Metadata of the chapter that will be visualized in SpringerLink

Book Title	Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications	
Series Title		
Chapter Title	Deep Learning Techniques for Segmenting Breast Lesion Regions and Classifying Mammography Images	
Copyright Year	2023	
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd.	
Author	Family Name	**Nguyen**
	Particle	
	Given Name	**Nam V.**
	Prefix	
	Suffix	
	Role	
	Division	Faculty of Information Technology
	Organization	Industrial University of Ho Chi Minh City
	Address	Ho Chi Minh City, Vietnam
	Email	
Author	Family Name	**Huynh**
	Particle	
	Given Name	**Hieu Trung**
	Prefix	
	Suffix	
	Role	
	Division	Faculty of Information Technology
	Organization	Industrial University of Ho Chi Minh City
	Address	Ho Chi Minh City, Vietnam
	Email	
Corresponding Author	Family Name	**Le**
	Particle	
	Given Name	**Phuc-Lu**
	Prefix	
	Suffix	
	Role	
	Division	Faculty of Information Technology
	Organization	Ho Chi Minh City University of Science
	Address	Ho Chi Minh City, Vietnam
	Email	lplu@fit.hcmus.edu.vn

Abstract

Breast cancer is currently one of the leading causes of death in many countries worldwide. Detecting breast masses early can provide higher chances of survival for patients. However, determining and segmenting benign or malignant breast masses is becoming a challenging issue. Currently, there are a wide range of Convolutional Neural Networks used to address breast mass segmentation and breast cancer classification issues, such as U-Net, SegNet, Mask R-CNN, for segmentation, and Convnet, CNN, ResNet, for

classification. However, these solutions are still not effective enough. Therefore, we have solved this problem by applying modern model called Segment Anything Model to predict breast tumor segmentation masks to help doctors identify and evaluate breast tumors and two models EfficientNet B0 combined with Focal Loss and Vision Transformer base to classify breast images as benign or malignant. The experimental results show those modern models achieved high performance with an Intersection over Union score of 96.59% on the CIBS-DDSM dataset. Additionally, the classification model achieved an accuracy of 100% and F1-scores of 100% on the DDSM dataset, outperforming other models. Our technique helps support doctors in identifying breast masses in images and provides reliable predictions for diagnostic purposes, thus improving the effectiveness of breast cancer detection.

Keywords (separated by '-')	Deep learning - Breast Cancer - EfficientNet B0 - Segment Anything Model - Vision Transformer - Focal Loss

Deep Learning Techniques for Segmenting Breast Lesion Regions and Classifying Mammography Images

Nam V. Nguyen[1], Hieu Trung Huynh[1], and Phuc-Lu Le[2(✉)]

[1] Faculty of Information Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam
[2] Faculty of Information Technology, Ho Chi Minh City University of Science, Ho Chi Minh City, Vietnam
lplu@fit.hcmus.edu.vn

Abstract. Breast cancer is currently one of the leading causes of death in many countries worldwide. Detecting breast masses early can provide higher chances of survival for patients. However, determining and segmenting benign or malignant breast masses is becoming a challenging issue. Currently, there are a wide range of Convolutional Neural Networks used to address breast mass segmentation and breast cancer classification issues, such as U-Net, SegNet, Mask R-CNN, for segmentation, and Convnet, CNN, ResNet, for classification. However, these solutions are still not effective enough. Therefore, we have solved this problem by applying modern model called Segment Anything Model to predict breast tumor segmentation masks to help doctors identify and evaluate breast tumors and two models EfficientNet B0 combined with Focal Loss and Vision Transformer base to classify breast images as benign or malignant. The experimental results show those modern models achieved high performance with an Intersection over Union score of 96.59% on the CIBS-DDSM dataset. Additionally, the classification model achieved an accuracy of 100% and F1-scores of 100% on the DDSM dataset, outperforming other models. Our technique helps support doctors in identifying breast masses in images and provides reliable predictions for diagnostic purposes, thus improving the effectiveness of breast cancer detection.

Keywords: Deep learning · Breast Cancer · EfficientNet B0 · Segment Anything Model · Vision Transformer · Focal Loss

1 Introduction

According to the WHO, in 2020, approximately 2.3 million women worldwide were diagnosed with breast cancer, and there were 685,000 deaths from breast cancer globally. The exact causes of breast cancer are still not fully understood, although genes and hormones appear to play a major role. Detecting and preventing the development of these cancer cells as early as possible is beneficial not only in increasing the chances of cure but also in improving the quality of life for

T. K. Dang et al. (Eds.): FDSE 2023, CCIS 1925, pp. 1–13, 2023.
https://doi.org/10.1007/978-981-99-8296-7_34

patients. Currently, imaging techniques such as Magnetic Resonance Imaging (MRI), Single-Photon Emission Computed Tomography (SPECT), Computed Tomography (CT), and X-ray mammography are used for screening and early detection of breast-related abnormalities and breast cancer. The earlier the diseases are detected, the higher the chances of successful treatment. Besides the benefits it brings, X-ray mammography also has certain limitations, including the inability to determine the benign or malignant state of a breast lesion. Moreover, manual image interpretation can lead to subjective results, errors, and burden the healthcare facility. Therefore, recently, image processing techniques combined with Convolutional Neural Networks (CNNs) have been introduced to assist doctors in breast cancer diagnosis. These techniques aim to address the limitations of traditional mammography by providing more objective and automated analysis, aiding in the early detection and accurate diagnosis of breast cancer. However, CNNs require considerable pre-processing to compensate for poor image quality [1]. The usage of low-quality and noisy mammography images can adversely affect the model's performance. Additionally, medical image data, including mammography images, is often scarce, leading to data imbalance between benign and malignant classes, which can bias the model's predictions towards the class with more data. As a result, selecting a suitable model for breast lesion segmentation and mammogram classification becomes one of the major challenges.

In this paper, we utilize the most advanced Segment Anything Model (SAM) model for breast tumor segmentation. This model is built on the largest segmentation dataset up to now, with over 1 billion masks on 11 million licensed and privacy-respecting images [2], making it exceptionally powerful. We performed model fine-tuning on the CIBS-DDSM dataset, starting from pre-trained weights and biases, which improved the breast tumor segmentation performance significantly. In addition, we developed a deep learning method for mammogram classification. We applied data augmentation techniques and utilized YOLOX-s models to remove redundant image regions. Furthermore, we employed mode; EfficientNet B0 combine with Focal Loss and Vision Transformer (ViT) base to classify breast images as benign or malignant tumors. Our solution enables comprehensive and detailed tumor segmentation, providing doctors with valuable insights into breast lesions. Moreover, our classification approach offers reliable predictions regarding the nature of the breast tumor, assisting doctors with trustworthy diagnostic recommendations for the patients.

2 Preliminaries

2.1 Dataset

In this study, we used two datasets for two different purposes. Firstly, we used the CBIS-DDSM dataset (Curated Breast Imaging Subset of DDSM) available at Cancerimagingarchive, and selected only the mass cases, to train the SAM segmentation model with 1,696 images from 892 patients. Secondly, we used the Digital Database for Screening Mammography (DDSM) dataset to train the classification models, which consists of 13,128 images (including the processed, rotated images

from the original), here we only used 2,188 raw and unprocessed images. This is available at: https://data.mendeley.com/datasets/ywsbh3ndr8/2.

2.2 Data Preprocessing

To train a good model, preprocessing the images is essential as it significantly impacts the model's performance. We used the CIBS-DDSM dataset in its original Digital Imaging and Communications in Medicine (DICOM) format, which contains important metadata such as brightness, contrast, image features, etc.

Converting DICOM images to PNG images: due to the large file size of DICOM images, this demands high-end hardware while our training on Kaggle with limited hardware resources, we converted the images to PNG format for convenience in processing and inputting them into the models. PNG maintains good image quality and compression capabilities without loss of data compared to other formats like JPEG, TIFF, etc. Additionally, we applied windowing, also known as gray-level mapping, to select specific pixel ranges from the image prior to normalization. This technique effectively increased the contrast between soft tissues and special tissue regions, and also allowed for a larger range when manually adjusting brightness/contrast later on. This technique should be applied when exporting images in PNG format (Fig. 1). AQ3

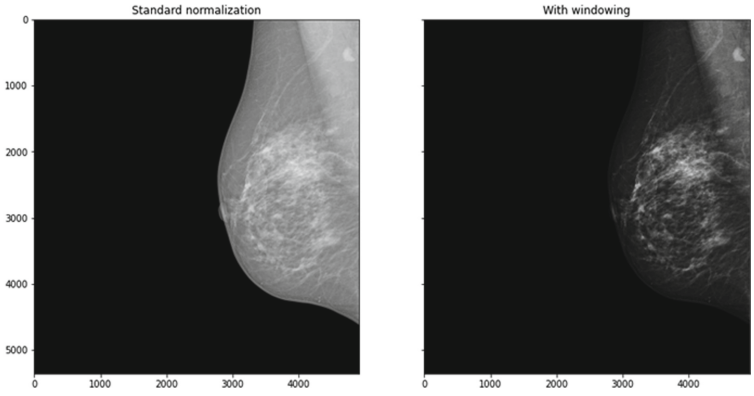

Fig. 1. Comparison of regular image and image using windowing.

Cropping images using the YOLOX model: with the original large-sized mammograms, we only need to extract the region of interest to feed into the model. Cropping and selecting the region of interest help reduce the size and file size of the images and reduce noise.

Recently, with the emergence of YOLOX for accurate lesion detection, we have utilized it for image cropping and resizing with a height of 1024 and a width of 512. YOLOX is a convolutional neural network model designed for object detection, recognition, and classification. Equipped with some recent advanced detection techniques, such as decoupled head, anchor-free, and advanced label assigning strategy, YOLOX achieves a better trade-off between speed and accuracy compared to other models of all sizes [3] (Fig. 2).

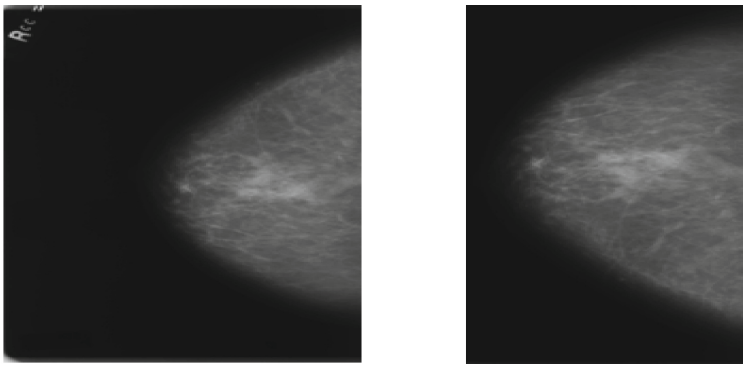

Fig. 2. Images before (left) and after (right) cropped by YOLOX-s.

In this study, we used the YOLOX-s version with 9 million parameters. Additionally, we annotated 1,000 mammogram images to train the YOLOX-s model, which helps improve accuracy in detecting regions of interest.

2.3 Transfer Learning

Transfer learning aims at improving the performance of target learners on target domains by transferring the knowledge contained in different but related source domains. In this way, the dependence on a large number of target domain data can be reduced for constructing target learners [4]. Therefore, this technique is particularly suitable for cases with limited data, especially for the set of mammography images, as models can be pre-trained on large datasets from other related medical image fields and then transfer this knowledge to the task of classifying and segmenting mammography images.

To achieve the best results when applying transfer learning, we should use pre-trained models with a large dataset that is relevant to the new target. This is because these models will have already learned useful features from the large dataset, and they can be used as a foundation for the transfer learning model.

After using the pre-trained model, we can fine-tune the model's parameters to fit the new dataset. This fine-tuning process can help improve the performance of the transfer learning model on the new dataset.

- For segmentation, we utilized the pre-trained SAM model, which was trained on the largest segmentation dataset to date, consisting of over 1 billion masks on 11 million images [2]. This significantly improves the efficiency of breast mass segmentation.
- For classification, we employed two pre-trained models, Vision Transformers and EfficientNet, which were trained on large natural image datasets like ImageNet. The objective was to use the pre-trained weights as a starting point for training the classification model to classify breast mammograms as benign or malignant.

3 Method Details

3.1 Focal Loss

Focal Loss is highly effective in addressing the issue of data imbalance among different classes. It focuses on harder-to-predict examples more than the easier ones. This aids in enhancing the prediction of challenging and hard-to-classify examples. Focal loss incorporates a modulating factor $(1 - p_t)^\gamma$ into the cross-entropy loss, with a tunable focusing parameter $\gamma \geq 0$ [5].

The formula for focal loss is defined as follows

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

If in the case of misclassified samples, p_t becomes small, approximately or very close to 1, and the loss function remains largely unaffected. As p_t approaches 1, the modulating factor tends towards 0, thereby down-weighting the loss value for well-classified examples. The focusing parameter γ smoothly adjusts the rate at which easy examples are down-weighted [5]. Subsequently, we add an α-balanced variant of the focal loss, which contributes to slightly improved accuracy compared to the unbalanced form.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{2}$$

In this study, we use the parameter $\gamma = 2$ and an α-balanced variant of the focal loss with a value of 0.25.

3.2 Balancing Data

Due to the limited and insufficient diversity of the data, we performed data augmentation using techniques such as Horizontal Flip, Vertical Flip, Rotation, and Color Jitter. Additionally, we divided the data into batches with a batch size of 64 or 32, depending on the model. Ensuring data balance within each batch is also an important concern. We increased the frequency of occurrence of classes with fewer samples in each batch. To achieve this, we assigned weights to each image using WeightedRandomSampler in PyTorch.

Initially, we calculated these weights using the formula one divided by the number of occurrences of each class. This way, classes with a higher number of occurrences were assigned smaller weights, while classes with fewer occurrences were assigned larger weights. This approach helped to balance the number of samples from different classes in each batch.

3.3 Vision Transformer Architecture

In this study, we utilized a pre-trained vision transformer model on the ImageNet dataset for benign and malignant classification. Additionally, we employed ensemble learning, which is the combination of multiple models, to further enhance the performance of our model.

Originating from the transformer model used in natural language processing (NLP) with remarkable effectiveness and success on one-dimensional word tokens, we applied its superior performance to images. To achieve this, we transformed the input images of size (W, H, C) into two-dimensional patches and flattened them into one-dimensional vectors. This sequence was then considered as the input sequence for the transformer encoder. The number of patches, N, was calculated using the formula $N = \frac{HW}{P^2}$, where (P, P) represents the resolution of each image patch. The transformer maintains a constant latent vector size D across all layers, and thus, we flattened the patches and mapped them to D dimensions using a trainable linear projection. The output of this projection is referred to as the patch embeddings [6] (Fig. 3).

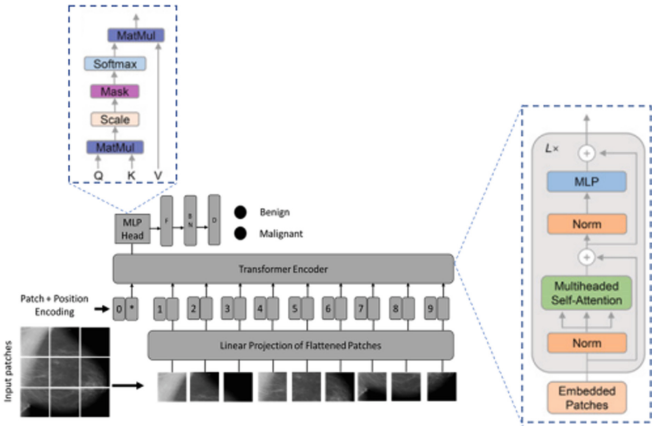

Fig. 3. The vision transformer-based transfer learning architecture for mammogram breast image detection [7]

The processing pipeline of the model can be summarized as follows: each image patch is flattened into a vector X_p^n of length $P \times P \times C$, where P is the patch size and C is the number of color channels, and there are a total of n image patches. The vectors X_p^n are then mapped to a D-dimensional space using a learnable linear projection denoted as E. As a result, we obtain a sequence of embedding vectors with size D. This sequence of embeddings is prefixed with a learnable class embedding called X_{class}, and the values of X_{class} correspond to the classification outcomes Y. Finally, the embedding vectors are combined with the positional embeddings E_{pos} (learned during training) to add positional information to the input [7]. The concatenated embedding vectors form the final input z_0, which will be fed into the encoder network of the vision transformer to perform the image classification process. This processing pipeline enables the

model to understand the spatial structure of the image and classify it based on the embedded information from the image patches and positional information.

$$z_0 = [X_{class};\ X_p^1 E;\ ...;\ X_p^N E] + E_{pos} \tag{3}$$

Finally, we feed z_0 into a transformer encoder network, which consists of a stack of L identical layers, to perform the classification process. The output of the L_{th} layer of the encoder is then fed into the classification head. Additionally, we use an MLP (Multi-layer Perceptron) with a single hidden layer to further process the classification by employing a single linear layer for the actual classification task. The GELU activation function is used in the MLP as the classification head. In summary, the transformer network helps abstract features from each image, and then the MLP is used for image classification based on the pre-abstracted features.

3.4 SAM Architecture

In this study, we employ SAM for breast tumor segmentation. With training on an extensive dataset, it has enabled the model to perform breast tumor segmentation effectively. SAM is released by the Meta AI Research group. The SAM model architecture consists of three main components: an image encoder, a flexible prompt encoder, and a fast mask decoder.

Image Encoder: We utilize a masked auto-encoder (MAE) [8] pre-trained ViT, which is minimally adapted to handle high-resolution inputs. The image encoder operates once per image and can be employed prior to triggering the model [2].

Prompt Encoder: We consider two groups of prompts: one group consists of sparse prompts (including points, boxes, and text), and the other group comprises dense prompts (including masks). To represent points and boxes, we utilize positional encodings combined with learned embeddings specific to each prompt type. For free-form text, we employ a pre-trained text encoder from the Contrastive Language-Image Pretraining (CLIP) framework. Dense prompts, namely masks, are embedded using convolutions and then element-wise summed with the image embedding.

Mask Decoder: The mask decoder efficiently maps the image embedding, prompt embeddings, and an output token to a mask. Our modified decoder block uses prompt self-attention and cross-attention in two directions (prompt-to-image embedding and vice-versa) to update all embeddings. After running two blocks, we up sample the image embedding and an MLP maps the output token to a dynamic linear classifier, which then computes the mask foreground probability at each image location [2] (Fig. 4).

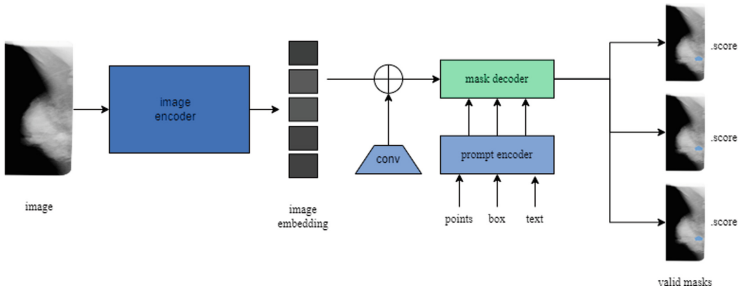

Fig. 4. Overview of SAM Application in Breast Tumor Segmentation.

3.5 Experimental Settings

In this study, our objective is to propose a solution for breast cancer segmentation and classification to aid physicians in diagnosing patients' conditions as benign or malignant. During the experimentation process, we divided our study into two stages. The first stage involved experiments on the SAM segmentation model. In the second stage, experiments were conducted on models EfficientNet B0 combined with Focal Loss and ViT applied for breast tumor classification.

3.6 Implementation Detail

During the model training process, we set the learning rate to 10^{-6} and utilize the AdamW optimizer with a weight decay setting of 0.05, which is an L2 regularization technique aimed at mitigating overfitting by applying a small penalty to the model's weight 'w' during the update process. We employ the Binary Cross-Entropy (BCE) loss function. In the context of the vision transformer model, GELU is employed as an activation function, while for other CNN-based models, we use the Rectified Linear Unit (ReLU). For the SAM model, the Intersection-over-Union (IoU) loss is utilized as the loss function. All experiments are conducted on the Kaggle platform using free GPUs.

4 Results

4.1 Breast Tumor Segmentation Using the SAM Model

In this phase, we utilize the CIBS-DDSM dataset to train the model. We select 1,318 images for the training set and 378 images for the testing set. Additionally, we conduct a pre-training process for the SAM model (Fig. 5).

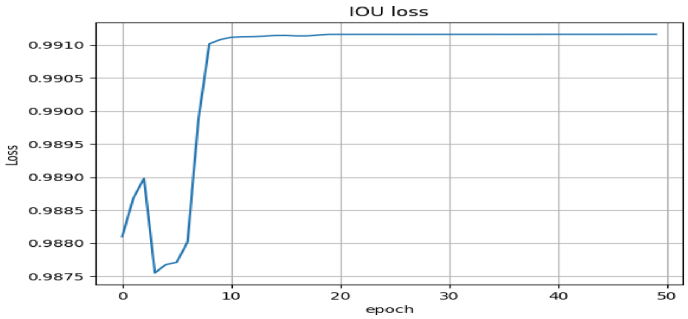

Fig. 5. The loss function during the training process.

The results of the training process with an IoU score exceeding 99% demonstrated that the SAM model adapted very well to the new training dataset. However, we also took the best weights achieved during training to perform evaluation on the test dataset using assessment metrics IoU. Denote S as the IoU values in data set, then we calculated some statistical quantities as follow: Mean is the average of S, Max $= \arg\max_{IoU_i \in S} IoU_i$ and Min $= \arg\min_{IoU_i \in S} IoU_i$ (Table 1).

Table 1. The results on the test sets

Type	IoU
Mean	96.59%
Max	99.4%
Min	53%

There are still some instances where the segmentation model does not perform well, achieving only 53% with IoU. However, overall, the SAM model is demonstrating excellent segmentation performance. It is indeed a valuable tool aiding physicians in detecting breast tumors and identifying any discrepancies during the diagnostic process. Comparison with previous studies (Table 2):

Table 2. The related literature and the method using our SAM model.

Literature	Dataset	IoU
Wessam at al. [9]	CBIS-DDSM	92.99%
Asma at al. [10]	CBIS-DDSM	80.02%
Our results	CBIS-DDSM	96.59%

Finally, compare the results with the state-of-the-art methods and models of previous studies. Our proposed SAM pre-trained model outperforms them (Fig. 6).

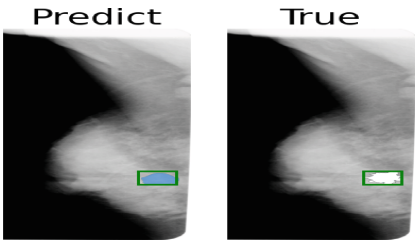

Fig. 6. Segmentation image of breast tumor as predicted by the model and segmentation image of actual breast tumor.

4.2 Classification of Breast Tumors

The results of the model training process for breast tumor classification have been presented in Table 3. Here, we employ two main approaches: (1) the 1st approach involves utilizing the ViT model for training, (2) the 2nd approach incorporates the EfficientNet B0 in conjunction with the Focal Loss function, which helps the model focus on incorrectly predicted samples rather than samples that the model predicts confidently. Both of these approaches yield promising results with Accuracy, Precision, and F1 score metrics all reaching 99.99% on the DDSM dataset.

Table 3. Classification results

Model	Test dataset	Accuracy (%)	Precision (%)	F1 Score (%)	AUC (%)
Efficientnet B0	DDSM	97.2	97.2	97.2	97.2
Efficientnet B0 + Haze Removal + Clahe	DDSM	98.6	98.6	98.6	98.6
Efficientnet B0 + Focal Loss	DDSM	99.9	99.9	99.9	99.9
Vision Transformer (ViT) base	DDSM	100	100	100	100

The following table shows the training time (by second) and weight size (by GB) between various models (Table 4).

Table 4. Model Training Time

Model	Train dataset	Time (s)	Weight size (GB)
Efficientnet B0	DDSM	1638	0.045
Efficientnet B0 + Haze Removal + Clahe	DDSM	9909	0.045
Efficientnet B0 + Focal Loss	DDSM	399	0.045
Vision Transformer (ViT) base	DDSM	604	0.97

We provide two confusion matrices, which are very similar, for two models: EfficientNet B0 + Focal Loss base and (ViT) Base (Fig. 7).

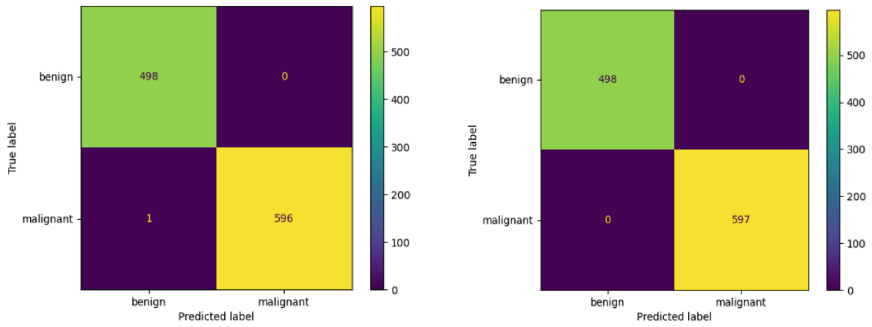

Fig. 7. The confusion matrices of EfficientNet B0 + Focal Loss (left) and ViT Base (right).

To achieve the results, we experimented various methods to select the best approaches. In Table 3, we present the best-performing experiments. Initially, we trained the EfficientNet B0 model on the DDSM dataset, achieving an accuracy of 97.2% and an F1 Score of 97.2%. Subsequently, during the training process, we recognized the pivotal role of data in enhancing model performance (Fig. 8).

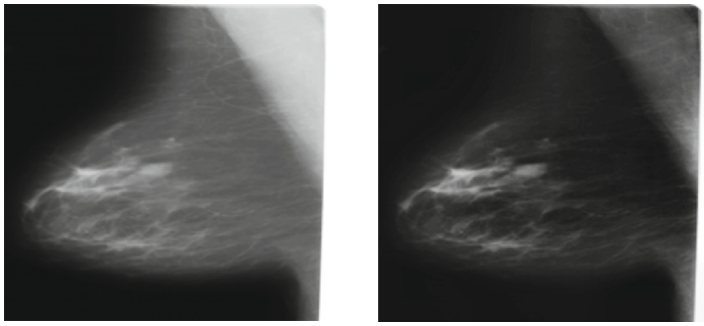

Fig. 8. Images before and after applying Haze Removal and CLAHE techniques

However, accessible data regarding mammographic images is limited. Consequently, we performed data augmentation by employing image processing techniques, namely Haze Removal to enhance clarity by eliminating noise and blurriness, thus sharpening the breast masses. This was combined with the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to improve contrast and image clarity. As a result, we improved the model by 1.4% in accuracy, achieving 98.6% accuracy and an F1 Score of 98.6%. However, the training time for this model increased sixfold compared to EfficientNet B0.

Subsequently, our proposed method combined EfficientNet B0 with the Focal Loss technique, enabling the model to focus on challenging prediction cases. This

led to a significant improvement of 1.3% in performance, resulting in an accuracy of 99.9% and an F1 Score of 99.9%, compared to the (EfficientNet B0 + Haze Removal + CLAHE) method.

Finally, we conducted experiments on a new model, ViT base, achieving nearly 100% accuracy and an F1 Score of 100%. In summary, for the task of breast image classification, we propose the EfficientNet B0 + Focal Loss model and the ViT base model due to their low training time and high performance. All our experiments were conducted on the DDSM dataset. The following is the comparison between our work and the recent studies (Table 5):

Table 5. Related literature and our Proposed Method 1 & 2

Literature	Test dataset	Accuracy (%)	Precision (%)	F1 Score (%)	AUC (%)
Mei-Ling Huang et al. [11]	DDSM	99.93	–	99.92	–
Ribli et al. [12]	DDSM	–	–	–	95
Chougrad et al. [13]	DDSM	97.35	–	–	98
Proposed Method 1	DDSM	99.9	99.9	99.9	99.9
Proposed Method 2	DDSM	100	100	100	100

With our two proposed methods, there are significant improvements in performance metrics such as Accuracy, Precision, F1 Score, and AUC compared to other studies. Our works including source code can be found on the following link: https://github.com/david-nguyen-S16/Segmentation-and-classification-of-mammographic-images.

5 Conclusions and Future Works

In this study, we performed breast mass segmentation on mammographic images using a pre-trained SAM that was fine-tuned with superior performance, achieving an IoU score of 96.59% on the CIBS-DDSM dataset. This supports doctors in detecting abnormal breast masses, with a significantly improved segmentation performance, given a partially pre-trained dataset. Secondly, we conducted the classification of mammographic images as benign or malignant using two proposed methods: EfficientNet B0 + Focal Loss and ViT base. Both models achieved outstanding performance with F1 Scores of 99.9% and 100% on the DDSM dataset, leveraging transfer learning to enhance training. A limitation of this study is the scarcity of mammographic image data, but its impact on model performance is substantial.

In the future, we expect to refine the new methods for mammographic segmentation and classification, while also gathering additional mammographic data to achieve better results across different datasets.

Acknowledgement. We would like to express our gratitude to Mr. Khoi Nguyen (Hajim School of Engineering & Applied Sciences: University of Rochester, USA), have guided and supported us in experimenting and completing this paper.

This article was funded in part by University of Science, VNU-HCM under Grant No. CNTT2022−11.

References

1. Ayana, G., Dese, K., Raj, H., Krishnamoorthy, J., Kwa, T.: De-speckling breast cancer ultrasound images using a rotationally invariant block matching based non-local means (RIBM-NLM) method. Diagnostics **12**(4), 862 (2022). https://doi.org/10.3390/diagnostics12040862
2. Kirillov, A., et al.: Segment anything. ArXiv abs/2304.02643v1. Meta AI Research, FAIR (2023)
3. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: exceeding YOLO series in 2021. ArXiv abs/2107.08430. Megvii Technology (2021)
4. Zhuang, F., et al.: A comprehensive survey on transfer learning. ArXiv abs/1911.02685. IEEE (2020)
5. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. ArXiv abs/1708.02002 (2017)
6. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. ArXiv abs/2010.11929v2. Google Research, Brain Team (2020)
7. Ayana, G., et al.: Vision transformer-based transfer learning for mammogram classification. Diagnostics **13**, 178 (2023). https://doi.org/10.3390/diagnostics13020178
8. He, K., Chen, X., Xie, S., Li, Y., Dollar, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: CVPR (2022)
9. Salama, W.M., Aly, M.H.: Deep learning in mammography images segmentation and classification: automated CNN approach. Alex. Eng. J. **60**(5), 4701–4709 (2021). https://doi.org/10.1016/j.aej.2021.03.048. ISSN 1110-0168
10. Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., et al.: Connected-UNets: a deep learning architecture for breast mass segmentation. Breast Cancer **7**, 151 (2021). https://doi.org/10.1038/s41523-021-00358-x
11. Huang, M.-L., Lin, T.-Y.: Double-dilation non-pooling convolutional neural network for breast mass mammogram image classification. Bahrain Med. Bull. **44**(4), 1144 (2022)
12. Ribli, D., Horváth, A., Unger, Z., et al.: Detecting and classifying lesions in mammograms with deep learning. Sci. Rep. **8**, 4165 (2018). https://doi.org/10.1038/s41598-018-22437-z
13. Chougrad, H., Zouaki, H., Alheyane, O.: Deep convolutional neural networks for breast cancer screening. Comput. Methods Program. Biomed. **157**, 19–30 (2018). https://doi.org/10.1016/j.cmpb.2018.01.011

Author Queries

Chapter 34

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author email address has been identified as per the information available in the Copyright form.	
AQ2	As Per Springer style, both city and country names must be present in the affiliations. Accordingly, we have inserted the city and country names in 1 and 2 affiliations. Please check and confirm if the inserted city and country names are correct. If not, please provide us with the correct city and country names.	
AQ3	Please check and confirm if the inserted citations of Figs. 1–8, Tables 1, 2, 4, 5 are correct. If not, please suggest an alternate citations.	