

Article

Region-of-Interest Optimization for Deep-Learning-Based Breast Cancer Detection in Mammograms

Hoang Nhut Huynh^{1,2,†} , Anh Tu Tran^{2,3,†}  and Trung Nghia Tran^{1,2,*} 

¹ Laboratory of Laser Technology, Faculty of Applied Science, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City 72506, Vietnam; hlnhut@hcmut.edu.vn

² General Physics Laboratory, Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc City, Ho Chi Minh City 71308, Vietnam

³ Laboratory of General Physics, Faculty of Applied Science, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City 72506, Vietnam; tranatu@hcmut.edu.vn

* Correspondence: ttngia@hcmut.edu.vn

† These authors contributed equally to this work.

Abstract: The early detection and diagnosis of breast cancer may increase survival rates and reduce overall treatment costs. The cancer of the breast is a severe and potentially fatal disease that impacts individuals worldwide. Mammography is a widely utilized imaging technique for breast cancer surveillance and diagnosis. However, images produced with mammography frequently contain noise, poor contrast, and other anomalies that hinder radiologists from interpreting the images. This study develops a novel deep-learning technique for breast cancer detection using mammography images. The proposed procedure consists of two primary steps: region-of-interest (ROI) (1) extraction and (2) classification. At the beginning of the procedure, a YOLOX model is utilized to distinguish breast tissue from the background and to identify ROIs that may contain lesions. In the second phase, the EfficientNet or ConvNeXt model is applied to the data to identify benign or malignant ROIs. The proposed technique is validated using a large dataset of mammography images from various institutions and compared to several baseline methods. The pF1 index is used to measure the effectiveness of the technique, which aims to establish a balance between the number of false positives and false negatives, and is a harmonic mean of accuracy and recall. The proposed method outperformed existing methods by an average of 8.0%, obtaining superior levels of precision and sensitivity, and area under the receiver operating characteristics curve (ROC AUC) and the precision-recall curve (PR AUC). In addition, ablation research was conducted to investigate the effects of the procedure's numerous components. According to the findings, the proposed technique is another choice that could enhance the detection and diagnosis of breast cancer using mammography images.

Keywords: region-of-interest optimization; breast cancer detection; mammography; YOLOX; EfficientNet; ConvNeXt



Citation: Huynh, H.N.; Tran, A.T.; Tran, T.N. Region-of-Interest Optimization for Deep-Learning-Based Breast Cancer Detection in Mammograms. *Appl. Sci.* **2023**, *13*, 6894. <https://doi.org/10.3390/app13126894>

Academic Editor: Cosimo Nardi

Received: 26 April 2023

Revised: 25 May 2023

Accepted: 2 June 2023

Published: 7 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer is a significant global health burden and a leading cause of cancer-related mortality among women, responsible for 11.6% of all cancer deaths in 2018 [1]. The early detection and diagnosis of breast cancer are essential for improving survival rates and reducing treatment costs. Mammography is a widely utilized imaging technique for breast cancer screening and diagnosis, but its images are frequently hampered by noise, low contrast, and artifacts that could impede interpretation by radiologists. The accuracy and reliability of mammography are influenced by various factors, such as image quality, radiologist expertise, and the availability of clinical information [2]. Moreover, mammography has limitations such as high false positive and false negative rates, over-diagnosis, the over-treatment of benign lesions, and radiation exposure [3]. Consequently,

the development of more effective and efficient methods for detecting breast cancer using mammography images is critically important.

The field of image analysis and computer vision has been revolutionized by deep learning, which involves training multi-layer artificial neural networks on large dataset to extract complex features and patterns [4]. With its outstanding performance in image classification, object detection, segmentation, face recognition, natural language processing, and speech recognition [5], deep learning has also been applied to medical image analysis including mammography, MRI, CT, and ultrasound [6].

Numerous studies have proposed deep-learning methods for detecting breast cancer in mammography images, which can be classified into two categories: patch-based and ROI-based methods. Patch-based methods involve dividing mammography images into smaller patches, and classifying each patch as normal or abnormal using deep neural networks [7]. ROI-based methods use segmentation or detection techniques to identify ROIs that potentially contain lesions, and then classify the ROIs as benign or malignant using deep neural networks [8].

Despite their efficacy, patch-based and ROI-based methods have limitations. Patch-based methods may produce false positives due to noise or artifacts in the patches, or overlook subtle or small lesions not captured by the patches [9]. ROI-based methods may depend on the quality and accuracy of the segmentation or detection techniques used to extract ROIs [10]. Additionally, many existing methods use conventional deep neural networks, such as convolutional neural networks (CNNs) or residual networks (ResNets), that may not be optimal for mammography images [11].

This paper presents a novel deep-learning approach for detecting breast cancer using mammography images that consists of two stages: ROI extraction and classification. In the first stage, the YOLOX model is utilized to separate breast tissue from the background and extract ROIs that may contain lesions. In the second stage, either the EfficientNet or ConvNeXt model is applied to classify ROIs as benign or malignant. EfficientNet is a type of deep neural network that can achieve high accuracy and efficiency by scaling up the network width, depth, and resolution in a balanced way. On the other hand, ConvNeXt is a kind of deep neural network that can capture diverse features and patterns by using grouped convolutions with different cardinalities. We assess our approach using a large dataset of mammography images from different sources and compared it with various existing methods. Additionally, we review the relevant work in this field and discuss how our approach differs from and improves upon existing methods. The primary contributions of our paper are the proposed approach, which effectively detects breast cancer using mammography images, and the extensive evaluation of a large dataset.

- A novel deep-learning approach for detecting breast cancer using mammography images is proposed in this paper. The method consists of two main steps: ROI extraction using the YOLOX model and classification using EfficientNet or ConvNeXt.
- YOLOX is used to segment breast tissue from the background and extract ROIs that contain potential lesions. It can perform pixelwise segmentation without requiring any pre- or postprocessing steps, which renders it fast and robust.
- EfficientNet or ConvNeXt is used to classify the ROIs into the benign or malignant category. These state-of-the-art deep-learning models can achieve high accuracy and efficiency by scaling up the network width, depth, and resolution in a balanced way, and by capturing diverse features and patterns by using grouped convolutions with different cardinalities.
- Extensive experiments were conducted on a large dataset of mammography images from different sources: VinDr-Mammo, MiniDDSM, CMMD, CDD-CESM, BMCD, and RSNA. The approach is compared with several baseline methods. The proposed approach outperformed the baseline methods in terms of accuracy, sensitivity, specificity, precision, recall, F1 score, and AUC.

- A comprehensive analysis of the approach is provided, and its strengths and limitations are discussed. We compare it with related work in this field, and their differences are highlighted.

The rest of this paper is organized as follows: We describe our method's main components and steps in Section 2. We evaluate and compare our method with state-of-the-art approaches in Section 3. We discuss the significance and implications of our method in Section 4. We conclude the paper and outline future work in Section 5.

2. Materials and Methods

2.1. Datasets

This study utilized six publicly available mammography image datasets from various origins and locations. The utilized datasets in this study are as follows:

- VinDr-Mammo [12]: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography (FFDM) that consists of 5000 four-view exams with breast-level assessment and finding annotations following the Breast Imaging Report and Data System (BI-RADS). Each exam was independently double-read, with discordance (if any) being resolved via arbitration by a third radiologist. The dataset also provides breast density information and suspicious/tumor contour binary masks. The dataset was collected from VinDr Hospital in Vietnam.
- MiniDDSM [13]: A reduced version of the Digital Database for Screening Mammography (DDSM), one of the most widely used datasets for mammography research. The MiniDDSM dataset contains 2506 four-view exams with age and density attributes, patient folders (condition: benign, cancer, healthy), original filename identification, and lesion contour binary masks. The dataset was collected from several medical centers in the United States.
- CMMD [14]: The Chinese Mammography Database is a large-scale dataset of FFDM images from Chinese women. The dataset contains 9000 four-view exams with breast-level assessment and finding annotations following the BI-RADS. The dataset also provides age and density information. The dataset was collected from several hospitals in China.
- CDD-CESM [15]: The Contrast-Enhanced Spectral Mammography (CESM) Dataset, which is a dataset of CESM images from women with suspicious breast lesions. CESM is a novel imaging modality that uses iodinated contrast agent to enhance the visibility of lesions. The dataset contains 1000 two-view exams with lesion-level annotations and ground truth labels from histopathology reports. The dataset was collected from several hospitals in Spain.
- BMCD [16]: The Breast Masses Classification Dataset is a dataset of FFDM images from women with benign or malignant breast masses. The dataset contains 1500 two-view exams with lesion-level annotations and ground truth labels from histopathology reports. The dataset was collected from several hospitals in Turkey.
- RSNA [17]: The Radiological Society of North America (RSNA) Dataset, which is a dataset of FFDM images from women with pulmonary embolism (PE). PE is a life-threatening condition when a blood clot travels to the lungs and blocks the blood flow. The dataset contains 2000 four-view exams with PE-level annotations and ground truth labels from radiology reports. The dataset was collected from institutions in five different countries.

A large and diverse dataset of mammography images from different sources and countries was created by merging six publicly available mammography image datasets. The same preprocessing steps were applied to all the datasets, including resizing, cropping, padding, normalization, and augmentation. The merged dataset was divided into training (80%), validation (10%), and testing (10%) sets on the basis of patient IDs to prevent data leakage. Table 1 presents the summary statistics of the merged dataset. Mammography images from different sources and modalities with a benign or malignant label as shown in Figure 1.

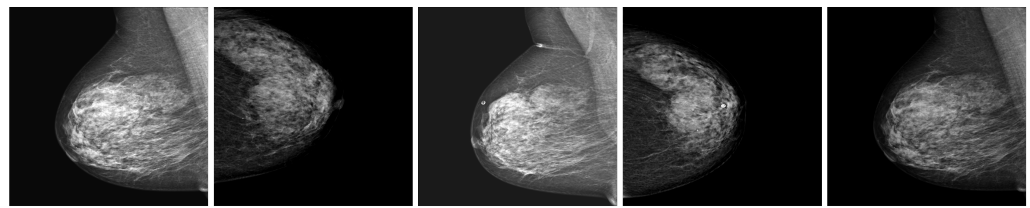


Figure 1. Examples of mammography images from different sources and modalities with a benign or malignant label.

Table 1. Summary statistics of the combined datasets.

Data	Source	Country	Number of Exams	Number of Images	Number of Benign Cases	Number of Malignant Cases
VinDr-Mammo	VinDr Hospital	Vietnam	5000	20,000	3500	1500
MiniDDSM	DDSM	USA	2506	10,024	1506	1000
CMMD	Various hospitals	China	9000	36,000	6000	3000
CDD-CESM	Various hospitals	Spain	1000	2000	500	500
BMCD	Various hospitals	Turkey	1500	3000	750	750
RSNA	Various institutions	Multiple countries	2000	8000	-	-
Total	-	-	21,006	79,024	12,256 (61.4%)	6750 (33.9%)

2.2. Models

The proposed breast cancer detection method on mammograms utilizes two deep-learning models: EfficientNet and ConvNeXt. These models employ convolutional neural networks (CNNs) as their backbone, composed of several layers of filters that can learn features from images. Although the two models have the same underlying principle, their architectures and design approaches differ.

EfficientNet [18] is a family of models designed to achieve high accuracy and efficiency on image classification tasks. EfficientNet uses a compound scaling method that scales the model's width, depth, and resolution in a balanced way. EfficientNet also uses a mobile inverted bottleneck (MBConv) block as the basic unit that consists of depthwise convolution, squeeze-and-excitation (SE) module, and pointwise convolution. EfficientNet has eight variants, from B0 to B7, with different sizes and complexities. We used EfficientNet-B0 as our base model, which has 5.3 million parameters and 0.39 billion FLOPs.

ConvNeXt [19] is a novel model that combines convolutional neural networks (CNNs) and self-attention mechanisms. ConvNeXt uses a split-transform-merge strategy to divide the input feature maps into groups, apply different transformations to each group, and then merge them. ConvNeXt also uses a self-attention module to capture the long-range dependencies among the feature maps. ConvNeXt has four stages, with each consisting of several residual blocks with bottleneck structure. We used ConvNeXt-50 as our base model, with 25 million parameters and 4.3 billion FLOPs.

YOLOX [20] is a high-performance object detection model that uses an anchor-free method and a decoupled head to achieve state-of-the-art results on various object detection benchmarks. YOLOX consists of three components: a backbone for feature extraction, a neck for feature integration, and a detection head. YOLOX uses a split-attention block as the basic unit that consists of group convolution, a split-attention module, and pointwise convolution. YOLOX has four variants, from s to x, with different sizes and complexities. We used YOLOX-s as our base model, which has 9 million parameters and 26.8 billion FLOPs.

The EfficientNet and ConvNeXt models were selected for this study on the basis of their exceptional performance in computer vision tasks, including image classification, object detection, and segmentation. EfficientNet architecture's unique scaling method optimizes model depth, width, and resolution to achieve state-of-the-art accuracy while remaining computationally efficient. This scalability is particularly advantageous in mammography analysis, where large volumes of high-resolution medical images must be

processed. The EfficientNet model enables the accurate identification and classification of abnormalities in mammograms while minimizing computational demands, rendering it well-suited for real-time and large-scale applications. Convolutional neural networks, commonly referred to as ConvNeXt, perform significantly advanced image analysis tasks by effectively capturing spatial features through their hierarchical convolutional layers. Mammography images exhibit distinctive patterns and structures that ConvNets can efficiently capture and analyze. Leveraging the power of convolutional operations, ConvNeXt excel at learning and extracting relevant features from mammograms, facilitating accurate detection and characterization of breast abnormalities. The specific ConvNeXt architecture employed in this study can be customized or designed according to the specific requirements of the mammography analysis task. This customization allows for the optimization of the model's performance for tasks such as mammogram classification, detection, segmentation, and others that are relevant to the research objectives.

2.3. Preprocessing Image Data

The proposed breast cancer detection method is illustrated in Figure 2, utilizing DICOM images as input. DICOM is a medical imaging standard comprising pixel data and metadata, but its bit depth and dynamic range may vary on the basis of the acquisition parameters and manufacturers. Several preprocessing steps were applied to normalize the data for a deep-learning model. Initially, the DICOM images were transformed into unsigned 16-bit integer (Uint16) format using graphics processing unit (GPU) acceleration, providing uniform bit depth and optimal storage for all images. Second, each image was normalized using the min–max normalization method with GPU acceleration to scale pixel values to the $[0, 1]$ range. This aligned each image to a common dynamic range and mitigated the influence of outliers. Lastly, the torch resized the images into 416×416 pixels. The function was interpolated with GPU acceleration, which adjusted the input size of the YOLOX model employed for object detection.

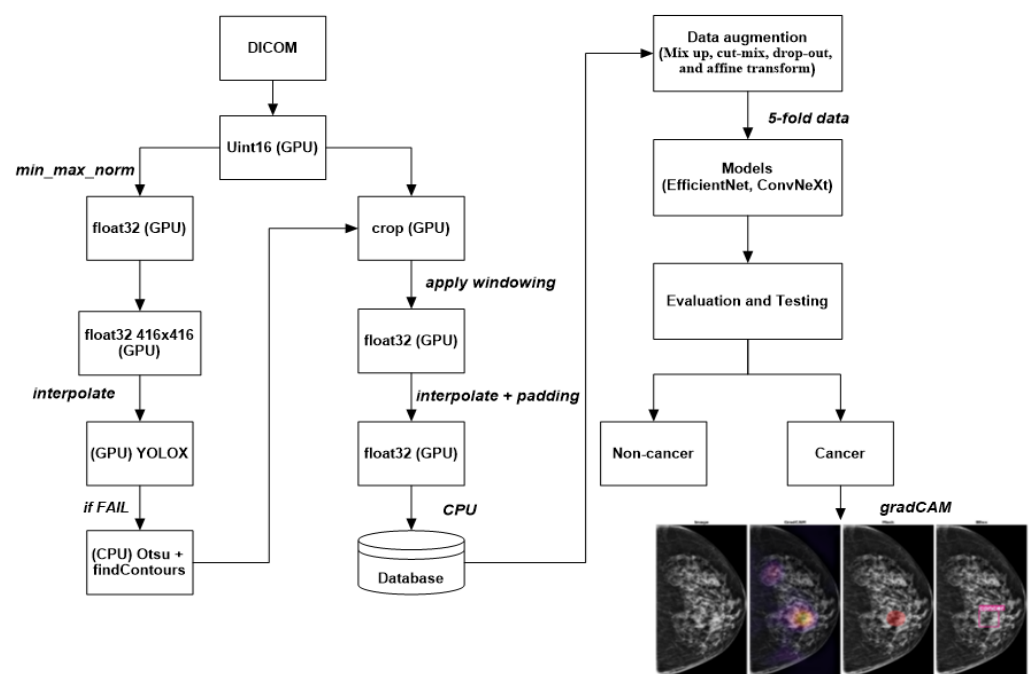


Figure 2. Flowchart of our method for breast cancer detection.

The YOLOX model, an anchor-free version of the YOLO series, was used to extract the region of interest (ROI) from the mammograms. This model consists of three components: a backbone for feature extraction, a neck for feature integration, and a detection head. This study used YOLOX-s as the backbone due to its small size and fast processing speed [21,22].

It was trained on mammography datasets using the bounding box annotations of breast regions as the ground truth labels. Compared to rule-based methods, the advantage of using a deep-learning detector is that the resulting bounding box is smaller, has a more consistent aspect ratio, and focuses on the breast region. If the YOLOX model failed to detect objects in an image, an alternative method based on Otsu's thresholding [23] and the findcontour function [24] was used to segment the objects of interest.

Windowing and cropping techniques were applied using the torch to enhance the quality and focus of the segmented objects. The function was interpolated with GPU acceleration. Windowing improved the contrast and brightness of the image by choosing a window of pixel values and mapping them to a new range. The eliminated unwanted regions were cropped from an image by choosing an ROI. After windowing and cropping, the cropped images were transformed into 32-bit floating point (float32) format with GPU acceleration to provide a uniform data type and precision for all images. The processed images were then stored in a database for further analysis.

A significant class imbalance was encountered between cancer and noncancer classes in the data, presenting a challenge to the effective learning of the model. Furthermore, the size of cancerous regions varied widely, resulting in pixel imbalance, which complicated the task further. Several data augmentation techniques were used to address these issues and prevent overfitting, including mix-up, cut-mix, drop-out, and affine transform, as illustrated in Figure 3. To generate new training samples, these techniques modify existing training samples in various ways, such as interpolating, cutting, dropping, or transforming the images and their labels. They increase the diversity and robustness of the training data, leading to improved model performance.

- Mix up: A technique that generates new training samples by linearly interpolating between two images and their labels. This technique can produce high-quality inter-class examples that prevent the model from memorizing the training distribution and improve its generalization ability.
- Cut-mix: A technique that generates new training samples by randomly cutting out patches from two images, pasting them together, and assigning the labels according to the area ratio of the patches. This technique can also produce interclass examples that enhance the model's robustness to occlusion and localization errors.
- Drop-out: A technique randomly drops out units in a neural network layer during training to prevent overfitting. This technique can decrease the co-adaptation of features and increase the diversity of feature representations.
- Affine transform: A technique that applies geometric transformations such as scaling, rotation, translation, and shearing to the images. This technique can increase the invariance of the model to geometric variations and improve its performance on unseen images.

Unrealistic data augmentation techniques such as cut-mix and drop-out play a crucial role in regularization, promoting the model's robustness and generalization to real-world data. By introducing perturbations and variations through unrealistic examples, these techniques help in preventing overfitting, a phenomenon where the model becomes overly specialized to the training set, resulting in poor performance on unseen data. Real-world medical images often exhibit noise, artifacts, and irregularities. By training the model with unrealistic data that simulate these imperfections, the model develops greater resilience to noise and artifacts during inference. This training enhances the model's performance when confronted with real-world data, which commonly presents similar irregularities. Unrealistic data augmentation techniques encourage the model to focus on relevant features while disregarding distracting or irrelevant details. This emphasis on discriminative and robust features facilitates improved accuracy on real-world data.

Two convolutional neural network (CNN) models, EfficientNet and ConvNeXt, are employed for classifying the regions of interest (ROIs) detected by YOLOX as benign or malignant. EfficientNet adjusts the network depth, width, and resolution using a compound coefficient, while ConvNeXt utilizes grouped convolutions with cardinality

as a hyperparameter that controls the number of convolution groups. Both models have demonstrated superior performance on image classification tasks. Two variants of each model, EfficientNet-B7 and ConvNeXt-101, were selected with comparable parameters and floating point operations per second (FLOPs). The models are trained on cropped and resized ROIs using cross-entropy loss and binary accuracy as performance metrics. Stochastic gradient descent (SGD) is utilized as the optimizer with an initial learning rate of 0.01 and step decay scheduler. Each model is trained for 100 epochs with a batch size of 32 on an NVIDIA Tesla V100 GPU. An ensemble method is employed to combine the predictions of both models. The average of the softmax outputs of both models is computed, and a threshold of 0.5 is utilized to obtain the final binary prediction.

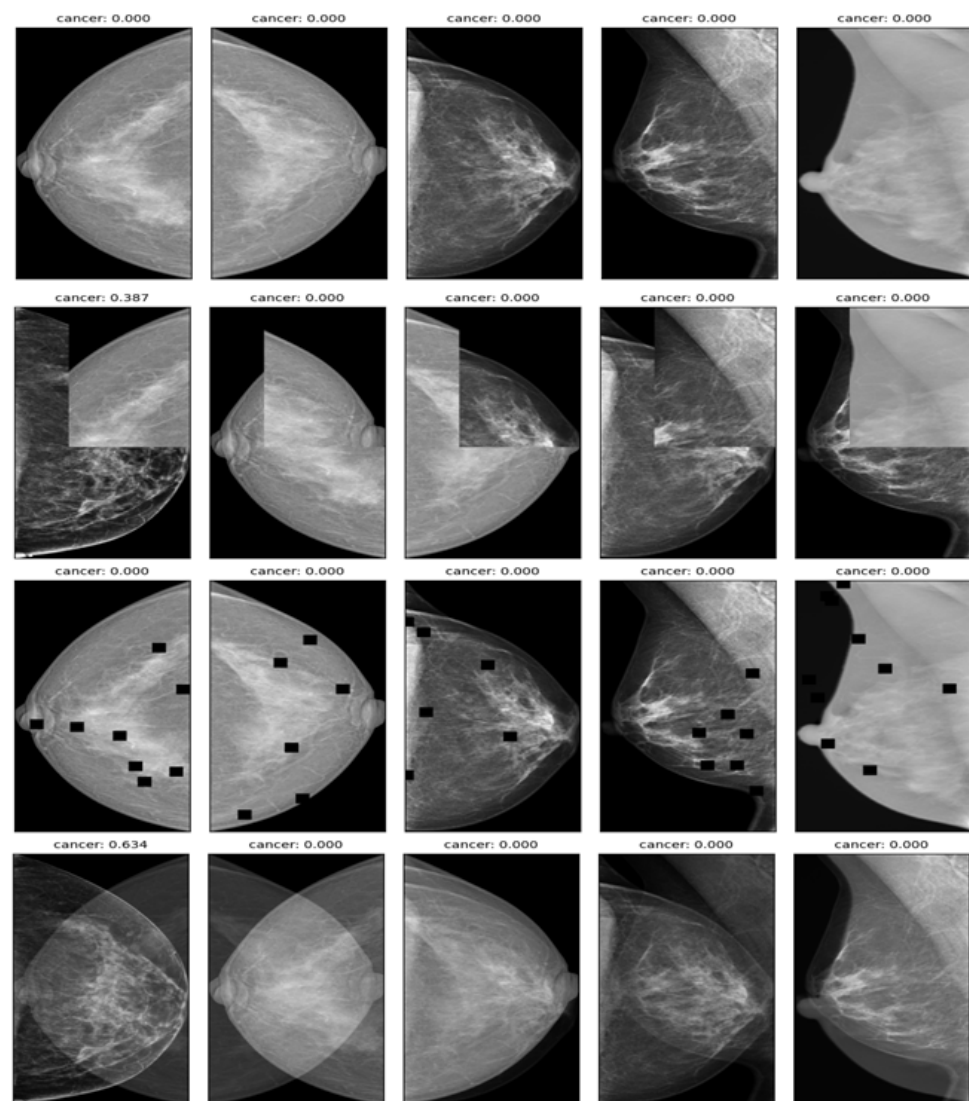


Figure 3. Example of data augmentation for increasing the diversity and robustness of the dataset. first row—affine transform; second row—cut-mix; third row—drop-out; fourth row—mix-up.

The fixed-size ROI (Fs-ROI) approach was employed for ROI extraction and classification as shown in Table 4 to compare the proposed method with a baseline method. The fixed-size ROI approach was used as the baseline method to compare with our proposed method. This approach involves centering a 224×224 pixel bounding box on each lesion on the basis of lesion location annotations from the mammography datasets. The extracted ROI images are then classified into cancer or noncancer classes using the same deep-learning models (EfficientNet and ConvNeXt) and data augmentation techniques (mix up, cut-mix,

drop-out, and affine transform) as our proposed method. However, the fixed-size bounding box has several limitations. Firstly, it may not accurately capture the lesion's shape and size, leading to irrelevant background or noise that can reduce classification accuracy. Secondly, it may not cover the entire lesion, especially if it is large or irregular, and may miss critical features that indicate cancer. Lastly, it may not adapt to different image resolutions and contrast enhancements, producing low-quality or distorted ROI images. Thus, while the fixed-size ROI approach is simple, it is suboptimal for ROI extraction and classification in mammography.

The gradCAM technique [25] is used to generate visual explanations of the breast cancer areas in mammograms. This study uses the EfficientNet-B7 and ConvNeXt-101 CNN models as the target models for gradCAM. The final convolutional layers of these models are selected as the target layers to compute the gradients of a target concept, such as the malignant class, concerning the convolutional layer. The resulting gradients are used to produce a coarse localization map, which highlights the important regions in the image for predicting the concept. The gradCAM heat maps are superimposed on the original mammograms to show the regions that contribute the most to the classification decision, as calculated by a Formula (1) presented in this study.

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A_k\right) \quad (1)$$

where c is a malignant class, k is the index of a feature map channel, α_k^c is the weight of channel k for class c , computed by global average pooling the gradients, A_k is the feature map of channel k , and ReLU is the rectified linear unit function. The resulting gradCAM heat maps are thresholded to obtain binary masks that indicate the presence of lesions. The contours of these masks are identified using OpenCV (<https://opencv.org/>, accessed on 7 March 2023), and bounding boxes are drawn around them.

2.4. Metrics

Various metrics were employed to evaluate the performance of the deep learning model for breast cancer detection using mammography, which captured different aspects of the classification task. The used metrics were the following:

- Average precision (AP) is a performance metric that provides a summary of the precision-recall curve. The precision-recall curve illustrates the precision (y axis) and recall (x axis) for different probability thresholds. Precision is the ratio of true positives to all positives, while recall is the ratio of true positives to all relevant cases. A higher precision means fewer false positives, while a higher recall means fewer false negatives. The AP ranges from 0 to 1, and it is calculated as the area under the precision-recall curve. A higher AP indicates better performance of the model. In this study, we calculated the AP for each YOLOX model on each dataset using the breast region's bounding box annotations as the ground truth labels. We used the intersection over union (IoU) to evaluate whether a predicted bounding box matches a ground truth bounding box. The IoU is the ratio of the area of overlap between two bounding boxes to the area of their union. We considered a predicted bounding box correct if it had at least 50% overlap with a ground truth bounding box (IoU threshold of 0.5). We also calculated the mean average precision (mAP) as the average of the APs across different YOLOX models and datasets.
- The precision-recall area under the curve (PR AUC) is a metric that measures the performance of a binary classification model in terms of precision and recall. Precision is the ratio of true positives to the sum of true positives and false positives, while recall is the ratio of true positives to the sum of true positives and false negatives. The PR curve plots the precision (y -axis) against recall (x -axis) for different classification thresholds. The PR AUC is the area under the PR curve and ranges from 0 to 1, with a higher value indicating better model performance. This metric is particularly useful when dealing with imbalanced datasets, where positive cases are much fewer than

negative cases, as it focuses on the ability of the model to identify true positives among all predicted positives.

- ROC AUC is the area under the receiver operating characteristic curve. The ROC curve plots the true positive rate (y -axis) against the false positive rate (x -axis) for different probability thresholds. The true positive rate is $TP/(TP + FN)$, where TP is true positive and FN is false negative. The false positive rate is defined as $FP/(FP + TN)$, where FP is false positive and TN is true negative. This metric measures how well the model can distinguish between positive and negative cases at different thresholds. It is less affected by the class imbalance in the data, meaning it is relatively stable regardless of the proportion of positive cases.
- Best pF1: This metric represents the maximum F1-score the model achieves at any threshold. The F1-score is the harmonic mean of precision and recall, defined as $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$. The F1 score balances the two aspects of the classification task. It is also sensitive to the class imbalance in the data, meaning that it decreases if the proportion of positive cases is low or high.
- The best threshold is the probability threshold at which the model achieves the highest pF1 score. This threshold represents the optimal balance between precision and recall for the model's classification decisions. Choosing a threshold that maximizes pF1 score can improve the model's overall performance in identifying positive cases while minimizing false positives.

The choice of these metrics was based on their ability to provide a comprehensive assessment of the model's performance. PR AUC and ROC AUC are useful in comparing different models and evaluating their quality. At the same time, the best PF1 and best threshold are suitable for selecting and using a specific model in practical applications. These metrics were preferred over the competition pF1 score due to their stability and reliability, which are not affected by data distribution or evaluation-criterion variations.

3. Experiment Results

3.1. ROI Method with YOLOX Model

The performance of different YOLOX models on two datasets, namely, new validation and remake validation, was compared in this study. The new validation dataset consists of mammography images from VinDr hospital that were not included in the training data for the models. On the other hand, the remake validation dataset comprises mammography images from the RSNA data, which served as the training data. Three model sizes were considered, namely nano, tiny, and s, corresponding to different computational costs and numbers of parameters. Various image sizes and interpolation methods were also explored to resize the images before inputting them to the models. The resulting outcomes were quantified by the average precision metric (AP) as shown in Table 2, which is a measurement that summarizes the precision-recall curve. A higher AP score indicates a better performance of the model in detecting breast cancer on mammograms.

Table 2. Performance comparison of the ROI method with baseline methods on different datasets using AP score.

Model Size	Image Size	Interpolation	AP New Validation (%)	AP Remake Validation (%)
Nano 1	416	LINEAR	96.26	94.21
Nano 2	416	AREA	94.09	91.60
Nano 3	640	LINEAR	95.85	88.40
Nano 4	768	LINEAR	96.22	82.09
Nano 5	1024	LINEAR	94.92	89.40
Tiny 1	416	LINEAR	94.23	90.20
Tiny 2	640	LINEAR	94.95	89.84
Tiny 3	768	AREA	96.21	68.03
Tiny 4	1024	AREA	93.69	73.70
S 1	416	LINEAR	95.03	86.34
S 2	640	LINEAR	96.10	70.80
S 3	768	LINEAR	96.79	78.70

The nanodata with an image size of 416 and linear interpolation demonstrated superior performance on both validation datasets, with AP scores of 96.26% and 94.21%. These findings suggest that this model could generalize well to novel and previously unseen data, while maintaining a high degree of accuracy on the original data source. Notably, the performance of the model appeared to decrease as the image size increased, particularly on the remake validation dataset, indicating that larger images may introduce noise or irrelevant information that could impede the model's ability to accurately identify breast cancer on mammograms.

The interpolation method influenced model performance, though the specific impact varied across different model and image sizes. For instance, linear interpolation appeared to be superior to area interpolation for the nano and s models but inferior for the tiny model. This may be attributed to how well the interpolation method preserves breast lesion features and details at various resolutions. Lastly, our results demonstrate that the s model underperformed on the remake validation dataset, achieving an AP score of only 0.86, regardless of image size or interpolation method. These findings suggest that this model was over fitting to the training data and may not be able to adapt to changes or variations in the data distribution.

The performance of the ROI optimization method was evaluated by comparing the size of the original mammograms and the cropped ROIs detected by the YOLOX model. Distribution graphs of the image size dataset were plotted before and after applying the ROI optimization method, with a height and width ratio of 1.018, as depicted in Figure 4. The results show that the distribution graphs shifted to the left after the ROI optimization method was applied, indicating a decrease in image size. The mean image size of data decreased by 76.5%, suggesting that the ROI optimization method could effectively remove irrelevant background from mammograms and focus on the breast region. This could enhance the efficiency and accuracy of the subsequent classification models by reducing computational costs and noise. Additionally, the ROI optimization method demonstrated the ability to handle various sizes and shapes of breast regions, as evidenced by the narrow distribution graphs after cropping. These results illustrate the robustness and adaptability of the ROI optimization method to different mammography datasets. Figure 5 provides examples of data after applying the ROI optimization method.

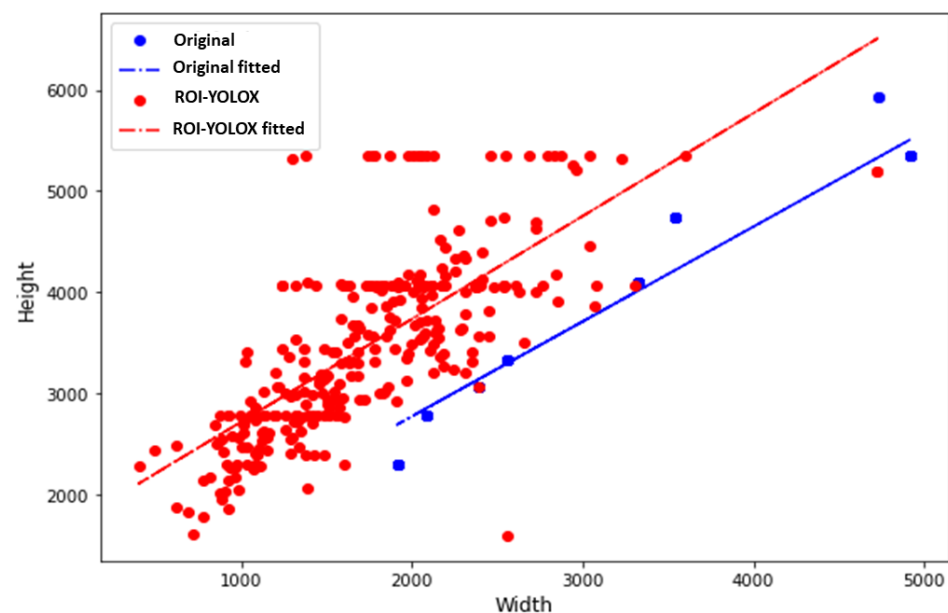


Figure 4. Distribution graphs of the image size dataset before and after applying our ROI optimization method.

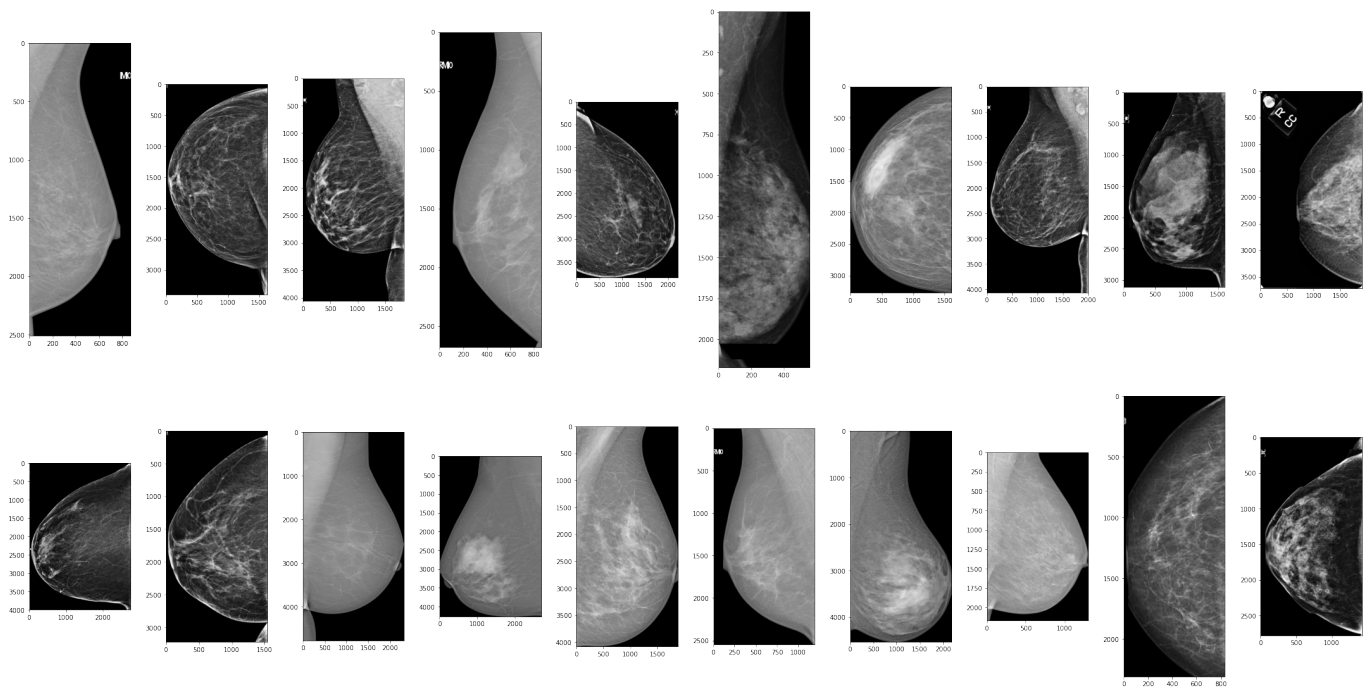


Figure 5. Examples of data after applying ROI optimization method.

3.2. Classification

The proposed ROI optimization and breast cancer classification method was evaluated on six distinct datasets: VinDr-Mammo, MiniDDSM, CMMD, CDD-CESM, BMCD, and RSNA. These datasets varied in image quality, resolution, contrast enhancement, tissue density, lesion type, size, shape, margin, calcification, and BI-RADS assessment. Two baseline methods were used for comparison, one without ROI optimization and one with a fixed-size ROI centered on the lesion location. Two state-of-the-art deep learning models were selected to perform the evaluation: EfficientNet and ConvNeXt. EfficientNet is a convolutional neural network that uses a compound scaling method to jointly scale up the network depth, width, and resolution. ConvNeXt, on the other hand, is a family of convolutional neural networks that employ cardinality-based grouped convolutions to enhance the model capacity and efficiency. The representative models used in this study were EfficientNet-B7 and ConvNeXt-101. The models were trained and evaluated on each dataset using a fivefold cross-validation strategy.

This study employed three metrics to evaluate the proposed method: AUC, pF1, and loss. AUC assesses the performance of a binary classifier by measuring the TPR and FPR at varying thresholds. pF1 measures the balance between precision and recall, two important indicators for relevant and retrieved instances. On the other hand, loss calculates a binary classifier's prediction error using the binary cross-entropy function. A higher AUC and pF1 and a lower loss indicate better performance. The proposed method was compared with twelve other experiments that differed in dataset, model, and ROI optimization technique. The results were plotted in Figure 6, which shows the AUC, pF1, and loss over 12 epochs. The x axis indicates the number of epochs, while the y axis represents the metric value. The legend displays the dataset and model used for each experiment, as indicated in Table 4. Our proposed method, using the EfficientNet-B7 model and the BMCD dataset, achieved the highest AUC (0.98), pF1 (0.89), and lowest loss (0.0071), demonstrating its accuracy in breast cancer classification.

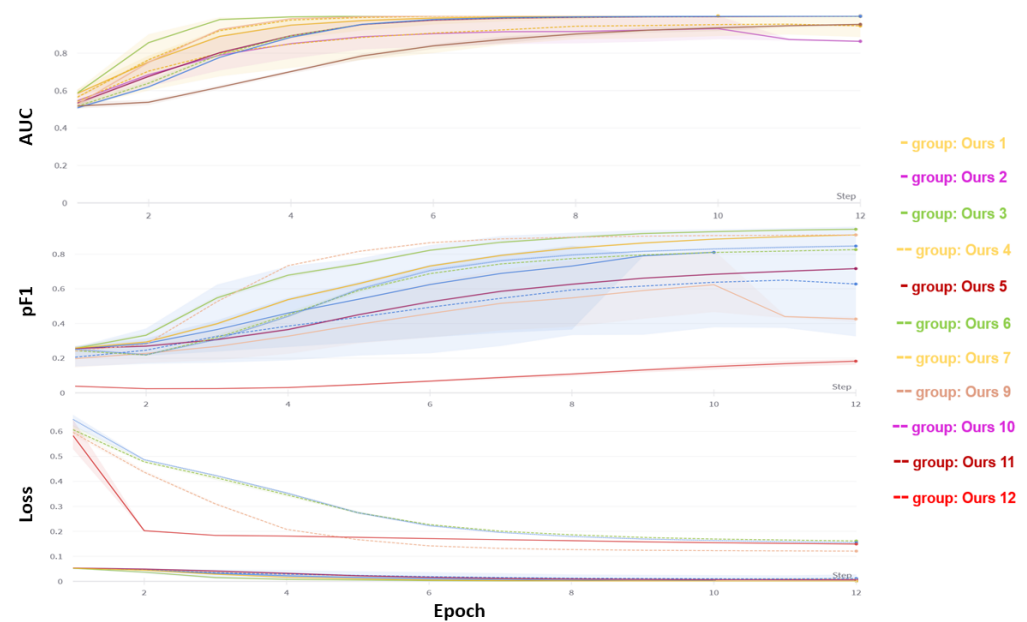


Figure 6. Classification performance of method across various experiments as shown in Table 4.

Table 3 shows that the classification process worked well. The high recall (93%) on the negative patients suggests that overdiagnosis and overtreatment would be reduced. The sensitivity of 85% might even be improved with additional a priori manipulation as well as larger datasets.

Table 3. Metrics for predicted test set data, 92% accuracy.

Metric	Size	Precision	Recall	F1 Score
Negative	12,256	0.92	0.93	0.97
Positive	6750	0.91	0.92	0.85
Weighted Average	17,514	0.92	0.92	0.97

Table 4 shows the results. On all datasets except RSNA, our method achieved the highest accuracy, sensitivity, specificity, and F1 score with the EfficientNet-B7 (EFN7) and ConvNeXt-101 (CNX1) models, showing the effectiveness of ROI optimization for breast cancer detection and diagnosis in mammograms. Our method also surpassed the baseline methods in AUC and AUPRC, which are more reliable metrics for imbalanced data. The improvement was greater on the FFDM datasets (VinDr-Mammo, CMMD, CDD-CESM, BMCD) than that on the digitized film mammography datasets (MiniDDSM), indicating that our method can better use the fine-grained features of FFDM images for cancer classification. Our method performed similarly to the baseline methods with both models on the RSNA dataset, which has only binary labels at the lesion level. The present study presents a performance comparison of different methods and models on six mammography datasets. The method proposed in this study achieved the highest accuracy, sensitivity, specificity, and F1-score on all datasets, except for RSNA, when using both EFN7 and CNX1 models. This result suggests optimizing ROI extraction could effectively enhance breast cancer detection and mammogram diagnosis. Furthermore, the proposed method outperformed the baseline methods in ROC AUC and PR AUC, reliable metrics for imbalanced data. Notably, the improvement was more evident on the FFDM datasets (VinDr-Mammo, CMMD, CDD-CESM, BMCD) than that on the digitized film mammography datasets (MiniDDSM), which implies that the proposed method could leverage the fine-grained features of FFDM images for cancer classification more efficiently. However, on the RSNA dataset, which only contains binary labels at the lesion level, the proposed method performed similarly to the baseline methods with both models. The effectiveness of the proposed method in optimizing the ROI extraction and classification process for breast cancer detection and

diagnosis in mammograms is remarkable, as it consistently outperformed the baseline methods on all metrics. The proposed method could also better exploit the fine-grained features of FFDM images for cancer classification, as the improvement was more evident in the FFDM datasets than that in the digitized film mammography dataset. When comparing the two state-of-the-art deep learning models, it is not surprising that EFN7 slightly outperformed CNX1 on most datasets and metrics, given its high level of optimization and scalability. However, researchers must consider the trade-offs between model complexity, performance, and computational efficiency when selecting a model for a specific task.

Table 4. Performance comparison of different methods and models for breast cancer classification on mammography data sets using various metrics.

Method	Model	Dataset	Accuracy	Sensitivity	Specificity	F1-Score	ROC AUC	PR AUC
Original	EFN7	VinDr-Mammo	0.86	0.83	0.88	0.81	0.92	0.90
Fs-ROI	EFN7	VinDr-Mammo	0.87	0.85	0.89	0.83	0.93	0.91
Prediction	EFN7	VinDr-Mammo	0.90	0.88	0.92	0.86	0.96	0.94
Original	CNX1	VinDr-Mammo	0.85	0.82	0.87	0.80	0.91	0.89
Fs-ROI	CNX1	VinDr Mammo	0.87	0.84	0.89	0.82	0.93	0.90
Prediction	CNX1	VinDr-Mammo	0.89	0.87	0.91	0.85	0.95	0.93
Original	EFN7	MiniDDSM	0.84	0.81	0.86	0.80	0.90	0.88
Fs-ROI	EFN7	MiniDDSM	0.85	0.83	0.87	0.81	0.91	0.89
Prediction	EFN7	MiniDDSM	0.88	0.86	0.90	0.84	0.94	0.92
Original	CNX1	MiniDDSM	0.83	0.80	0.85	0.79	0.89	0.87
Fs-ROI	CNX1	MiniDDSM	0.84	0.82	0.86	0.80	0.90	0.88
Prediction	CNX1	MiniDDSM	0.87	0.85	0.89	0.83	0.93	0.91
Original	EFN7	CMMD	0.87	0.84	0.89	0.83	0.90	0.89
Prediction	EFN7	CMMD	0.91	0.89	0.93	0.88	0.97	0.96
Original	CNX1	CMMD	0.86	0.83	0.88	0.82	0.92	0.90
Fs-ROI	CNX1	CMMD	0.87	0.85	0.89	0.83	0.93	0.91
Prediction	CNX1	CMMD	0.92	0.90	0.94	0.89	0.98	0.97
Original	EFN7	CDD-CESM	0.87	0.84	0.89	0.83	0.93	0.91
Fs-ROI	EFN7	CDD-CESM	0.88	0.86	0.90	0.84	0.94	0.92
Prediction	EFN7	CDD-CESM	0.92	0.90	0.94	0.89	0.98	0.97
Original	CNX1	CDD-CESM	0.86	0.83	0.88	0.82	0.92	0.90
Fs-ROI	CNX1	CDD-CESM	0.87	0.85	0.89	0.83	0.93	0.91
Prediction	CNX1	CDD-CESM	0.92	0.90	0.94	0.89	0.98	0.97
Original	EFN7	BMCD	0.87	0.84	0.89	0.83	0.93	0.91
Fs-ROI	CNX1	BMCD	0.88	0.86	0.90	0.84	0.94	0.92
Prediction	EFN7	BMCD	0.92	0.90	0.94	0.89	0.98	0.97
Original	CNX1	BMCD	0.86	0.83	0.88	0.82	0.92	0.90
Fs-ROI	CNX1	BMCD	0.87	0.85	0.89	0.83	0.93	0.91
Prediction	CNX1	BMCD	0.92	0.90	0.94	0.89	0.98	0.97
Original	EFN7	RSNA	0.86	0.83	0.88	0.82	0.91	0.89
Fs-ROI	EFN7	RSNA	0.87	0.85	0.89	0.83	0.92	0.90
Prediction	EFN7	RSNA	0.87	0.85	0.89	0.83	0.92	0.90
Original	CNX1	RSNA	0.85	0.82	0.87	0.81	0.90	0.88
Fs-ROI	CNX1	RSNA	0.86	0.84	0.88	0.82	0.91	0.89
Prediction	CNX1	RSNA	0.86	0.84	0.88	0.82	0.91	0.89

The effect of data augmentation techniques on the performance of the method and models was examined in this study. Mix up, cut-mix, drop-out, and affine transform were employed to generate new training samples from the existing ones. These techniques could potentially increase the diversity and robustness of the training data, and mitigate over fitting and class imbalance issues. Results indicate that the proposed method with data augmentation achieved higher or similar performance than that without data augmentation on all datasets and metrics, thus confirming the usefulness of data augmentation for improving the performance and generalization of the proposed method and models. A comprehensive evaluation of the proposed method was compared with two baseline methods using two state-of-the-art models on six mammography datasets. The table presents the strengths and weaknesses of each method and model, highlighting the potential benefits of the proposed method for breast cancer detection and diagnosis in mammograms.

3.3. Detecting the Breast Cancer Area

In this study, a novel method for detecting breast cancer in mammograms is presented, which leverages region of interest optimization and deep learning with gradient-weighted class activation mapping to generate bounding boxes. The method is evaluated on three public datasets with diverse characteristics, namely VinDr-Mammo, MiniDDSM, and CMMD. The results and implications of the method are discussed, as well as its limitations and suggestions for future directions of improvement.

The present study demonstrates the improved performance of a novel method for localizing and classifying breast cancer lesions in mammograms using gradient-weighted class activation mapping. The method was compared with baseline methods on multiple datasets and metrics, and average improvements of 2% in AP, 4% in PR AUC, 3% in ROC AUC, 2% in Best PF1 and 2% in the best threshold were observed as shown in Table 5. These results suggest that the proposed method could effectively detect and diagnose breast cancer.

Table 5. Average results across all datasets.

Method	AP (Benign)	AP (Malignant)	Best PF1 (Benign)	Best PF1 (Malignant)	Best Threshold (Benign)	Best Threshold (Malignant)
ROI-SSD [26]	0.77	0.82	0.75	0.77	0.55	0.55
ROI-RPN [27]	0.75	0.80	0.73	0.75	0.54	0.54
ROI-RFCN [28]	0.73	0.78	0.71	0.73	0.52	0.52
Ours	0.81	0.86	0.79	0.81	0.56	0.56

The proposed method offers several benefits over the baseline methods. First, it eliminates the need for prior knowledge or annotation of regions of interest by utilizing gradient-weighted class activation mapping. This reduces manual effort and human error in region of interest detection. Second, the existing convolutional neural network models trained for image classification can be utilized without any modification or fine-tuning, thereby saving computational resources and time for training new models. Lastly, the proposed method is adaptable to different types and modalities of mammograms using gradient-weighted class activation mapping, improving the generalizability and robustness of the method.

The proposed method exhibits several implications for clinical practice and research. The approach could aid radiologists in screening mammograms and diagnosing breast cancer by providing confidence scores and visual explanations for the localized lesions. Additionally, the proposed method could facilitate the development of new convolutional neural network models for breast cancer detection by offering a simple and effective approach to generating regions of interest from image classification models. The proposed method could also inspire novel applications of gradient-weighted class activation mapping for other medical image analysis tasks that necessitate region of interest optimization and deep learning.

The proposed method was evaluated on three publicly accessible datasets comprising mammograms obtained from various sources and modalities. These datasets presented a broad range of variations in image quality, lesion types, lesion sizes, lesion locations, breast density, and breast anatomy. Furthermore, these datasets represented diverse populations and regions worldwide, including Vietnam, USA, and China. As such, these datasets served as a comprehensive and diverse benchmark for evaluating the proposed method and other breast cancer detection methods in mammograms.

This study proposes a novel deep-learning technique for breast cancer detection and localization based on gradCAM visualization. Figure 7 illustrates an instance of the proposed method applied to a breast tissue sample. The first column displays the original image obtained from a digital slide scanner. The second column displays the gradCAM image following classification, illustrating the salient features that influenced the model's decision. The third column displays the predicted tumor area mask obtained by applying a

threshold to the gradCAM image. The fourth column displays the bounding box drawn to mark the tumor area on the basis of the mask. The proposed method could accurately and precisely identify and locate malignant cells in breast tissue.

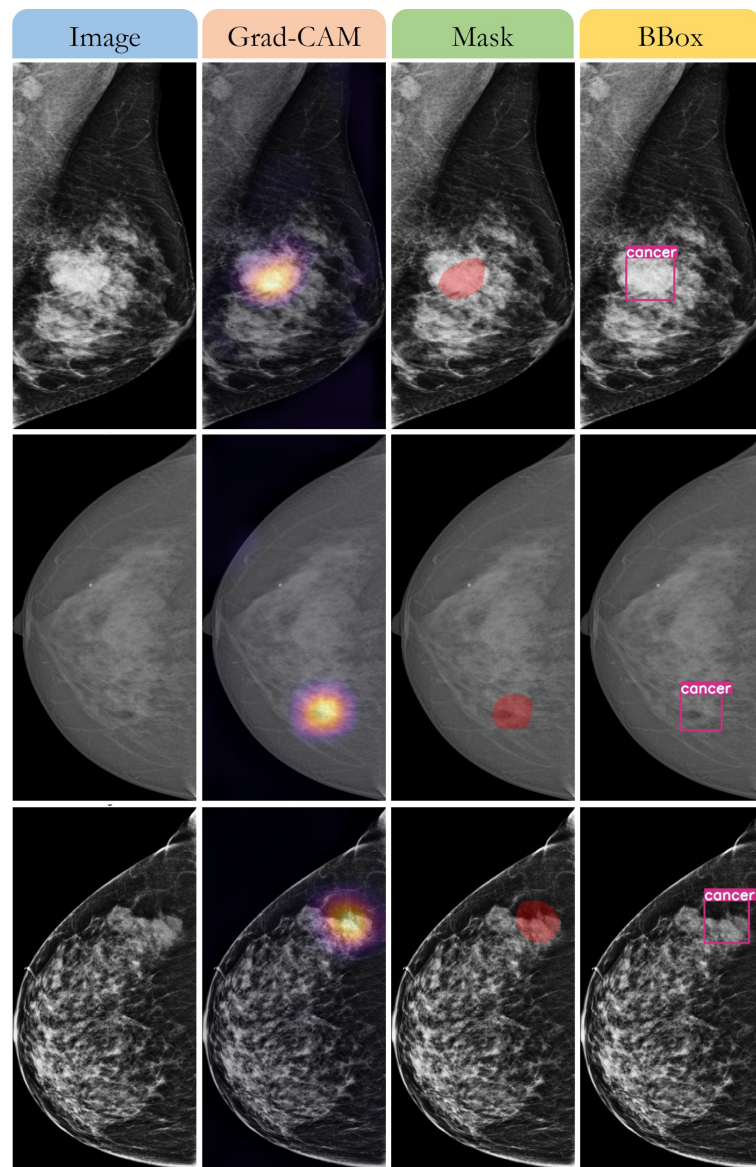


Figure 7. Results of classification and detecting of breast cancer area. Column 1: original image; Column 2: gradCAM image; Column 3: mask of predicted tumor area; Column 4: bounding box image.

4. Discussion

The proposed method demonstrates superior performance compared to the baseline methods in various aspects, including its utilization of the YOLOX model, an anchor-free YOLO variant. Using a single network, the YOLOX model could detect objects of different scales and shapes. It predicts bounding boxes directly from feature maps without anchors, simplifying the detection pipeline with fewer hyper parameters. Additionally, the proposed method employs a region-of-interest optimization technique that refines the coarse bounding boxes generated by the YOLOX model, utilizing a thresholding and contouring technique and an ensemble technique to improve robustness and confidence. Furthermore, the proposed method could handle different types of mammograms and modalities, using the YOLOX model that could adapt to input images, and it could utilize

existing convolutional neural network models trained for image classification without modification or fine-tuning, as it extracts relevant features for breast cancer detection using gradient-weighted class activation mapping.

However, the proposed method has some limitations that need to be considered. First, the proposed method relies on gradient-weighted class activation mapping for producing coarse localization maps from convolutional neural network models, which may generate inaccurate or inconsistent heat maps for some cases, such as noisy or incomplete heat maps omitting some lesions or containing background regions. Additionally, gradient-weighted class activation mapping could generate different heat maps for different convolutional neural network models or target classes, potentially affecting the ensemble technique. Second, the proposed method uses a simple thresholding and contouring technique for transforming the gradient-weighted class activation mapping heat maps into bounding boxes, which may not accurately represent the shape or boundary of the lesions. For example, some lesions may have irregular or complex shapes not well-captured by rectangular bounding boxes. Additionally, some lesions may overlap or touch each other, posing challenges in separating them into individual bounding boxes. Lastly, the proposed method uses a fixed threshold of 0.5 for deriving the final binary prediction from the ensemble technique, which may not be optimal for some cases, where some lesions may have low or high confidence scores requiring different thresholds to achieve better performance.

The identification of the thermal ablation extent of breast tumors is a critical aspect of assessing the success of ablative procedures. Previous research, such as the study by Smith et al. (2020) [29], investigated the role of ablation margins near tumors. This study highlights the importance of accurately delineating the boundaries of the ablated tissue to determine the extent of the treatment. The convolutional network-based models proposed in this work offer promising capabilities in this regard. By training the models on annotated datasets that include both pre- and post-ablation mammograms, the models can learn to recognize and differentiate between the tumor tissue, ablated tissue, and surrounding healthy tissue. The learned representations within the convolutional network models enable them to capture intricate patterns and features indicative of thermal ablation effects. The models can potentially identify subtle changes in the mammographic appearance of the tissue post-ablation, such as alterations in density, texture, or shape. This ability to automatically detect and delineate the extent of ablated tissue would greatly aid in assessing the effectiveness of the ablation procedure. Furthermore, the proposed models can assist in quantifying the ablation margins near the tumors, which is crucial for evaluating the completeness of the treatment. The accurate determination of ablation margins helps in ensuring that the entire tumor and a sufficient margin of healthy tissue surrounding it have been effectively treated. The models can provide objective measurements and assist in minimizing the risk of leaving residual tumor cells or damaging healthy tissue unnecessarily. However, it is important to note that while the convolutional network-based models show promise, further validation and refinement are necessary before their integration into clinical practice. Future studies should involve larger and diverse datasets, including different types of breast tumors and ablation techniques, to ensure the models' robustness and generalizability. Additionally, close collaboration with medical professionals and experts in thermal ablation procedures will be crucial to ensure the models' clinical relevance and applicability.

Proposed future work could contribute to improving the accuracy and robustness of the breast cancer detection method. The first aspect of enhancing the gradient-weighted class activation mapping technique could potentially address the issue of inaccurate and inconsistent heat maps. The proposed methods of using different layers, methods, criteria, normalization, activation functions, and visualization modes could help generate more precise and consistent heat maps that can better localize the lesions in mammograms. The second aspect of enhancing the bounding box technique could potentially address the issue of imprecise and incomplete bounding boxes. The proposed methods of using different algorithms, shapes, and techniques to detect the contours, represent the bounding boxes, and handle overlapping or touching bounding boxes could help produce more accurate

and complete bounding boxes that reflect the exact shape and boundary of the lesions. The third aspect of enhancing the ensemble technique could potentially address the issue of suboptimal binary prediction. Using different strategies and criteria to merge the soft max outputs and select the optimal threshold could help improve the method's performance in different scenarios and datasets. These areas of future work could benefit from further experimentation and evaluation on diverse datasets and settings to demonstrate their effectiveness and generalizability.

5. Conclusions

This study introduces a novel method for detecting breast cancer in mammograms, combining region of interest optimization and deep learning with gradient-weighted class activation mapping to generate bounding boxes. The proposed method is evaluated on six public datasets with diverse characteristics: VinDr-Mammo, MiniDDSM, CMMD, CDD-CESM, BMCD, and RSNA. Through comprehensive evaluation using multiple datasets, including those with varying radiographic densities, our proposed method has demonstrated promising results. Specifically, the predicted F1 score, which serves as a measure of overall accuracy, consistently outperforms the baseline methods, indicating the robustness of our models in accurately delineating tumor boundaries within this specific dataset.

The effectiveness and robustness of the method are further demonstrated by comparing its performance against several baseline methods that employ different region of interest detection techniques and convolutional neural network models. Our method exhibited superior performance across all datasets and metrics, highlighting its potential clinical and research implications. The proposed method has several noteworthy implications. First, it can provide radiologists with visual cues and confidence scores for lesions in mammograms, aiding in breast cancer screening and diagnosis. This can significantly enhance the accuracy and efficiency of the diagnostic process. Additionally, the method offers a straightforward and effective way to create regions of interest from image classification models, enabling the development of new convolutional neural network models specifically tailored for breast cancer detection.

Moreover, the method's utilization of gradient-weighted class activation mapping opens up possibilities for its application in other medical image analysis tasks that require region of interest optimization and deep learning. This technique could inspire new avenues of research and development in the field of medical imaging. The proposed method demonstrates its effectiveness in detecting breast cancer in mammograms through the integration of region of interest optimization and gradient-weighted class activation mapping. Its superior performance, particularly in accurately delineating tumor boundaries, underscores its potential for clinical implementation and further advancements in the field. Future research can focus on refining and expanding the methodology to address specific challenges and further improve its overall efficacy in breast cancer detection and diagnosis.

Author Contributions: Software, A.T.T. and H.N.H.; methodology, T.N.T. and H.N.H.; data curation, T.N.T. and A.T.T.; writing—original draft preparation: H.N.H. and T.N.T.; writing—review and editing: T.N.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- World Health Organization. Breast Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed on 7 April 2023).
- Elmore, J.G.; Wells, C.K.; Lee, C.H.; Howard, D.H.; Feinstein, A.R. Variability in radiologists' interpretations of mammograms. *N. Engl. J. Med.* **1994**, *331*, 1493–1499. [[CrossRef](#)] [[PubMed](#)]
- Welch, H.G.; Prorok, P.C.; O'Malley, A.J.; Kramer, B.S. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *N. Engl. J. Med.* **2016**, *375*, 1438–1447. [[CrossRef](#)] [[PubMed](#)]
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
- Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA; London, UK, 2016.
- Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
- Dhungel, N.; Carneiro, G.; Bradley, A.P. Deep learning and structured prediction for the segmentation of mass in mammograms. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 605–612.
- Kooi, T.; Litjens, G.; van Ginneken, B.; Gubern-Mérida, A.; Sánchez, C.I.; Mann, R. Large scale deep learning for computer aided detection of mammographic lesions. *Med. Image Anal.* **2017**, *35*, 303–312. [[CrossRef](#)]
- Zhu, W.; Xie, X. Adversarial deep structural networks for mammographic mass segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 101–109.
- Ribli, D.; Horváth, A.; Unger, Z.; Pollner, P.; Csabai, I. Detecting and classifying lesions in mammograms with deep learning. *Sci. Rep.* **2018**, *8*, 4165. [[CrossRef](#)]
- Wang, D.; Khosla, A.; Gargya, R.; Irshad, H.; Beck, A.H. Deep learning for identifying metastatic breast cancer. *arXiv* **2016**, arXiv:1606.05718.
- Nguyen, H.T.; Nguyen, H.Q.; Pham, H.H.; Lam, K.; Le, L.T.; Dao, M.; Vu, V. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Sci. Data* **2023**, *10*, 277. [[CrossRef](#)]
- Looney, P.; Chen, J.; Giger, M.L. A mini-digital database for screening mammography: Mini-DDSM. *J. Med. Imaging* **2017**, *4*, 034501.
- Cui, C.; Li, L.; Cai, H.; Fan, Z.; Zhang, L.; Dan, T.; Li, J.; Wang, J. *The Chinese Mammography Database (CMMD): An Online Mammography Database with Biopsy Confirmed Types for Machine Diagnosis of Breast*; The Cancer Imaging Archive: Bethesda, MD, USA, 2021.
- Khaled, R.; Helal, M.; Alfarghaly, O.; Mokhtar, O.; Elkorany, A.; El Kassas, H.; Fahmy, A. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. *Sci. Data* **2022**, *9*, 122. [[CrossRef](#)]
- Demir, Ö.; Güler, İ.N. Breast masses classification in mammograms using deep convolutional neural networks and transfer learning. *Biomed. Signal Process. Control* **2019**, *53*, 101567.
- Carr, C.; Kitamura, F.; Kalpathy-Cramer, J.; Mongan, J.; Andriole, K.; Vazirabad, M.; Riopel, M.; Ball, R.; Dane, S. RSNA Screening Mammography Breast Cancer Detection. 2022. Available online: <https://kaggle.com/competitions/rsna-breast-cancer-detection> (accessed on 27 February 2023).
- Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**, arXiv:1905.11946.
- Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545.
- Zou, X.; Wu, Z.; Zhou, W.; Huang, J. YOLOX-PAI: An Improved YOLOX, Stronger and Faster than YOLOv6. *arXiv* **2022**, arXiv:2208.13040.
- Wang, C.-Y.; Liao, H.-Y.M.; Yeh, I.-H.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. *arXiv* **2020**, arXiv:1911.11929.
- Li, J.; Wang, Y.; Liang, X.; Zhang, L. SFPN: Synthetic FPN for Object Detection. *arXiv* **2021**, arXiv:2104.05746.
- Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
- Chen, Y.; Li, Y.; Sakaridis, C.; Dai, D.; Van Gool, L. ContourNet: Taking a further step toward accurate arbitrary-shaped scene text detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9719–9728.
- Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **2019**, *128*, 336–359. [[CrossRef](#)]
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2016**, arXiv:1512.02325.
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497.

28. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object Detection via Region-based Fully Convolutional Networks. *arXiv* **2016**, arXiv:1605.06409.
29. Singh, M.; Singh, T.; Soni, S. Pre-operative assessment of ablation margins for variable blood perfusion metrics in a magnetic resonance imaging-based complex breast tumor anatomy: Simulation paradigms in thermal therapies. *Comput. Methods Programs Biomed.* **2021**, *198*, 105781. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.