in different views, making alignment difficult. Since different views may not be aligned, concatenation-based methods fuse multi-view information after the global pooling layer. Nevertheless, they lose a large amount of local information and cannot fully exploit the multi-view information (Sun et al. 2019).

Generative Adversarial: Adversarial learning has demonstrated meaningful potential in breast cancer classification. One approach involves the use of *Generative Adversarial Networks* (GANs). For example, Wang et al. (Wang et al. 2021) utilized a conditional GAN-based method to generate breast images with diverse lesion characteristics, showcasing the effectiveness of the generated images in improving classifier performance. Another technique involves employing adversarial training to bolster model robustness. For instance, Mukherjee and Kundu (Jochelson and Lobbes 2021) incorporated adversarial training by introducing perturbations to breast images, yielding a more resilient classification model and favorable results. Moreover, some studies have applied adversarial learning to merge breast images from different modalities, such as mammography and MRI, to enhance classification performance (Hao et al. 2024).

Methodology

In Figure 2, an overview of our NaFV-Net is depicted. This network leverages the CC and MLO views from both affected and non-affected sides to aid radiologists in the detection and delineation of breast mass malignancies. Our model is designed to perform the dual tasks of breast mass segmentation and subsequently classifying them.

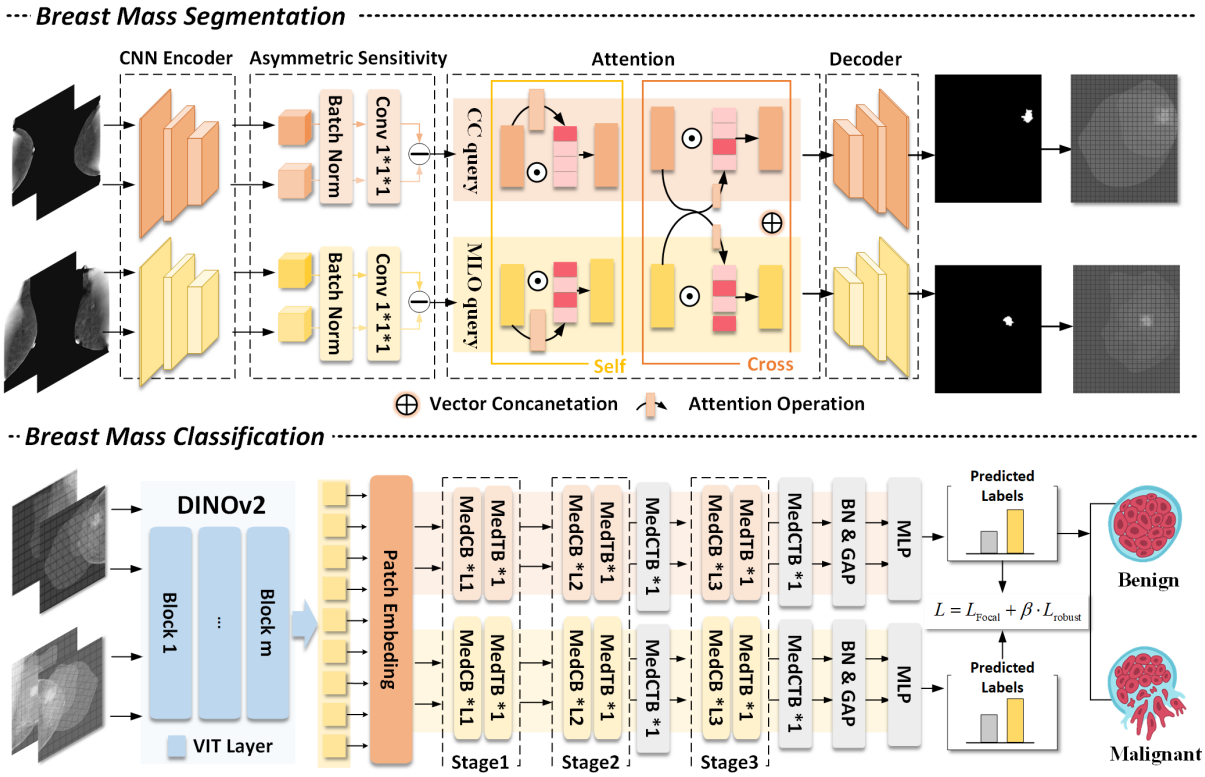

Figure 2: Overview of NaFV-Net

The Breast Mass Segmentation

We initially preprocess four RC view images and input them into the model's encoder to convert them into feature vectors: $\{L_{cc}, R_{cc}, L_{mlo}, R_{mlo}\} \in \mathbb{R}^{H\times W\times C_1}$. The encoder comprises three concatenated convolutional blocks, each containing a 3×3 convolutional layer, a group normalization layer, and a ReLU layer. The encoder converts the $H \times W$ image into an $\frac{H}{P} \times \frac{H}{P}$ feature map, while the decoder converts the $\frac{H}{P} \times \frac{H}{P}$ feature map back into an $H \times W$ image.

Then, the model utilizes an asymmetric sensitive method to merge the left and right RC images. Since the likelihood of a tissue area appearing in the exact location in both images is minimal, our approach capitalizes on the asymmetry of the pathology. It enhances attention to places that may exhibit significant pathology while identifying the lesion area's substantial enhancement from *Breast Parenchymal Enhancement* (BPE) caused by various factors.

Utilizing an asymmetric sensitive approach, we have developed a method for merging the left and right RC image features: $\{\mathcal{RC}^{\mathcal{L}}_{cc}, \mathcal{RC}^{\mathcal{L}}_{mlo}, \mathcal{RC}^{\mathcal{R}}_{cc}, \mathcal{RC}^{\mathcal{R}}_{mlo}\} \subseteq \mathbb{R}^{H\times W}$. This methodology holds excellent potential for improving the accuracy and effectiveness of findings detection in clinical practice. Using the affected and unaffected side's CC views in LE images ($Z^{\mathcal{A}s}_{cc}$ and $Z^{\mathcal{U}a}_{cc}$) as the example, the calculation process is expressed as follows:

$$M_s = Sig(Conv(AvgPool(Z^{\mathcal{U}a}_{cc}))) \tag{1}$$

$$F^{out}_{cc} = Z^{\mathcal{A}s}_{cc} \cdot M_s \tag{2}$$

where $M_s \in R^{H\times W\times 1}$ is the obtained spatial attention map for $Z^{\mathcal{A}s}_{cc}$, and $F^{out}_{cc} \in R^{\frac{H}{P}\times\frac{W}{P}\times C_1}$ is the input of the next convolution block reassigned by M_s and $Z^{\mathcal{A}s}_{cc}$.

$$Z^{out}_{cc} = Concat(Z^{\mathcal{A}s}_{cc}, F^{out}_{cc}) \tag{3}$$

where Z^{out}_{cc} is the concatenated feature map of the affected and unaffected side.

Upon concatenating the feature maps of the affected side, the Z^{out}_{cc} and Z^{out}_{mlo} are subsequently tokenized into a sequence of flattened 2D patches $\{x^i_P \in \mathbb{R}^{P^2C} | i = 1, 2.., N\}$, where each patch is of size $P \times P$ and N is the number of image patches. After position encoding,

$$\tilde{Z}^0_{cc/mlo} = [x^1_P E; x^2_P E; ...; x^N_P E] + E_{pos} \tag{4}$$

Take the CC views and the output of the l-th layer as an example. The multi-head self-attention mechanism can be defined as follows. Additionally, the output of the l-th layer can be expressed through this mechanism.

$$\hat{Z}^l_{cc} = MSA(LN(\tilde{Z}^{l-1}_{cc})) + \tilde{Z}^{l-1}_{cc} \tag{5}$$

$$\dot{Z}^l_{cc} = MLP(LN(\hat{Z}_l)) + \hat{Z}_l \tag{6}$$

Similarly, the multi-head cross-attention mechanism can be defined in the same context. This is defined in *Appendix* with $\dot{Z}^l_{cc}$. In hyperparameter learning, $Z^{\mathcal{S}}_{cc}$, $Z^{\mathcal{C}}_{cc}$, are respectively the outputs of the last layer for $\dot{Z}^l_{cc}$, $\dot{Z}^l_{cc}$, while $\lambda^{\mathcal{S}}$, $\lambda^{\mathcal{C}}$

Algorithm 1: Lesion-Enriched Image Training and Adversarial Sample Generation

Input: Mass location $\mathcal{R}$, image matrix $\mathcal{I}_{\text{matrix}}$
Output: $\mathcal{D}_{\text{matrix}}, \mathcal{L}$
Extract $\mathcal{X}_{\min}, \mathcal{X}_{\max}, \mathcal{Y}_{\min}, \mathcal{Y}_{\max}$

$$\mathcal{M}(x,y) = \begin{cases} 1 & \text{if } (x,y) \in \mathcal{R} \\ 0.9 & \text{if not } \mathcal{R} \\ 0.9 - \text{dist}(x,y,\text{center}(\mathcal{R})) & \text{if outside box} \end{cases}$$

$\mathcal{D}_{\text{matrix}} = \mathcal{I}_{\text{matrix}} \circ \mathcal{M}$
Generate adversarial sample:
$\mathcal{F}(x,y) = \alpha \cdot \mathcal{O}(x,y) + (1-\alpha) \cdot \mathcal{R}(x,y)$
Composite loss function:

$$\mathcal{L} = \mathcal{L}_{\text{Focal}} + \beta \cdot \mathcal{L}_{\text{robust}}$$

Focal Loss:

$$\mathcal{FL}(p_t) = -\alpha_t (1-p_t)^{\gamma} \log(p_t)$$

Robust Loss:

$$\mathcal{L}_{\text{robust}} = \text{KL}(\mathcal{O}_{\text{image}}(x,y), \mathcal{F}_{\text{image}}(x,y))$$

return $\mathcal{D}_{\text{matrix}}, \mathcal{L}$

are the weight parameters corresponding to $Z_{cc}^{\mathcal{S}}$, $Z_{cc}^{\mathcal{C}}$.

$$\tilde{Z}_{cc} = \sum_{i \in \{\mathcal{S},\mathcal{C}\}} \lambda^i \cdot Z_{cc}^i \tag{7}$$

Finally, the feature encoder upsamples $\tilde{Z}_{cc}$ to an $H \times W$ image.

The Breast Mass Classification

Generative Adversarial After pinpointing the precise mass location $\mathcal{R}$ from the mass segmentation, we extract the boundary coordinates $\mathcal{X}_{\min}$, $\mathcal{X}_{\max}$, $\mathcal{Y}_{\min}$, and $\mathcal{Y}_{\max}$, thereby defining the bounding box of the breast mass. This bounding box, combined with the pathological ground truth labels, formulates the label set for LE image training. Given the potential influence of non-lesion areas on mass classification, we construct an image-level mask matrix. Specifically, the mask value for coordinates within $\mathcal{R}$ is set to 1, while those within the bounding box but outside $\mathcal{R}$ are assigned a value of 0.9. For points outside the mass bounding box, mask values decrease from 0.9 to 0 based on their Euclidean distance from the center of $\mathcal{R}$. The input data matrix for the network is the product of the image matrix $\mathcal{I}_{\text{matrix}}$ and the mask matrix $\mathcal{M}_{\text{matrix}}$:

$$\mathcal{D}_{\text{matrix}} = \mathcal{I}_{\text{matrix}} \circ \mathcal{M}_{\text{matrix}} \tag{8}$$

To improve the feature extraction capabilities of the model and differentiate it from other methods of generating adversarial samples, we have employed authentic images and implemented a mixup-like algorithm 1 for creating samples for adversarial training. Our hypothesis is based on the observation that human tissue exhibits significant overall symmetry, with subtle variations at the organ or tissue level, which can serve as natural disturbances for model training. As shown in Figure 3, to put this into practice, after identifying the bounding box of the breast mass on the affected side, we combine it with the symmetrical area from the non-lesion image. Before overlaying the bounding box of the mass from the affected side onto the symmetrical area on the unaffected side, we use a soft contour mask to seamlessly blend the original and target regions. The resulting merged image is denoted by:

$$\mathcal{F}(x,y) = \alpha \cdot \mathcal{O}(x,y) + (1-\alpha) \cdot \mathcal{R}(x,y) \tag{9}$$

where α is the value from the soft contour mask.

After generating the fused image on the unaffected side, we fine-tune the coordinates of the lesion bounding box to introduce slight misalignments, thereby adding perturbations to the model's perception of the precise lesion location. Similarly, the adversarial sample matrix input to the network is the product of the image matrix $\mathcal{I}_{\text{matrix}}$ and the mask matrix $\mathcal{M}_{\text{matrix}}$.

During the loss computation and backpropagation phase of our multi-view classification network, we opte to substitute the conventional cross-entropy loss function with a composite loss function comprising a penalty mechanism and adversarial loss. This methodology is designed to augment the model's generalization capability, resilience to perturbations, and feature extraction prowess. Our loss function $\mathcal{L}$ is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{Focal}} + \beta \cdot \mathcal{L}_{\text{robust}} \tag{10}$$

The Focal Loss $\mathcal{L}_{\text{Focal}}$, designed to address class imbalance by down-weighting easy examples and focusing on hard ones, is defined as:

$$\mathcal{FL}(p_t) = -\alpha_t (1-p_t)^{\gamma} \log(p_t) \tag{11}$$

where p_t is the predicted probability for the true class, α_t is a weighting factor for class t, and γ is a focusing parameter adjusting the rate at which easy examples are down-weighted.

$\mathcal{L}_{\text{robust}}$ represents the KL divergence loss between the original image $\mathcal{O}_{\text{image}}(x,y)$ and the fused image $\mathcal{F}_{\text{image}}(x,y)$, incorporating perturbations for adversarial robustness. This integration helps create realistic adversarial samples. Consequently, our composite loss function ensures that the model is not only accurate in classification but also resilient to adversarial perturbations, thereby enhancing its overall performance in multi-view classification tasks.

Multi-view Classification Network After obtaining images with clear mass position information through the breast mass segmentation stage, we will conduct adversarial training by treating the affected side's CC and MLO position images and the unaffected side's CC and MLO position images as a whole. These images are then fed into the fine-tuned DINOv2 to extract image features (We selected three publicly available breast mammography datasets—CBIS-DDSM, INbreast, and CMMD—and fine-tuned DINOv2 on them to ensure compatibility with the requirements of deep learning models, as detailed in the *appendix*.), which serve as inputs for training a quadruple-view breast tissue classifier. To facilitate this, we have designed a mixed model for

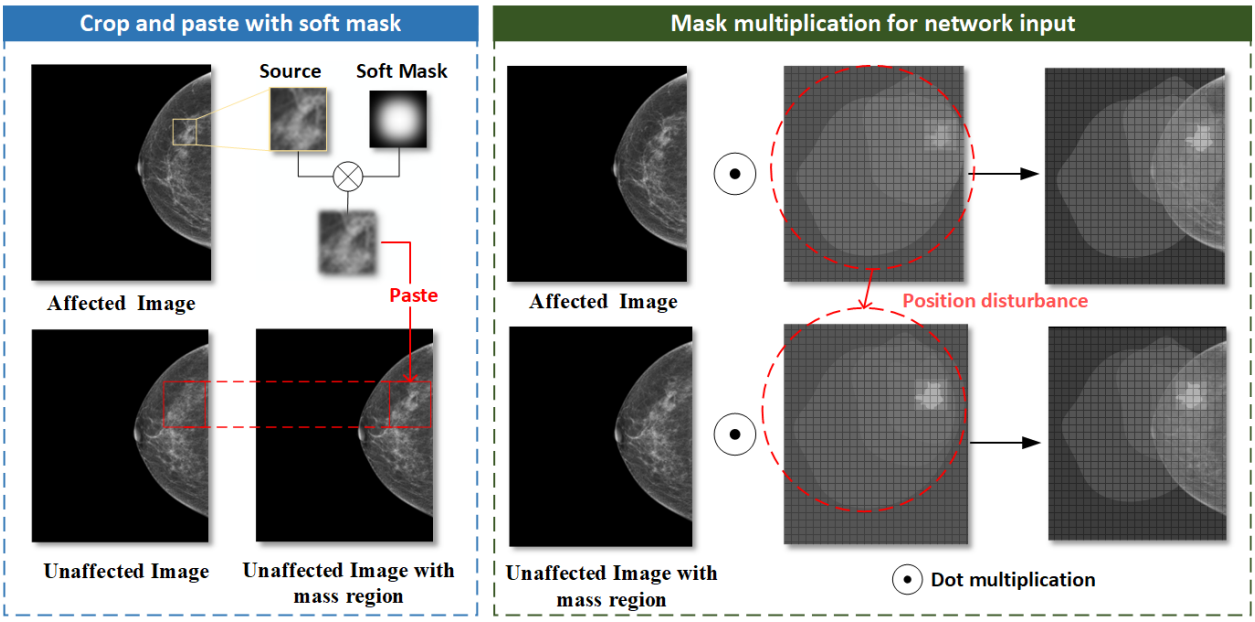

Figure 3: The generation of adversarial training samples involves a two-step process. First, a soft mask is used to crop and paste the segmented mass from the affected side image onto the symmetrical position of the unaffected side image. Second, mask multiplication is performed on the affected and unaffected images to create the input images for the network.

breast tissue classification, which combines the strengths of the CNN and Transformer. The model ultimately outputs a prediction label for four images.

After patch embedding, the classifier first adopts the Medical Convolution Block (MedCB), which draws from the residual block in ResNet and combines convolution and transformer (Manzari et al. 2023) design but does not include self-attention, thus enhancing performance and capturing higher-level structures, which applies to medical imaging. In stage 1, affected images and un-affected images pass through L1 times MedCB (Medical Convolution Block), followed by a pass through a time Medical Transformer Block (MedTB), which is combined to capture and mix multi-frequency information, providing depth and width to the model. In stage 2, after undergoing L2 times MedCB and a time MedTB, both affected images and un-affected images pass through a time Medical Cross Transformer Block (MedCTB), which cross-correlates the CC view and MLO view of either affect or un-affect images to capture and mix multi-view information. In stage 3, following L3 times MedCB and a time MedTB for both affected images and un-affected images, they jointly pass through a time Medical Cross Transformer Block (MedCTB). After jointly passing through a Batch Normalization Layer, a Global Average Pooling Layer ,and a MLP layer, affect images output a label indicating whether the finding is benign or malignant. See the *appendix* for detailed designs of the MedCB, MedTB, MedCTB blocks, and the specific settings of L1, L2, and L3 times. *Appendix* and more details can be released at https://github.com/CGCL-codes/NaFV-Net.

Experiments and Results

Internal and Public Datasets

We trained our model on Tongji dataset from Tongji Hospital. This internal dataset contains CEM images of 479 breast cancer patients, with 284 malignant cases and 195 benign cases. Each patient has 4 RC images as well as 4 LE images. The *Institutional Review Board* (IRB) of Tongji Hospital approved this study and waived the requirement for patient informed consent as part of the study approval (IRB number: 20230790). We validated our model on both the internal dataset and the public dataset. The public dataset is the CDD-CESM dataset from Egypt (Khaled et al. 2022), and to our knowledge, it is currently the only public CEM mammography dataset in the world. It contains 210 breast mass images, 184 non-mass enhancement images, as well as corresponding medical reports and pathological diagnoses. There are 167 positive cases and 92 negative cases with complete 4 RC images and 4 LE images available for our external validation. Just like previous breast cancer classification studies, we label images as two classes: the BIRADS scores belonging to 1, 2, 3 as normal or benign, and 4, 5, 6 as malignant.

Experimental Environment Setting

Our NaFV-Net has two independent components, so we adopted different strategies for each part during the training process. For breast mass segmentation, we employed an optimizer for 200 training epochs. We selected the batch with the highest Dice and mIoU metrics in the validation set after 100 epochs as our primary model. Testing was confined to our internal dataset due to the absence of mass annotations

Method	Views	Data Division	AUC	Precision	Recall
Internal Dataset					
Deep MIL (Zhu et al. 2017)	Single	Image	0.824 ± 0.022	0.800 ± 0.029	0.799 ± 0.025
Nan Wu et.al (Wu et al. 2019)	Dual	Patient	0.852 ± 0.015	0.822 ± 0.019	0.804 ± 0.024
GMIC (Shen et al. 2021)	Dual	Patient	0.878 ± 0.019	0.848 ± 0.020	0.844 ± 0.022
Sun et.al (Sun et al. 2022)	Dual	Patient	0.865 ± 0.020	0.835 ± 0.023	0.831 ± 0.027
PHBreast (Lopez et al. 2022)	Dual	Patient	0.874 ± 0.021	0.839 ± 0.020	0.825 ± 0.022
BRAIxMVCCL (Chen et al. 2022)	Dual	Patient	0.893 ± 0.018	0.862 ± 0.021	0.860 ± 0.027
DCHA-Net (Wang et al. 2023)	Dual	Patient	0.907 ± 0.015	0.874 ± 0.018	0.870 ± 0.023
DINOv2 (Oquab et al. 2023)	Single	Image	0.764 ± 0.029	0.758 ± 0.026	0.753 ± 0.030
Ours	Four	Patient	**0.935± 0.015**	**0.902 ± 0.018**	**0.899 ± 0.019**
Public Dataset					
Deep MIL (Zhu et al. 2017)	Single	Image	0.801± 0.019	0.779± 0.027	0.780± 0.027
Nan Wu et.al (Wu et al. 2019)	Dual	Patient	0.829 ± 0.019	0.784 ± 0.020	0.798 ± 0.022
GMIC (Shen et al. 2021)	Dual	Patient	0.843 ± 0.020	0.826 ± 0.021	0.819 ± 0.022
Sun et.al (Sun et al. 2022)	Dual	Patient	0.813 ± 0.021	0.782 ± 0.026	0.790 ± 0.027
PHBreast (Lopez et al. 2022)	Dual	Patient	0.839 ± 0.016	0.809 ± 0.017	0.811 ± 0.019
BRAIxMVCCL (Chen et al. 2022)	Dual	Patient	0.860 ± 0.017	0.852 ± 0.017	0.840 ± 0.020
DCHA-Net (Wang et al. 2023)	Dual	Patient	0.871 ± 0.015	0.857 ± 0.017	0.843 ± 0.021
DINOv2 (Oquab et al. 2023)	Single	Image	0.752 ± 0.020	0.740 ±0.026	0.743 ±0.028
Ours	Four	Patient	**0.883 ± 0.014**	**0.865 ± 0.018**	**0.848 ± 0.018**

Table 1: Quantitative comparison of different benchmark models on both internal dataset and public dataset based on AUC, Precision, Recall. The AUC, Precision, and Recall values in the table are the average values after 10-fold cross-validation for each method.

Model	Dice	Miou
Internel dataset		
UNet (Ronneberger and Fischer 2015)	77.45	69.15
UNet++ (Zhou et al. 2019)	81.61	72.12
TransUNet (Chen et al. 2021)	84.64	74.85
Unext (Valanarasu and Patel 2022)	84.52	76.03
SwinUNet (Cao et al. 2022)	85.14	76.31
Medsam (Ma and Wang 2023)	81.97	71.83
Ours	**88.18**	**79.16**

Table 2: Comparison with SOTA medical image segmentation methods based on dice (%) and miou (%)

in the external dataset. The model for this component was initialized with a learning rate of 0.01, a weight decay of 1e-4, and a momentum of 0.9.

For breast mass classification, we used an optimizer for 300 training epochs. After observing no further increase in AUC for 30 epochs, we stopped training, subsequently selecting the model with the highest AUC. The model for this stage was initialized with a learning rate of 0.001 and a weight decay of 0.01. Similar to mass segmentation, we randomly partitioned the dataset into training, validation, and test sets in a 7:2:1 ratio and used 10-fold cross-validation for evaluation. The PyTorch framework was utilized to implement this model, with an NVIDIA A100 and a 40GB GPU.

Performance Benchmarks

In assessing our model's breast mass segmentation performance, we conducted a comparative analysis with SOTA methods, including UNet (Ronneberger and Fischer 2015), TransUNet (Chen et al. 2021), UNext (Valanarasu and Patel 2022), UNet++ (Zhou et al. 2019), Medsam (Ma and Wang 2023), and SwinUNet (Cao et al. 2022).

To evaluate the performance of our model's breast mass classification, we utilized SOTA breast cancer classification models, including Deep MIL (Zhu et al. 2017), Nan Wu et.al (Wu et al. 2019), Sun et.al (Sun et al. 2022), PHbreast (Lopez et al. 2022), GMIC (Shen et al. 2021), BRAIxMVCCL (Chen et al. 2022), and DCHA-Net (Wang et al. 2023). We aimed to maintain consistency by aligning our training strategy and parameter settings as closely as possible with those of our model.

Comparison of Classification

We conducted a comprehensive comparative analysis utilizing both internal and public datasets to assess the efficacy of our approach in breast cancer classification. Our method was benchmarked against seven SOTA breast cancer classification models and the DINOv2 model from the natural image domain. The findings presented in Table 1 demonstrate that the majority of the compared approaches are dual view-based methods, with the exception of Deep MIL, which is a single view-based method. Our method effectively utilizes the four-view features of patients' CEM RC and LE images, outperforming all benchmark models in terms of AUC (0.935, a 3% to 22.4% improvement), Precision (0.902, a 3.2% to 18.9% improvement), and Recall (0.899, a 3.3% to 19.4% improvement). Furthermore, upon evaluation using a public dataset, our method surpassed all benchmark models in AUC (0.883, an improvement of 1.4% to 17.4%), Precision (0.855, an improvement of 0.9% to 15.5%), and Recall (0.848, an improvement of 0.6% to 10.5). Figure 4 depicts

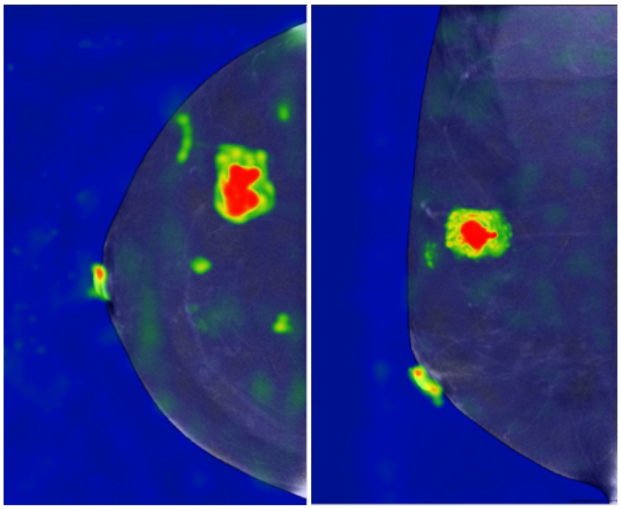

Figure 4: It offers a visual representation. In a specific case, an initial diagnostic report from a medical professional suggested, "This is mostly indicative of malignant tumoral changes. BI-RADS 4c." Subsequent pathology results, however, revealed "Fibroadenoma of the breast, benign mass." This necessitated a correction to the doctor's initial assessment. Our model provides accurate classification, outlines the lesion area in red, and generates heat maps on LE images during breast mass classification, thereby aiding radiologists in making precise diagnoses.

the ROC curves that compare our method with benchmark models on internal and external datasets. The visualization also highlights a scenario where our model contributes to assisting radiologists in achieving precise diagnoses through the use of heat maps and accurate classification.

Comparison of Segmentation

The comparative results for segmentation models can be found in Table 2. Compared to the baseline model, Unet, our model exhibited significant improvement, achieving a 10.7% increase in the Dice coefficient and a 10.0% improvement in the mIoU metric. Our model demonstrated a 4% improvement in the Dice coefficient and a roughly 3% enhancement in the mIoU metric compared to recent leading methodologies. Furthermore, contrasted with the unsupervised large medical imaging segmentation model Medsam, our model showed a 6.2% increase in the Dice coefficient and a 7.3% enhancement in the mIoU metric. Additionally, a comprehensive comparative analysis is depicted in *Appendix* Figure 2, showing our methodology's performance against other approaches in segmenting tissue within the same context.

Ablation Study

To assess the rationality of each module, we conducte ablation experiments and have outlined the results in Table 3.

In the phase of breast mass segmentation, we have incorporated two distinct modules: an asymmetric sensitivity module and an attention module, aiming to improve the model's segmentation accuracy. As evidenced by the findings presented in Table 3, the asymmetric sensitivity module yielded a 4.0% enhancement in the segmentation metric Dice and a 1.2% improvement in the classification performance AUC. Additionally, the designed attention module demonstrated significant capability in incorporating the morphological characteristics of breast tissue, resulting in a notable 9.1% improvement in the segmentation metric Dice and a 1.6% enhancement in the classification performance AUC.

Asymmetry sensitivity	Attention	Four-view	Position mask	Dice (%)	AUC (%)
	✓	✓	✓	84.186	92.34
✓		✓	✓	79.090	91.92
✓	✓			-	88.24
✓	✓		✓	-	91.58
✓	✓	✓	✓	88.182	93.52
Not Fine-tuning on DINOv2				-	91.47
Fine-tuning on DINOv2				-	93.52

Table 3: Ablation study on the internal dataset. ✓ indicates the presence of the corresponding module. "-" denotes that the metric was not tested in this experiment.

During the breast mass classification phase, our ablation experiments focused on two primary considerations. Firstly, we exclusively utilized the affected side CC and MLO images while omitting unaffected side images to develop a dual-view classification network. The loss function segment of this approach employed the cross-entropy loss function. We ensured that the intricacies of the dual-view classification network were consistent with the affected side component of the four-view classification network, resulting in a 5.3% reduction in classification performance compared to the four-view classification method, as indicated in the third row of Table 3. Secondly, we refrained from manipulating the masks of the unaffected side images, thus retaining the symmetrical alignment of positional information contained in the masks of the input affected and unaffected side images. This approach yielded a 2.0% decrease in classification performance compared to the four-view classification method, as demonstrated in the fourth row of Table 3.

Conclusion

Our work introduces an innovative adversarial multi-view classification network designed to enhance the accuracy of breast cancer diagnosis. The network combines images from both affected and non-affected sides, utilizes an asymmetric sensitivity algorithm, a cross-attention mechanism for precise segmentation, and adversarial learning with real contralateral images to prevent overfitting. The hybrid multi-view classifier effectively discerns between benign and malignant lesions. In our future endeavors, we aim to expand the dataset and undertake multi-center validation.

Acknowledgments

This work is supported by the Key Project of the National Natural Science Foundation of China (62232012) and the Hubei Big Data Analysis Platform and Intelligent Service Project for Medical and Health.

References

Alukić, E.; Homar, K.; Pavić, M.; Žibert, J.; and Mekiš, N. 2023. The impact of subjective image quality evaluation in mammography. *Radiography*, 29(3): 526–532.

Beuque, M. P.; Lobbes, M. B.; van Wijk, Y.; Widaatalla, Y.; Primakov, S.; Majer, M.; Balleyguier, C.; Woodruff, H. C.; and Lambin, P. 2023. Combining deep learning and handcrafted radiomics for classification of suspicious lesions on contrast-enhanced mammograms. *Radiology*, 307(5): e221843.

Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; and Wang, M. 2022. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, 205–218. Springer.

Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A. L.; and Zhou, Y. 2021. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Chen, Y.; Wang, H.; Wang, C.; Tian, Y.; Liu, F.; Liu, Y.; Elliott, M.; McCarthy, D. J.; Frazer, H.; and Carneiro, G. 2022. Multi-view local co-occurrence and global consistency learning improve mammogram classification generalisation. In *Proceeding of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 3–13. Springer.

Ding, J.; Peng, W.; Jiang, Z.; Xu, L.; Hu, D.; and Zheng, X. 2009. Diagnostic value of full-field digital mammography for breast carcinoma. *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, 31(11): 854–857.

Duffy, S. W.; Tabár, L.; Chen, H.-H.; Holmqvist, M.; Yen, M.-F.; Abdsalah, S.; Epstein, B.; Frodis, E.; Ljungberg, E.; Hedborg-Melander, C.; et al. 2002. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties: a collaborative evaluation. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 95(3): 458–469.

Gøtzsche, P. C. 2002. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer*, 94(2): 578–578.

Gotzsche, P. C. 2002. The mammographic screening trials: Commentary on the recent work by Olsen and Gotzsche-Invited response.

Hao, D.; Arefan, D.; Zuley, M.; Berg, W.; and Wu, S. 2024. Adversarially Robust Feature Learning for Breast Cancer Diagnosis. *arXiv preprint arXiv:2402.08768*.

Hu, H.; Gu, J.; Zhang, Z.; Dai, J.; and Wei, Y. 2018. Relation networks for object detection. In *Proceeding of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3588–3597. Salt Lake City: IEEE.

Jochelson, M. S.; and Lobbes, M. B. 2021. Contrast-enhanced mammography: state of the art. *Radiology*, 299(1): 36–48.

Khaled, R.; Helal, M.; Alfarghaly, O.; Mokhtar, O.; Elkorany, A.; El Kassas, H.; and Fahmy, A. 2022. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. *Scientific Data*, 9(1): 122.

Khan, H.; Shahid, A.; Raza, B.; Dar, A.; and Alquhayz, H. 2019. Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. *IEEE Access*, 7: 165724–165733.

Kopans, D. B. 2015. An open letter to panels that are deciding guidelines for breast cancer screening. *Breast Cancer Research and Treatment*, 151: 19–25.

Li, C.; Xu, J.; Liu, Q.; Zhou, Y.; Mou, L.; Pu, Z.; Xia, Y.; Zheng, H.; and Wang, S. 2020. Multi-view mammographic density classification by dilated and attention-guided residual learning. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(3): 1003–1013.

Lopez, E.; Grassucci, E.; Valleriani, M.; and Comminiello, D. 2022. Hypercomplex Neural Architectures for Multi-View Breast Cancer Classification. *arXiv preprint arXiv:2204.05798*.

Ma, J.; Li, X.; Li, H.; Wang, R.; Menze, B.; and Zheng, W.-S. 2021. Cross-view relation networks for mammogram mass detection. In *Proceeding of 2020 25th International Conference on Pattern Recognition (ICPR)*, 8632–8638. Milan: IEEE.

Ma, J.; and Wang, B. 2023. Segment anything in medical images. *arXiv preprint arXiv:2304.12306*.

Manzari, O. N.; Ahmadabadi, H.; Kashiani, H.; Shokouhi, S. B.; and Ayatollahi, A. 2023. MedViT: a robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157: 106791.

Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. 2023. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*.

Ronneberger, O.; and Fischer. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceeding of International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Shen, Y.; Wu, N.; Phang, J.; Park, J.; Liu, K.; Tyagi, S.; Heacock, L.; Kim, S. G.; Moy, L.; Cho, K.; et al. 2021. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. *Medical Image Analysis*, 68: 101908.

Sun, L.; Wang, J.; Hu, Z.; Xu, Y.; and Cui, Z. 2019. Multi-view convolutional neural networks for mammographic image classification. *IEEE Access*, 7: 126273–126282.

Sun, Z.; Jiang, H.; Ma, L.; Yu, Z.; and Xu, H. 2022. Transformer Based Multi-view Network for Mammographic Image Classification. In *Proceeding of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 46–54. Springer.

Valanarasu, J. M. J.; and Patel, V. M. 2022. Unext: Mlp-based rapid medical image segmentation network. In *Proceeding of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 23–33. Springer.

Van Tulder, G.; Tong, Y.; and Marchiori, E. 2021. Multi-view analysis of unregistered medical images using cross-view transformers. In *MICCAI 2021*, volume 12903 of *LNCS*, 104–113. Cham: Springer.

Wang, C.; Li, J.; Zhang, F.; Sun, X.; Dong, H.; Yu, Y.; and Wang, Y. 2021. Bilateral asymmetry guided counterfactual generating network for mammogram classification. *IEEE Transactions on Image Processing*, 30: 7980–7994.

Wang, Z.; Xian, J.; Liu, K.; Li, X.; Li, Q.; and Yang, X. 2023. Dual-view Correlation Hybrid Attention Network for Robust Holistic Mammogram Classification. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, 1515–1523. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Wu, N.; Phang, J.; Park, J.; Shen, Y.; Huang, Z.; Zorin, M.; Jastrzkbski, S.; Févry, T.; Katsnelson, J.; Kim, E.; et al. 2019. Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Transactions on Medical Imaging*, 39(4): 1184–1194.

Wu, N. e. a. 2020. Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Trans. Med. Imaging*, 39: 1184–1194.

Yang, Z.; Cao, Z.; Zhang, Y.; Tang, Y.; Lin, X.; Ouyang, R.; Wu, M.; Han, M.; Xiao, J.; Huang, L.; et al. 2021. MommiNet-v2: Mammographic multi-view mass identification networks. *Medical Image Analysis*, 73: 102204.

Zhao, X.; Yu, L.; and Wang, X. 2020. Cross-view attention network for breast cancer screening from multi-view mammograms. In *Proceeding of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1050–1054. Barcelona: IEEE.

Zhou, Z.; Siddiquee, M. M. R.; Tajbakhsh, N.; and Liang, J. 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6): 1856–1867.

Zhu, W.; Lou, Q.; Vang, Y. S.; and Xie, X. 2017. Deep multi-instance networks with sparse label assignment for whole mammogram classification. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, 603–611. Springer.