Enhancing Decision Boundaries for Mammographic BIRADS Classification using a Weighted Double Constraint Classifier

Oğuz DENİZ
Artificial Intelligence Engineering
TOBB ETU
Ankara, Türkiye
o.deniz@etu.edu.tr

***Abstract*—Mammographic image classification is critical for the early detection of breast cancer. This study presents a novel approach for the binary classification of mammograms into benign and malignant cases by leveraging a dual-classifier architecture with a weighted double constraint (WDC) loss function. The methodology utilizes the Ministry of Health of the Republic of Türkiye's expert-labeled mammography dataset, which comprises DICOM images of RCC/RMLO and LCC/LMLO views. Initially, the high-resolution images are converted to PNG format using the pydicom library. Subsequent preprocessing steps include breast region extraction using a pretrained YOLO model and contrast enhancement through the application of the CLAHE filter. Six state-of-the-art backbone configurations—EfficientNetb0, EfficientNetb3, ResNet18, ResNet50, DenseNet121, and DenseNet169—are employed with both a conventional classifier head and the proposed WDC classifier. The dual-classifier mechanism is designed to adjust decision boundaries dynamically by applying a standard binary cross-entropy (BCE) loss when the classifiers agree, and a modified loss function to better distinguish hard-to-classify samples when they diverge. Experimental results obtained from 12 different models demonstrate the effectiveness of the proposed approach, potentially paving the way for more robust mammographic diagnostic tools.**

***Index Terms*—Mammographic Image Classification, BIRADS, Dual Classifier, Weighted Loss Function, Deep Learning, Breast Cancer Detection**

I. Introduction

Breast cancer remains a leading cause of mortality among women globally, underscoring the critical importance of early and accurate detection. Mammography is a widely adopted and effective screening modality for breast cancer, and the Breast Imaging Reporting and Data System (BIRADS) provides a standardized framework for radiologists to categorize findings, assess risk, and guide patient management. Accurate classification of mammographic images according to the BIRADS system is crucial for timely diagnosis and treatment.

Despite advancements in medical imaging and computer-aided diagnosis, challenges persist in accurately classifying mammographic images, particularly in distinguishing between benign and malignant lesions in complex cases. Early and accurate detection of breast cancer is fundamental to reducing mortality and improving treatment outcomes. Mammography remains the most effective screening modality; however, interpreting these images poses challenges due to the subtle differences between benign and malignant findings. The approach detailed in this paper addresses these challenges by formulating the problem as a binary classification task that distinguishes between benign (BIRADS1 and BIRADS2) and malignant (BIRADS4 and BIRADS5) cases.

Advances in deep learning have significantly contributed to medical image analysis, particularly through the use of convolutional neural networks (CNNs) that excel in feature extraction. This study extends these advancements by introducing a weighted double constraint classifier (WDCC) that incorporates a dual-classifier architecture. When both classifiers produce the same prediction, a standard binary cross-entropy (BCE) loss is applied; when predictions differ, a modified loss function is activated to enhance the discrimination of challenging samples, thereby refining the system's decision boundaries.

The dataset employed originates from the Ministry of Health of the Republic of Türkiye and consists of 5439 mammographic images from 2721 patients. Each patient's data includes paired views (RCC/RMLO or LCC/LMLO), stored in DICOM format and curated by expert radiologists. To prevent potential data leakage, the dataset is partitioned into training, validation, and test sets based on patient IDs, ensuring that images from a single patient are confined to one subset.

Preprocessing plays a crucial role in this methodology. DICOM images are first converted to PNG format via the pydicom library. A pretrained YOLO model is then employed to crop the breast region from the high-resolution images (approximately 4000×4000 pixels), followed by the application of the Contrast Limited Adaptive Histogram Equalization (CLAHE) filter to enhance features such as calcifications and tumors.

The performance of the approach is evaluated using six widely adopted backbone networks—EfficientNetb0, EfficientNetb3, ResNet18, ResNet50, DenseNet121, and DenseNet169—each tested with both a standard classifier and the WDC classifier, resulting in a total of 12 model configurations.

II. Related Works

Mammography remains unavoidable for early breast cancer detection, yet its interpretation is inherently challenging. Computer-Aided Diagnosis (CAD) systems have increasingly relied on deep learning techniques to support radiologists by automating and enhancing the diagnostic process [1], [2]. In particular, accurate BIRADS classification is essential for guiding treatment decisions.

Recent research has demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in mammographic image analysis. CNN architectures such as EfficientNet, ResNet, and DenseNet have been widely adopted due to their ability to learn complex imaging features. For instance, EfficientNet variants have achieved high F1 scores in malignant classification [3], while ResNet models, especially when combined with contrast enhancement methods like CLAHE, have demonstrated impressive accuracy in abnormality detection [4]. Similarly, DenseNet-based approaches have shown robust performance in BIRADS-based breast density assessment [5].

Loss function design has also received significant attention, particularly in addressing class imbalances common in mammography datasets. Although the Binary Cross-Entropy (BCE) loss is the standard, its limitations have led to the exploration of alternatives such as Focal Loss and Weighted BCE, which have improved the classification performance on minority classes [6].

Preprocessing techniques have proven critical to enhancing model performance. Techniques such as breast region extraction using models like YOLO [7] and contrast enhancement via the CLAHE filter have been shown to improve the quality of input images significantly. In parallel, multi-view analysis methods have been explored to integrate information from different imaging angles, thereby boosting classification accuracy [8], [9].

More recent efforts have focused on improving decision boundaries through dual-classifier architectures. For example, WDCCNet employs a weighted double-constraint strategy to refine decision boundaries and enhance feature discrimination, resulting in improved performance for mammographic image classification [10]. Complementary approaches have also investigated the use of nearest neighbors to adjust classifier boundaries dynamically [11].

The availability of large-scale mammography datasets, such as those provided by Mammo-Bench and VinDr-Mammo, has further accelerated progress in this field by enabling the training of more robust models [12], [13]. Together, these advancements form the foundation for ongoing research aimed at developing more reliable CAD systems and improving the accuracy of BIRADS classification.

III. Methods

A thorough methodology was established to ensure high-quality data and robust model training for the binary classification of mammographic images. The approach encompasses exploratory data analysis, a comprehensive data preprocessing pipeline, and model training using both standard and weighted double constraint classifiers.

A. Exploratory Data Analysis (EDA)

The dataset comprises 5439 mammographic images from 2721 patients, collected from the Ministry of Health of the Republic of Türkiye. Each patient folder contains two DICOM images representing either the RCC/RMLO or LCC/LMLO views. The images are labeled with BIRADS classes 1, 2, 4, and 5.

The dataset exhibits a balanced distribution of imaging views (RCC/RMLO and LCC/LMLO). However, the BIRADS 1 category is significantly larger, containing twice the number of samples as all other BIRADS categories combined. Categories BIRADS 2, 4, and 5 are evenly distributed.

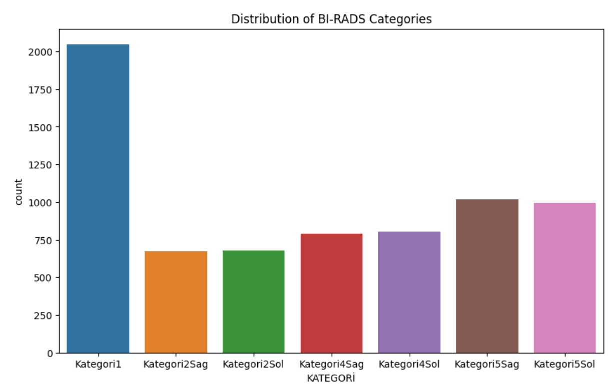


Fig. 1: BIRADS Category Distribution of Dataset

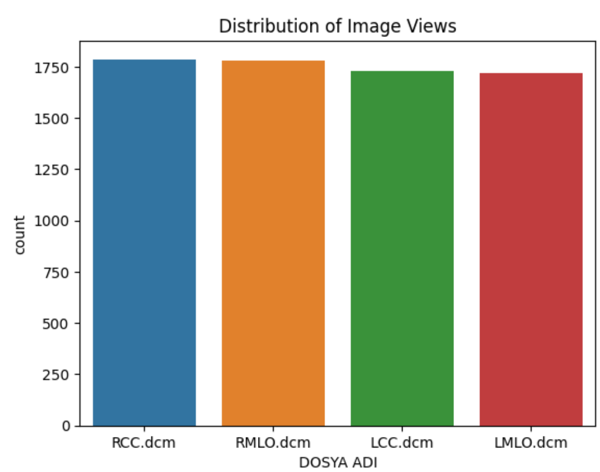


Fig. 2: Imaging View Distribution of Dataset

B. Data Preprocessing Pipeline

A series of preprocessing steps were implemented to ensure high-quality input for the classification models:

- **DICOM to PNG Conversion:** DICOM images were converted to PNG format using the pydicom library. This conversion facilitated further processing and compatibility with common deep learning frameworks.
- **Breast Region Extraction:** Given the high resolution (approximately 4000×4000 pixels) of the original images, a pretrained YOLO model was employed to automatically crop the breast region. This step effectively removed irrelevant background areas and focused on the region of interest.
- **Contrast Enhancement:** The Contrast Limited Adaptive Histogram Equalization (CLAHE) [14] filter was applied to the cropped images. This technique enhanced the visibility of discriminative features such as veins, tumors, and calcifications, thereby improving the quality of the input data.
- **Dataset Splitting:** To avoid patient-level data leakage, the dataset was partitioned into training, validation, and test sets using an 80:10:10 ratio based on patient IDs. This ensured that images from the same patient did not appear across different subsets.

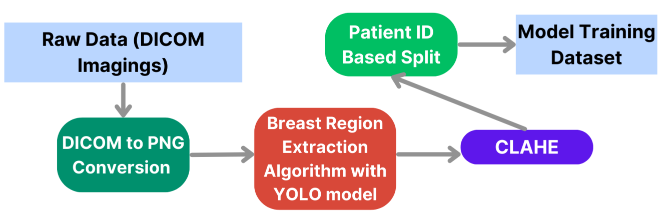


Fig. 3: Data Preprocessing Pipeline Workflow Diagram

The preprocessing pipeline transforms raw DICOM images into refined, high-quality inputs for deep learning analysis. By converting images to PNG, isolating the breast region with a pretrained YOLO model, enhancing contrast using CLAHE, and enforcing patient-level splits to prevent data leakage, the process standardizes and optimizes the dataset.

These steps collectively establish a robust foundation for model training, ensuring that the classifiers are developed on representative, high-fidelity data that is crucial for accurate benign versus malignant differentiation.

C. Model Training

The experimental framework involves training two distinct classifier architectures across three widely adopted backbone networks in mammographic image classification. These backbones were chosen for their proven capability in capturing complex features from medical images, and they include **EfficientNet, ResNet**, and **DenseNet**.

For each backbone family, two configurations were implemented to explore the impact of network complexity and scaling on classification performance. Specifically, **EfficientNetb0** and **EfficientNetb3** represent lightweight and moderately scaled models respectively; **ResNet18** and **ResNet50** offer variations in depth and complexity; and **DenseNet121** and **DenseNet169** provide different levels of feature reuse and information flow within the network.

For every backbone configuration, two separate classifier heads were designed and trained. The first is a **Normal Classifier**, which utilizes a conventional classifier head optimized with the binary cross-entropy (BCE) loss function—a standard approach for binary classification tasks. The second is a **Weighted Double Constraint (WDC) Classifier** from [10] that integrates a dual-classifier structure.

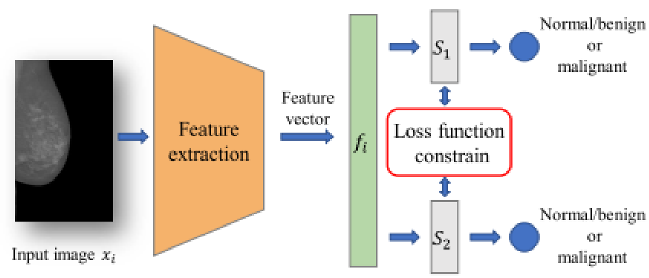


Fig. 4: Visualization of the WDC Classifier Architecture from WDCCNet Paper [10]

This architecture features two identical classifier branches that process features from the backbone network. When both branches agree on a prediction, conventional BCE loss is applied. If their predictions differ, a weighted loss function activates, dynamically adjusting decision boundaries to emphasize challenging cases and foster a more discriminative feature space between benign and malignant samples.

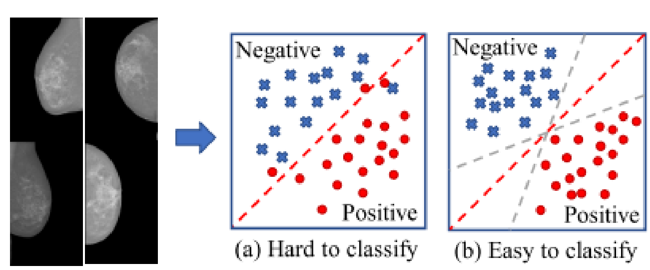


Fig. 5: Visualization of the Decision Boundaries from WDC-CNet Paper [10]

The systematic training of these models under identical conditions across 12 configurations (derived from 6 backbone variants multiplied by 2 classifier types) ensures a fair and robust comparison. This setup is critical for evaluating the relative strengths of the normal and WDC classifier heads in the context of mammographic image classification.

*1) **Training Procedure**:* Pretrained weights accelerate convergence and improve feature extraction performance in medical imaging tasks. The training setup employs a learning rate of 0.0001, the Adam optimizer, and a total of 10 epochs. A batch size of 8 is maintained for most configurations, with a reduced size of 4 for larger backbones to prevent out-of-memory errors.

Data augmentation plays a crucial role in enhancing model robustness. For the training dataset, the following transformations are applied: ColorJitter to simulate variations in imaging conditions, resizing to (520,520) pixels, RandomCrop to (500,500) pixels to introduce spatial variability, RandomRotation up to 30 degrees, and RandomHorizontalFlip. These augmentations aim to increase the diversity of training samples and improve generalization.

Given that each patient provides two different mammographic images, a **double fusion** strategy is employed during feature extraction. Specifically, features are independently extracted from each image and then concatenated before being passed to the classifier head. This fusion ensures that the model leverages complementary information from both views for a more robust prediction.

*2) **Model Architectures**:*

- **Normal Classifier:** A conventional classifier head with a single layer of 1024 neurons, optimized using the standard binary cross-entropy (BCE) loss.
- **Weighted Double Constraint (WDC) Classifier:** A dual-classifier structure with an identical classifier head (also comprising one 1024-neuron layer) in each branch. The same feature fusion strategy is applied prior to classification.

This experimental design results in a total of 12 models (6 backbone configurations × 2 classifier types), all trained under uniform conditions to enable a fair and direct comparison of performance in benign versus malignant classification.

D. Weighted Double Constraint Classifier Architecture

The Weighted Double Constraint (WDC) classifier is specifically designed to enhance performance on challenging cases by dynamically adjusting decision boundaries. In this dual-classifier framework, features extracted from the two mammographic images (after concatenation) are fed simultaneously into two identical classifier branches. Both branches consist of a fully connected layer with 1024 neurons.

When both classifiers in the WDC structure produce the same prediction, the system applies the standard BCE loss, treating the sample as straightforward. In instances where the predictions diverge, a modified weighted loss function is activated. This function imposes a higher penalty on discrepancies, compelling the network to focus on learning more distinctive features that differentiate between benign and malignant cases.

$$l_{1,i} = \begin{cases} -\log \dfrac{e^{s(m\cos(\theta_{1,i}))}}{e^{s(m\cos(\theta_{1,i}))} + e^{s(\cos(\theta_{2,i}))}}, & y_i = 0, \\ -\log \dfrac{e^{s(\cos(\theta_{2,i}))}}{e^{s(\cos(\theta_{2,i}))} + e^{s(m\cos(\theta_{1,i}))}}, & y_i = 1, \end{cases}$$

Fig. 6: Loss Function of the 1st Classifier from [10]

$$l_{2,i} = \begin{cases} -\log \dfrac{e^{s(\cos(\theta_{1,i}))}}{e^{s(\cos(\theta_{1,i}))} + e^{s(m\cos(\theta_{2,i}))}}, & y_i = 0, \\ -\log \dfrac{e^{s(m\cos(\theta_{2,i}))}}{e^{s(m\cos(\theta_{2,i}))} + e^{s(\cos(\theta_{1,i}))}}, & y_i = 1. \end{cases}$$

Fig. 7: Loss Function of the 2nd Classifier from [10]

$$\mathcal{L}_{WDCC} = \lambda \mathcal{L}'_1 + (1-\lambda)\,\mathcal{L}'_2.$$

Fig. 8: Combined WDCC Loss Function from [10]

The architecture thus effectively mitigates the limitations of standard loss functions in scenarios with class imbalance and subtle inter-class differences.By integrating a dual fusion strategy, rigorous data augmentation, and the WDC mechanism, the model training process is tailored to harness complementary information from dual-view inputs and address the complexities inherent in mammographic image classification.

IV. Evaluation

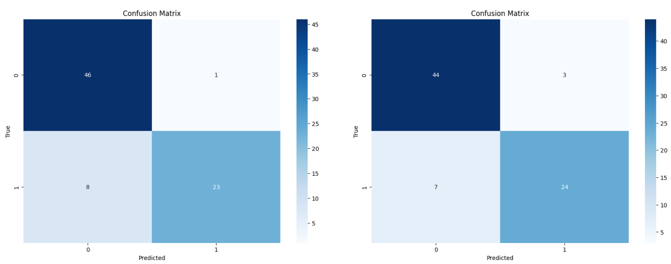
The evaluation of the models was conducted on a fixed test set comprising 78 samples, ensuring that each patient's two mammographic images contributed to the final prediction through the proposed double fusion strategy. The performance of two classifier head architectures—a Normal Classifier and the proposed Weighted Double Constraint (WDC) Classifier—was compared across six backbone configurations: EfficientNet-b0, EfficientNet-b3, ResNet-18, ResNet-50, DenseNet121, and DenseNet161. All models were trained under identical settings (pretrained weights, a learning rate of 0.0001, Adam optimizer, 10 epochs, and a batch size of 8 or 4 depending on backbone size) with identical data augmentation pipelines.

A. Quantitative Results

*1) **EfficientNet Architectures**:*

- **EfficientNet-b0 with WDC Classifier:** The WDC classifier achieved an overall accuracy of 88.46% with a precision of 0.85 and 0.96 for the benign and malignant classes, respectively. The f1-scores were 0.91 for benign and 0.84 for malignant cases, reflecting robust performance in handling hard-to-classify samples.

- **EfficientNet-b3 with WDC Classifier:** Results were similar, with an overall accuracy of 87.18%, and slightly lower f1-scores (0.89 for benign and 0.81 for malignant), indicating consistent performance across EfficientNet variants.
- **Normal Classifier on EfficientNet-b0 and b3:** For EfficientNet-b0, the Normal Classifier yielded an overall accuracy of 87.18%, with 91.49% accuracy on benign cases and 80.65% on malignant cases. Similarly, EfficientNet-b3 with the Normal Classifier reported an overall accuracy of 88.46%, achieving a high benign class accuracy (95.74%) but a lower malignant class accuracy (77.42%).



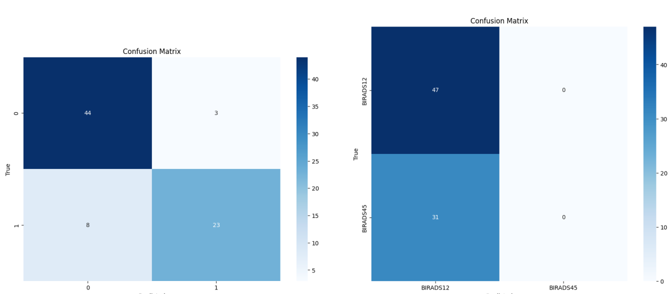
(a) EfficientNet-b0 with WDC Head Results

(b) EfficientNet-b3 with WDC Head Results

Fig. 9: EfficientNet Architectures with WDC Head Results

2) *ResNet Architectures:*

- **ResNet-18 and ResNet-50 with WDC Classifier:** Both configurations achieved an overall accuracy of 86% with similar class-wise performances (approximately 85.90% accuracy for benign predictions).
- **Normal Classifier on ResNet-18 and ResNet-50:** In contrast, the Normal Classifier failed to learn from the ResNet-based configurations. Both ResNet-18 and ResNet-50 with the Normal Classifier resulted in an overall accuracy of 60.26%, where benign cases were perfectly classified (100% accuracy) while malignant cases were not recognized at all (0% accuracy).



(a) ResNet18 with WDC Head Results

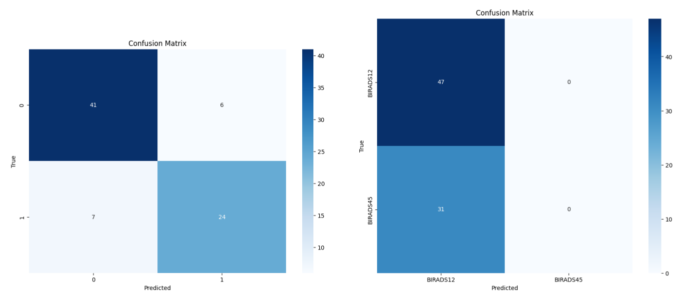
(b) ResNet18 with Normal Head Results

Fig. 10: ResNet Architectures with WDC and Normal Head Results

These results demonstrate that the WDC mechanism helps the model learn better decision boundaries even in relatively deep networks. The clear imbalance highlights the limitation of the standard BCE loss in handling hard-to-classify samples.

3) *DenseNet Architectures:*

- **DenseNet121 with WDC Head:** The WDC classifier on DenseNet121 achieved an overall accuracy of 85%, with an 84.62% class-wise accuracy for benign cases, reflecting balanced performance.
- **DenseNet161 with WDC Head:** A slightly lower overall accuracy of 83% was observed with DenseNet161, with a benign class accuracy of 83.33%.
- **Normal Classifier on DenseNet Architectures:** Similar to the ResNet configurations, DenseNet models trained with the Normal Classifier produced poor results, with an overall accuracy of 60.26%, 100% accuracy on benign cases, and 0% on malignant cases.



(a) DenseNet161 with WDC Head Results

(b) DenseNet161 with Normal Head Results

Fig. 11: DenseNet Architectures with WDC and Normal Head Results

B. *Comparison of Classifier Heads*

The evaluation clearly indicates that the Normal Classifier struggles with the inherent challenges of the dataset, particularly in distinguishing malignant cases. Across ResNet and DenseNet configurations, the standard approach resulted in complete failure to classify malignant cases despite high benign classification accuracy. This behavior is typical when decision boundaries are not sufficiently adjusted to handle hard-to-classify samples.

In contrast, the Weighted Double Constraint (WDC) Classifier consistently outperformed the Normal Classifier across all backbone architectures. By incorporating a dual-classifier mechanism and applying a specialized weighted loss function in cases of prediction divergence, the WDC head was able to dynamically adjust decision boundaries. This resulted in a more balanced classification performance and significantly improved recognition of malignant cases, as reflected in higher f1-scores and overall accuracies.

V. Conclusion

A robust framework for mammographic image classification was developed and evaluated using a dual-classifier architecture with a Weighted Double Constraint (WDC) mechanism. The study addressed the challenges inherent in binary classification of benign and malignant cases by employing extensive data preprocessing—including DICOM-to-PNG conversion, breast region extraction using a pretrained YOLO model, and contrast enhancement with CLAHE—to optimize the quality of the inputs. A double fusion strategy was adopted to integrate complementary features from dual-view images, thereby enhancing the overall information available for classification.

The experimental evaluation, conducted across 12 model configurations utilizing EfficientNet, ResNet, and DenseNet backbones, demonstrated that standard classifier heads, optimized with binary cross-entropy loss, struggle to adequately learn decision boundaries, particularly for malignant cases. In contrast, the WDC classifier consistently achieved superior performance. For instance, EfficientNet models with the WDC head reached accuracies as high as 88.46%, while ResNet and DenseNet architectures trained with conventional classifiers failed to recognize malignant cases, resulting in significantly lower overall accuracies. These results underscore the effectiveness of the dual-classifier strategy and adaptive loss function in addressing class imbalance and hard-to-classify samples.

Overall, the findings suggest that integrating a weighted double constraint mechanism can substantially improve classification performance in challenging medical imaging tasks. The approach not only refines decision boundaries but also provides a more balanced performance across classes, making it a promising solution for computer-aided diagnosis in mammography. Future work may explore further optimization of loss functions and the integration of additional imaging modalities to continue advancing the diagnostic accuracy and clinical applicability of such systems.The code is available on GitHub repository: GitHub

References

[1] Abdel-Zaher, A. M., and H. M. Abdel-Kader. ”Breast Cancer Classification Using Deep Learning.” Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 33, no. 1, 2020, pp. 135-46, doi:10.1007/s10334-019-00799-1.

[2] Ragab, Dalia A., et al. ”Deep Convolutional Neural Networks for Computer-Aided Detection of Breast Cancer in Mammography: Current Status and Future Challenges.” Diagnostics, vol. 10, no. 12, 2020, p. 988, doi:10.3390/diagnostics10120988.

[3] Afzal, Farheen, et al. ”Classification of BI-RADS Using Convolutional Neural Network and EfficientNet-B7.” International Journal of Science and Research Archive, vol. 11, no. 1, 2024, pp. 1022-28. Bhole, Gaurav, et al. ”Mammo-Bench: A Large-Scale Benchmark Dataset of Mammography Images.” medRxiv, 2025, doi:10.1101/2025.01.31.25321510.

[4] Li, Hong, et al. ”Benign and Malignant Classification of Mammogram Images Based on Deep Learning.” Biomedical Signal Processing and Control, vol. 51, 2019, pp. 347-54, doi:10.1016/j.bspc.2019.03.005.

[5] Parekh, Nita, et al. ”Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection.” Diagnostics, vol. 12, no. 5, 2022, p. 1147, doi:10.3390/diagnostics12051147.

[6] Scholz, D., et al. ”Imbalance-Aware Loss Functions Improve Medical Image Classification.” Proceedings of the 7th International Conference on Medical Imaging with Deep Learning, 2024, pp. 1341-56.

[7] Su, Luhao, et al. ”YOLO-LOGO: A Transformer-Based YOLO Segmentation Model for Breast Mass Detection and Segmentation in Digital Mammograms.” Computer Methods and Programs in Biomedicine, vol. 221, 2022, p. 106903, doi:10.1016/j.cmpb.2022.106903.

[8] Khan, Hafiz Ali, et al. ”Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network.” IEEE Access, vol. 7, 2019, pp. 165724-33, doi:10.1109/access.2019.2953488.

[9] Nguyen, Huyen T. X., et al. ”A Novel Multi-View Deep Learning Approach for BI-RADS and Density Assessment of Mammograms.” arXiv, 2021, doi:10.48550/arXiv.2112.04490.

[10] Wang, Yan, et al. ”WDCCNet: Weighted Double-Classifier Constraint Neural Network for Mammographic Image Classification.” IEEE Transactions on Medical Imaging, vol. 41, no. 3, 2022, pp. 559-70, doi:10.1109/tmi.2021.3117272.

[11] Ortiz-Jimenez, et al. ”Improving Classifier Decision Boundaries Using Nearest Neighbors.” arXiv, 2023, doi:10.48550/arXiv.2310.03927.

[12] Bhole, Gaurav, et al. ”Mammo-Bench: A Large-Scale Benchmark Dataset of Mammography Images.” medRxiv, 2025, doi:10.1101/2025.01.31.25321510.

[13] Pham, H. H., et al. ”VinDr-Mammo: A Large-Scale Benchmark Dataset for Computer-Aided Detection and Diagnosis in Full-Field Digital Mammography.” PhysioNet, 2022, doi:10.13026/br2v-7517.

[14] Pisano, E.D., Zong, S., Hemminger, B.M. et al. Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms. J Digit Imaging 11, 193 (1998). https://doi.org/10.1007/BF03178082