Deep learning algorithms for breast cancer detection from ultrasound scans

Lawysen, Gede Putra Kusuma
Department of Computer Science, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Sep 13, 2024
Revised Dec 2, 2024
Accepted Dec 15, 2024

Keywords:

Breast cancer detection
Comparative analysis
Deep learning algorithm
Resnet-50 model
Ultrasound scan

ABSTRACT

Breast cancer is a highly dangerous disease and the leading cause of cancer-related deaths among women. Early detection of breast cancer is considered quite challenging but can offer significant benefits, as various treatment interventions can be initiated earlier. The focus of this research is to develop a model to detect breast cancer based on ultrasound results using deep learning algorithms. In the initial stages, several preprocessing processes, including image transformation and image augmentation were performed. Two types of models were developed: utilizing mask files and without using mask files. Two types of models were developed using four deep learning algorithms: residual network (ResNet)-50, VGG16, vision transformer (ViT), and data-efficient image transformer (DeiT). Various algorithms, such as optimization algorithms, loss functions, and hyperparameter tuning algorithms, were employed during the model training process. Accuracy used as the performance metric to measure the model's effectiveness. The model developed with ResNet-50 became the best model, achieving an accuracy of 94% for the model using mask files. In comparison, the model developed with ResNet-50 and DeiT became the best model for the model without mask files, with an accuracy of 80%. Therefore, it can be concluded that using mask files is crucial for producing the best-performing model.

Corresponding Author:

Lawysen
Department of Computer Science, BINUS Graduate Program - Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia, 11480
Email: lawysen@binus.ac.id

1. INTRODUCTION

Breast cancer is one of the most dangerous diseases and is one of the leading causes of cancer-related deaths among women [1]. Diagnosing breast cancer manually can be time-consuming and costly. Additionally, detecting abnormalities at an early stage is very challenging due to the subtlety of the initial symptoms [2]. Early detection of cancer is crucial because the sooner a cancer is diagnosed, the sooner appropriate treatment can be administered. Therefore, having a system that can detect breast cancer at an early stage allows patients to receive appropriate treatment and management earlier, significantly increasing their chances of survival.

There are two methods for detecting breast cancer, namely mammography and ultrasound scanning. Many studies have been conducted by previous researchers to detect breast cancer using various machine learning and deep learning algorithms. Mammography is one of the best methods for diagnosing breast cancer, where images are taken from multiple angles to provide a comprehensive view of the patient's breast condition [1]. The X-ray results are then examined to determine if there is a tumor in the breast. If a tumor is present, it is further investigated to ascertain whether it is benign or malignant (cancer) [2]. In this research, various algorithms are used, including machine learning algorithms such as support vector machine (SVM) [3], [4], extreme learning machine [5], [6], as well as deep learning algorithms like convolutional neural network (CNN) [4], [7]-[10], residual network (ResNet)-50 [10], [11], and you only look once (YOLO) [12] to classify whether the mammogram images contain benign tumors, cancer, or are normal (no tumor).

In addition to utilizing mammogram image datasets, there are also studies that utilize ultrasound scan image datasets. Ultrasound is one of the most widely used methods in the medical field due to its ability to produce high-quality images of a patient's internal organs [13]. In previous studies, several researchers have employed various machine learning algorithms, such as K-nearest neighbor (KNN), decision tree (DT), random forest (RF) classifier, and SVM [14]-[16], for classification. In addition to machine learning algorithms, there are also researchers who have utilized deep learning algorithms, including ResNet-50 [13] and DenseNet-121 [17], for classification purposes.

Previous research has predominantly utilized mammogram data rather than ultrasound scan data. However, in recent years, researchers have begun to shift towards using ultrasound scan data over mammogram data. According to several studies, ultrasound scanning offers certain advantages compared to mammography. For example, ultrasound scans are considered safer than mammograms as they do not involve any radiation [18]. The absence of radiation makes periodic examinations of patients much safer. Additionally, ultrasound scanning has a higher sensitivity for detecting breast cancer in younger women [19], allowing for earlier intervention and action. Therefore, this study employs ultrasound scan datasets.

The algorithms used in this paper to develop models for detecting breast cancer based on ultrasound scan results are deep learning algorithms, specifically ResNet-50, VGG16, vision transformer (ViT), and data-efficient image transformer (DeiT). These algorithms were chosen due to their widespread use in various image processing applications, including medical image classification (X-ray, Ultrasound, and magnetic resonance imaging (MRI)). Additionally, these algorithms have been employed in numerous previous studies and have demonstrated the capability to produce models with commendable performance [10], [11], [13], [17], [20]. The models generated by these four algorithms will be compared, and the model with the highest performance among them will be selected. All four models will undergo identical preprocessing, training, and testing stages to ensure a fair comparison of their performance.

The dataset used for developing the model consists of ultrasound scan images obtained from Kaggle. This dataset is unique because it includes mask image files. These mask images represent the shape, size, and location of tumors from the ultrasound scans. Therefore, this study utilizes the mask files with the assumption that, in practical applications, such mask files can be obtained through image segmentation processes. Before model development, the dataset will undergo preprocessing process to prepare the data through data transformation. This process aims to prepare and adjust the data for training and testing the model. Additionally, various image augmentation techniques will be applied to balance the number of samples across different classes and enhance the data variability [14], with the goal of achieving good model performance. After completing the preprocessing steps, the model training and testing phases will be conducted, and the performance of the resulting models were compared against each other.

2. LITERATURE REVIEW

For the literature review discussed, there are two types: literature review for various papers that focus on detecting breast cancer using ultrasound results and papers that focus on detecting breast cancer using mammogram results. The purpose of discussing these two types of papers is to explore the latest trends in breast cancer detection. The insights gained will help guide the selection of the most suitable algorithm for developing the detection model.

2.1. Literature review based on mammogram

The solution proposed in this study utilizes cellular neural networks to segment suspicious regions in mammogram images and employs the SVM algorithm for classifying the previously processed images. Sampaio *et al.* [4] specifically explores this approach, using the public DDSM dataset, which consists of 2620 mammogram screening images. However, only 623 images from this dataset were used in the research. The performance metrics used to evaluate the model are accuracy (AC) and area under the ROC curve (AUC). The model's performance achieved an accuracy of 84.62% and an AUC of 0.87.

A different approach is explored by combining deep learning with case-based reasoning (CBR) Systems to enhance the accuracy of classification models. This method utilizes deep learning for precise segmentation of mammogram images and integrates CBR for both accurate and explainable classification. The deep learning algorithms employed for image segmentation include a hybrid of deep neural networks (DNNs) and SE-ResNet. The study uses the CBIS-DDSM dataset, which contains mammogram images, to test the model. The performance metrics for evaluation include accuracy and recall, with the model achieving an accuracy of 86.71% and a recall of 91.34%, as discussed in Benlabiod *et al.* [7].

Vijayarajeswari *et al.* [3], a solution is proposed that employs maximization estimation during the preprocessing stage and utilizes hough transform for feature extraction from the processed images. The SVM algorithm is then applied to classify breast cancer based on these processed images. The dataset used in this study is the MIAS database, which contains a total of 417 mammographic images. The performance of the model is evaluated using accuracy as the metric. The authors compared their model to various previous approaches and found that their SVM-based model achieved the highest accuracy, reaching 94%.

A solution is developed for detecting breast cancer from screening mammograms using deep learning algorithms with end-to-end training. This study utilizes two CNN architectures: ResNet-50 and VGG (VGG16). The research employs two datasets: CBIS-DDSM, which contains 2,478 mammogram images, and INbreast, with 410 mammogram images. The model's performance is evaluated using the area under the curve (AUC) as the primary metric, achieving an AUC value of 0.95, as detailed in paper [11].

A solution is introduced for diagnosing breast cancer from digital mammography results using the YOLO algorithm. Aly *et al.* [12] specifically compares the performance of three YOLO architectures: YOLO-V1, YOLO-V2, and YOLO-V3. The study utilizes the INbreast dataset, which contains 410 digital mammography images. The performance of the three architectures is evaluated using accuracy as the metric, with YOLO-V3 achieving the highest accuracy at 95.5%.

The development of a model using a lightweight deep CNN named DisepNet is proposed, incorporating innovative feature extraction techniques called Disep block and Incep-L block. Yu *et al.* [8] focuses on this approach, utilizing the MINI-MIAS and INbreast datasets. The model's performance is evaluated through specificity, sensitivity, and accuracy metrics, with results showing a specificity of 0.9744, a sensitivity of 0.9371, and an accuracy of 0.956.

An automated solution is presented for classifying mammography images without the need for labeling the region-of-interest (ROI), using a multi-scale DNN model. Xie *et al.* [9] explores this approach, incorporating a breast region segmentation (BRS) module for preprocessing and a CNN algorithm for feature extraction and classification. The study employs DenseNet and MobileNet algorithms, utilizing the INbreast dataset, which includes 410 digital mammography images. The performance of the model is evaluated based on accuracy and AUC, with the model built using DenseNet, Multi-scale, and the BRS module achieving the highest accuracy of 96.34% and an AUC of 0.9713.

Al- Antari *et al.* [10] proposes the development of a model utilizing the YOLO algorithm to detect breast lesions in mammogram images, followed by classification using CNN, ResNet-50, and InceptionResNet-V2 algorithms. The study employs two datasets, DDSM and INbreast, consisting of two classes: malignant and benign. The paper uses k-fold cross-validation to train, validate, and test the resulting model. Before training, validation, and testing, the datasets undergo data transformation and data augmentation processes to address class imbalance within the datasets. The performance metric used in this paper to evaluate the model is accuracy, with performance measured separately for each algorithm and dataset. For the DDSM dataset, the model built with CNN achieved an accuracy of 94.5%, ResNet-50 achieved 95.83%, and InceptionResNet-V2 achieved 97.5%. For the INbreast dataset, the model built with CNN achieved an accuracy of 88.74%, ResNet-50 achieved 92.55%, and InceptionResNet-V2 achieved 95.32%.

A model development solution using extreme learning machine, enhanced by the crow-search optimization algorithm, is proposed. Chakravarthy and Rajaguru [5] discusses this approach, which first employs the ResNet-18 algorithm for feature extraction. The study utilizes three mammogram datasets-DDSM, MIAS, and INbreast-each comprising three classes: normal, malignant, and benign. The performance of the proposed model is evaluated using accuracy, with results showing 97.193% on the DDSM dataset, 98.137% on the MIAS dataset, and 98.266% on the INbreast dataset.

A model development solution is introduced using lifting wavelet transform (LWT) for feature extraction, along with principal component analysis (PCA) and linear discriminant analysis (LDA) for feature vector reduction. The classification process combines extreme learning machine with moth flame optimization (MFO-ELM). Prior to these steps, the images undergo ROI extraction using the cropping method. The study utilizes the MIAS dataset, which includes normal and abnormal images, and the DDSM dataset, comprising benign and malignant images. The model is evaluated using accuracy, achieving 99.76% on the MIAS dataset and 98.80% on the DDSM dataset, as described in Maduli *et al.* [6].

2.2. Literature review based on ultrasound

Lee *et al.* [17], a deep learning model development solution is proposed, where object detection is first applied using mask R-CNN to locate and identify the tumor region in ultrasound images. For classification, the paper develops a model using the DenseNet-121 algorithm along with the ADAM optimization algorithm. The dataset used in this study was created by the authors using data from 153 patients at National Taiwan University Hospital (NTUH). The performance metrics used to evaluate the model include accuracy, recall, specificity, precision, NPV, and AUC. The model achieved an accuracy of 81.05%, a recall of 81.36%, a specificity of 80.85%, a precision of 72.73%, an NPV of 87.36%, and an AUC of 0.8054. The performance of the model was also compared with other models developed using different machine learning algorithms such as logistic regression (LR), SVM, and XGBoost.

A computer-aided diagnosis solution is presented for classifying ultrasound images into benign, malignant, or normal cases, using a combination of feature extraction and classification methods. Khanna *et al.* [21] explores this approach, utilizing the pre-trained CNN algorithm ResNet-50 for feature extraction, followed by binary gray wolf optimization for feature selection, and finally employing the SVM algorithm for classification. The dataset used is the BUSI dataset, containing 780 images, including 437 benign, 210 malignant, and 133 normal cases. The model is evaluated using accuracy and AUROC, achieving an accuracy of 84.9% and an AUROC of 0.97.

Another model development solution is proposed, leveraging a hybrid CNN-ViT architecture combined with an MLP mixer. Tagnamas *et al.* [22] discusses the use of two encoders, EfficientNet V2 and ViT, for feature extraction. The extracted features are fused using channel attention fusion, with the MLP-mixer handling classification. This study also utilizes the BUSI dataset, and the performance is measured by accuracy, with the model achieving an accuracy of 86%.

A computer-aided diagnosis (CAD) solution is developed using the speckle reducing anisotropic diffusion (SRAD) method for preprocessing, along with the active contour model for image segmentation. Feature extraction is performed using the Gray Level Co-occurrence Matrix, and various machine learning algorithms, including KNN, DT, and RF classifier, are employed for classification. The study utilizes a publicly available ultrasound image dataset containing 780 breast ultrasound images, though only 160 images are used, 100 for training and 60 for testing. The model is evaluated using accuracy as the performance metric, with the RF algorithm achieving the highest accuracy of 88%, as discussed in Pavithra *et al.* [14].

Abhisheka *et al.* [23], a solution is proposed that combines histogram-oriented gradient (HOG) and local binary pattern (LBP) for extracting local features and employs ResNet-50 for extracting global features. For the classification model, the paper uses the SVM algorithm. The dataset used in this study is the BUSI dataset, and 5-fold cross-validation is applied to train the model. The performance metric used is accuracy, with the proposed model achieving an accuracy of 88.87%.

A solution is introduced that leverages both supervised and unsupervised algorithms for feature selection, combining ResNet-34 as the supervised algorithm and a convolutional autoencoder (CAE) as the unsupervised one. Song and Kim [24] discusses this approach, utilizing several machine learning algorithms for classification, including DT, KNN, SVM, and RF. The dataset used in the study is the BUSI dataset, comprising 780 images, with 437 benign, 210 malignant, and 133 normal cases. The model is evaluated using performance metrics such as accuracy, sensitivity, specificity, AUC, and ACA, achieving an accuracy of 88.18%, a sensitivity of 85.25%, a specificity of 100%, an AUC of 85.7%, and an ACA of 88.18%.

A model creation solution is presented utilizing four machine learning algorithms: LDA, SVM, RF, and DT. Yao *et al.* [15] outlines the process, starting with ROI segmentation on ultrasound images to manually determine the tumor position using open-source software. Feature extraction and feature selection follow, as the model is built using these machine learning algorithms. The dataset includes breast cancer patient data with a total of 278 patients. The performance metrics used for evaluation are AUC, sensitivity, specificity, and accuracy, with the SVM-based model achieving the best results: an AUC of 0.934, sensitivity of 86.7%, specificity of 89.9%, and accuracy of 91%.

Ding *et al.* [16] proposes a model creation solution using multiple-instance learning (MIL) to classify breast cancer from ultrasound images. This paper first processes the images, including segmentation and texture classification to estimate the ROI, which then undergoes various other processes before being classified using the SVM algorithm to determine whether the tumor is benign or malignant (cancer). The dataset used in this paper consists of ultrasound images, including 168 images with 72 malignant and 96 benign. The performance metric used to evaluate the model is accuracy, with the model achieving an accuracy of 91%.

A breast cancer detection network called BCDNet is introduced, utilizing deep learning for both feature extraction and classification. The paper tests and compares several deep learning algorithms to identify the one with the best performance. The study uses the breast ultrasound images dataset, which consists of 813 images, including 166 normal samples and 647 cancer samples. The performance of the model is evaluated using sensitivity, precision, F1-score, and accuracy. ResNet-50 is identified as the top-performing algorithm, achieving an accuracy of 93.97%, along with the highest sensitivity (95.24%) and F1-score (96.42%). However, its precision is the second highest at 97.68%, as detailed in Lu *et al.* [13].

2.3. Summary of literature review

This study utilizes the ultrasound image dataset (BUSI dataset). This dataset is a public dataset available from Kaggle [25]. Among the various ultrasound datasets available, this one was chosen because it has a relatively large amount of data compared to other ultrasound datasets. Additionally, this dataset includes tumor mask images on the ultrasound images, making it quite unique compared to other ultrasound datasets. The presence of mask image files in this dataset allows this study to evaluate the impact of using mask images in model creation.

The primary reason for using ultrasound images in this paper is that researchers initially preferred to use mammogram image datasets for their studies. However, in recent years, researchers have started using ultrasound image datasets for their research. Therefore, this study uses ultrasound images for its dataset. A review of the literature shows that earlier methods for this type of research involved feature extraction using either specialized algorithms or deep learning algorithms before applying machine learning algorithms for classification, with SVM being a popular choice. Recently, researchers have increasingly relied on deep learning algorithms, with popular choices including ResNet, VGG, and CNN.

Considering these developments, this study utilizes several well-known algorithms used by previous researchers, including ResNet-50 and VGG16, and also explores two relatively new algorithms in computer vision: ViT and DeiT. This study evaluates the models created with these four algorithms to determine which algorithm produces the best-performing model. The comparison of models is conducted fairly, with each image undergoing the same preprocessing steps. In addition to comparing models, this study also evaluates models created with and without the use of masks to assess the impact of masks on model performance. The masks are images obtained from the image segmentation process during preprocessing, and this dataset already includes mask files.

3. RESEARCH METHOD

The methods utilized in this research will be discussed in this methodology chapter. Figure 1 shows an overview of the sequence of various steps in the proposed solution design. The initial stage of this process is to prepare the dataset that will be used to create the model. Next, if the developed model utilizes a mask file, a center crop process is performed on the ultrasound scan images based on the mask image, with the aim of focusing the ultrasound image solely on the tumor. If the developed model does not utilize a mask file, the process continues to the next stage. The following step is to split the center-cropped images into training data, validation data, and testing data. After the data has been split, various preprocessing steps are carried out, which include image transformation and image augmentation.

The initial stage of this process is to prepare the dataset that will be used to create the model. Next, if the developed model utilizes a mask file, a center crop process is performed on the ultrasound scan images based on the mask image, with the aim of focusing the ultrasound image solely on the tumor. If the developed model does not utilize a mask file, the process continues to the next stage. The following step is to split the center-cropped images into training data, validation data, and testing data. After the data has been split, various preprocessing steps are carried out, which include image transformation and image augmentation.

Once the data is ready to be used, the process of model creation begins. This model is trained using the training and validation data that was prepared in the previous stages. The resulting model undergoes fine-tuning using optimization algorithms, loss functions, and Bayesian optimization, aimed at finding parameters that can produce a model with the best performance, characterized by a low loss value. After the model has completed various training processes, the next step is to test the model by predicting the testing data that was prepared earlier, and the prediction results are measured using performance metrics for evaluation. A more detailed explanation of the steps depicted in Figure 1 will be provided in the following sub-sections.

3.1. Dataset

The dataset used for model creation is an ultrasound image dataset obtained from Kaggle [25]. This dataset contains 3 classes: normal, benign, and malignant. It includes 1,578 images, divided into 421 malignant images, 891 benign images, and 266 normal images. Each ultrasound image in this dataset has a corresponding mask image, as shown in Figure 2 (Figure 2(a) and Figure 2(b)). The mask images are utilized to perform a center crop on the ultrasound images. The purpose of this center crop is to ensure that the ultrasound images focus solely on the tumor present in the ultrasound image. By performing the center crop based on the mask image, all ultrasound images in the normal class are not subjected to center cropping. This is because no tumor is detected in the mask images. An example of the center crop process applied to ultrasound images based on the mask image can be seen in Figure 3 (Figure 3(a) and Figure 3(b)). The following are the data counts after the cropping process: the malignant class contains 210 images, the normal class contains 133 images, and the benign class contains 437 images.

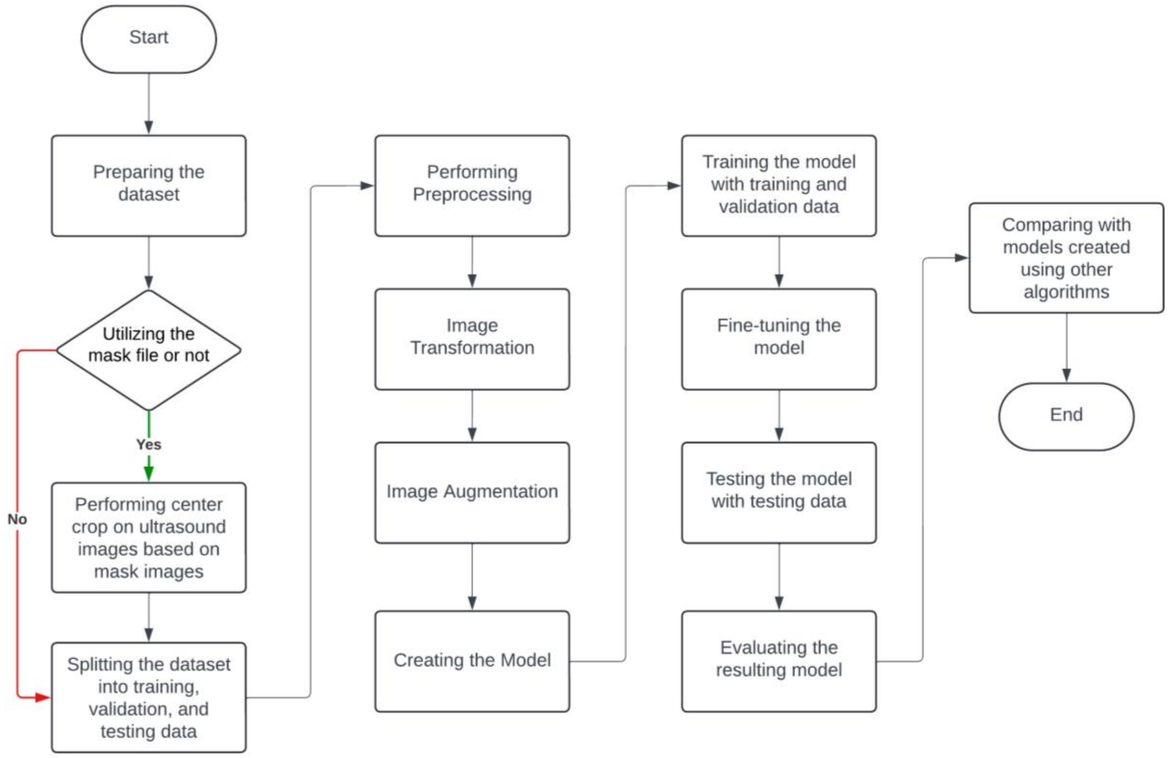


Figure 1. Overview of the various stages carried out in the research

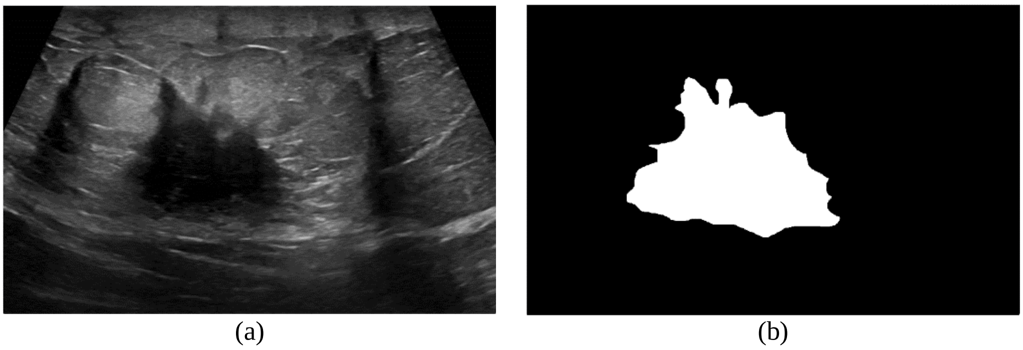


Figure 2. Examples of images in the dataset (a) example image of an ultrasound result in malignant class and (b) example image for mask file

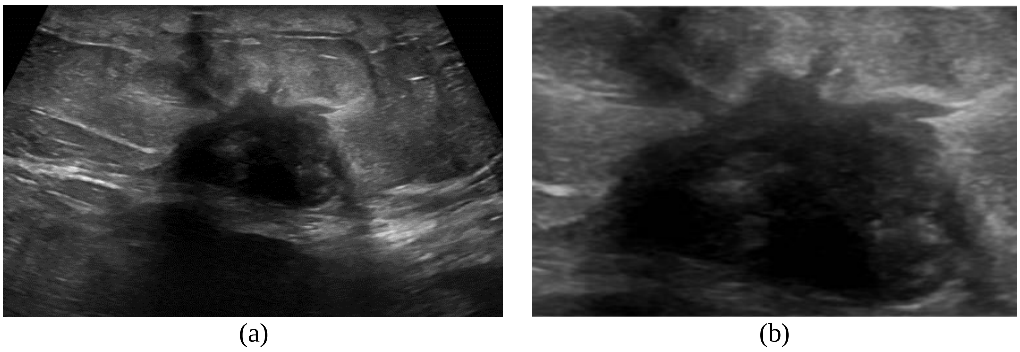


Figure 3. An example of the results from the cropping process (a) example image of the original image and (b) the center-cropped image

3.2. Preprocessing

In the preprocessing stage, the images used for training the model are processed first, with the expectation that the resulting model will have good performance. There are two processes conducted during the preprocessing stage: image transformation and image augmentation. Here are some of the steps taken during the image transformation stage:

– Resizing all images in the dataset to ensure that every image has a uniform size of 256×256 pixels.
– Converting the images to tensor format.
– Normalizing the pixel values in the images using the specified mean and standard deviation values. The mean values used are (0.485, 0.456, 0.406) and the standard deviation values used are (0.299, 0.224, 0.225).

The image augmentation process is also carried out with the aim of increasing the variation in the data used. Here are some steps taken during the image augmentation stage:

– Flipping the images horizontally with a 90% probability ($p = 0.9$).
– Randomly rotating the images by 15 degrees without enlarging the image size to ensure that the images remain the same size as the others.
– Adjusting the color of the existing images by setting several parameters such as brightness with a value of 0.2, contrast with a value of 0.2, saturation with a value of 0.2, and hue with a value of 0.2.

3.3. Model development and evaluation

For model development, four deep learning algorithms were utilized: Resnet-50, VGG16, ViT, and DeiT. These four algorithms were implemented as pretrained models using the torchvision library. The choice of the torchvision library is due to its being part of the PyTorch ecosystem, which is specifically designed to handle computer vision tasks such as classification. The models were trained using prepared training and validation datasets.

During the model training process, several algorithms were employed to maximize the model's training efficiency, including the SGD optimizer, the cross-entropy loss function to measure the loss value during model training, and the Bayesian optimization algorithm for hyperparameter tuning. Additionally, an early stop function was implemented during training, designed to halt the training process if there is no decrease in validation loss. The training process will stop if there is no change in validation loss for 10 consecutive epochs. The purpose of using the early stop function is to prevent model overfitting.

After completing the entire model training process and obtaining the best-performing model, the model was tested by classifying the prepared testing data. The performance of the resulting model is measured by how well it classifies the images from the testing data. The model selected is the one with the best performance among the four models developed. The confusion matrix and classification report are used to measure the performance of the developed model. The confusion matrix provides an overview of the model's predictions on the testing data, while the classification report presents various performance metrics such as accuracy, precision, recall, and F1-score. However, for comparing the performance between different models, the focus will primarily be on the accuracy value.

4. RESULT AND DISCUSSION

The results discussed from the research are the outcomes of the training and validation processes conducted for all models during the training phase, as well as the testing results of all the models that have been developed. The loss and accuracy values of the model during the training process are displayed using plots and tables to display the detailed numbers. While the accuracy values of the model at the testing stage are displayed using tables only.

4.1. Training and validation results

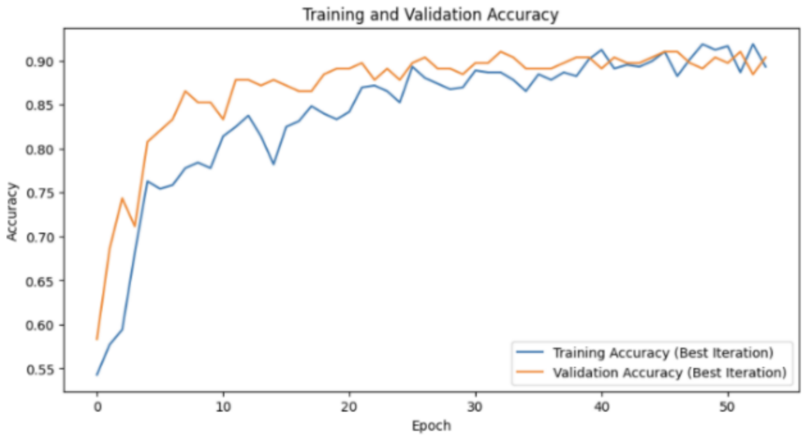
Figure 4 (Figure 4(a) and Figure 4(b) show an example of a plot displaying the loss and accuracy values during the training process of a model. This plot details the loss and accuracy values of the model from the beginning to the end of the training epoch. The purpose of displaying the loss and accuracy plot during training is to check for any occurrence of overfitting in the model during the training process.

Table 1 shows the performance of the model during the training and validation process when using the mask file, while Table 2 displays the performance of the model during training and validation without using the mask file. In the training data, the model developed with the mask file shows that the DeiT performed the best, with the lowest loss value of 0.128164 and the highest accuracy of 0.965812. Conversely, in the model developed without using the mask file, the ViT had the lowest loss value of 0.222382 and the highest accuracy of 0.929487.

For the validation process, the model developed with the mask file shows that the ResNet-50 had the highest accuracy of 0.903846, although it did not have the lowest loss value. Meanwhile, the model developed without using the mask file shows that the VGG16 performed best, with the highest accuracy of 0.903846 and the lowest loss value of 0.275931. The impact of using the mask file is evident in the training data, where models tend to perform better compared to models developed without the mask file. However, in the validation process, the VGG16 performed better in the model developed without the mask file.

Overall, using the mask file significantly improves model performance in the training data. The VGG16 showed good and consistent performance in both conditions, especially in the validation data. The ViT and DeiT demonstrated excellent performance in the training data with the mask file, but did not yield the same results in the validation data. The ResNet-50 had more stable performance in the validation data when using the mask file.

(a)



(b)

Figure 4. An example of a model performance plot showing the results from both the training and validation process, (a) example plot for loss values and (b) example plot for accuracy values

Table 1. The results of the model training and validation utilizing the mask file

	Utilizing the mask file			
	Train		Validation	
Model	Loss	Accuracy	Loss	Accuracy
ResNet-50	0.284669	0.899573	0.304066	0.903846
VGG16	0.236413	0.910256	0.277411	0.884615
Vision transformer	0.139979	0.957265	0.305617	0.871795
Data-efficient image transformer	0.128164	0.965812	0.325701	0.878205

Table 2. The results of the model training and validation without utilizing the mask file

	Without utilizing the mask file			
	Train		Validation	
Model	Loss	Accuracy	Loss	Accuracy
ResNet-50	0.397846	0.837607	0.334732	0.858974
VGG16	0.300331	0.867521	0.275931	0.903846
Vision transformer	0.222382	0.929487	0.592379	0.788462
Data-efficient image transformer	0.243050	0.910256	0.553510	0.775641

4.2. Testing results

Table 3 presents a summary of the testing results for several models tested under two conditions: models utilizing the mask file and models not utilizing the mask file. The tested models were developed using four algorithms: ResNet-50, VGG16, ViT, and DeiT. In the first condition, the performance of all the tested models can be considered very good. The model developed with ResNet-50 achieved an accuracy of 94% and a weighted accuracy of 93%. Similarly, the model developed with VGG16 performed just as well as ResNet-50, also achieving an accuracy and weighted accuracy of 93%. The model developed with ViT, although not performing as well as ResNet-50 and VGG16, still showed good performance with an accuracy and weighted accuracy of 90%. Lastly, the model developed with DeiT also exhibited strong performance, achieving accuracy and weighted accuracy of 92%.

Table 3. Summary of the testing results for all models

	Utilizing the mask file		Without utilizing the mask file	
Model	Accuracy	Weighted	Accuracy	Weighted
ResNet-50	94%	93%	80%	80%
VGG16	93%	93%	76%	76%
Vision transformer	90%	90%	76%	76%
Data-efficient image transformer	92%	92%	80%	80%
Average	92%	92%	78%	78%

On the other hand, the performance of models that did not utilize the mask file experienced a significant decline compared to the models' performance in the first condition. The models developed with ResNet-50 and DeiT only achieved an accuracy and weighted accuracy of 80%. This represents a performance drop of 14% for the ResNet-50 model and a 12% drop for the DeiT model. Furthermore, the models developed with VGG16 and ViT only achieved an accuracy and weighted accuracy of 76%. With these figures, the VGG16 model experienced the largest performance decline, dropping by 17%, while the ViT model saw a decline of 14%. From the average performance of all models developed with and without the mask file, it was found that, overall, models developed with the mask file performed significantly better compared to models not utilizing the mask file. Based on these results, it can be concluded that the utilization of the mask file is crucial in improving the predictive capability of the developed models.

5. CONCLUSION

In this study, we developed models to detect breast cancer using ultrasound scan results, focusing on two approaches: with and without mask files. Both approaches utilized four deep learning algorithms: ResNet-50, VGG16, ViT, and DeiT. The models underwent extensive training to optimize performance, with results showing that models using mask files significantly outperformed those without. The model that demonstrated the best performance among all those developed was the one using mask files in conjunction with the ResNet-50 algorithm, achieving an impressive accuracy of 94%.This research underscores the potential of deep learning in breast cancer detection, particularly with the use of mask files. However, it also highlights the need for future studies to explore newer algorithms and leverage larger datasets to enhance model generalization and accuracy.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Lawysen	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Gede Putra Kusuma	✓			✓		✓				✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis

I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting

Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state there is no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset.

REFERENCES

[1] A. Sahu, P. K. Das, and S. Meher, "Recent advancements in machine learning and deep learning-based breast cancer detection using mammograms," *Physica Medica*, vol. 114, p. 103138, Oct. 2023, doi: 10.1016/j.ejmp.2023.103138.

[2] A. Sahu, P. K. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," *Biomedical Signal Processing and Control*, vol. 87, p. 105377, Jan. 2024, doi: 10.1016/j.bspc.2023.105377.

[3] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement: Journal of the International Measurement Confederation*, vol. 146, pp. 800–805, Nov. 2019, doi: 10.1016/j.measurement.2019.05.083.

[4] W. B. Sampaio, E. M. Diniz, A. C. Silva, A. C. de Paiva, and M. Gattass, "Detection of masses in mammogram images using CNN, geostatistic functions and SVM," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 653–664, Aug. 2011, doi: 10.1016/j.compbiomed.2011.05.017.

[5] S. R. S. Chakravarthy and H. Rajaguru, "Automatic detection and classification of mammograms using improved extreme learning machine with deep learning," *Irbm*, vol. 43, no. 1, pp. 49–61, Feb. 2022, doi: 10.1016/j.irbm.2020.12.004.

[6] D. Muduli, R. Dash, and B. Majhi, "Automated breast cancer detection in digital mammograms: a moth flame optimization based ELM approach," *Biomedical Signal Processing and Control*, vol. 59, p. 101912, May 2020, doi: 10.1016/j.bspc.2020.101912.

[7] L. B.-Benlabiod, K. Harrar, L. Yamoun, M. Y. Khodja, and M. A. Akhloufi, "A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification," *Computers in Biology and Medicine*, vol. 163, p. 107133, Sep. 2023, doi: 10.1016/j.compbiomed.2023.107133.

[8] X. Yu, K. Xia, and Y. D. Zhang, "DisepNet for breast abnormality recognition," *Computers and Electrical Engineering*, vol. 90, p. 106961, Mar. 2021, doi: 10.1016/j.compeleceng.2020.106961.

[9] L. Xie, L. Zhang, T. Hu, H. Huang, and Z. Yi, "Neural networks model based on an automated multi-scale method for mammogram classification," *Knowledge-Based Systems*, vol. 208, p. 106465, Nov. 2020, doi: 10.1016/j.knosys.2020.106465.

[10] M. A. Al-antari, S. M. Han, and T. S. Kim, "Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105584, Nov. 2020, doi: 10.1016/j.cmpb.2020.105584.

[11] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.

[12] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO based breast masses detection and classification in full-field digital mammograms," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105823, Mar. 2021, doi: 10.1016/j.cmpb.2020.105823.

[13] S. Y. Lu, S. H. Wang, and Y. D. Zhang, "BCDNet: an optimized deep network for ultrasound breast cancer detection," *IRBM*, vol. 44, no. 4, p. 100774, Aug. 2023, doi: 10.1016/j.irbm.2023.100774.

[14] S. Pavithra, R. Vanithamani, and J. Justin, "Computer aided breast cancer detection using ultrasound images," *Materials Today: Proceedings*, vol. 33, pp. 4802–4807, 2020, doi: 10.1016/j.matpr.2020.08.381.

[15] J. Yao *et al.*, "Machine learning-based breast tumor ultrasound radiomics for pre-operative prediction of axillary sentinel lymph node metastasis burden in early-stage invasive breast cancer," *Ultrasound in Medicine and Biology*, vol. 50, no. 2, pp. 229–236, Feb. 2024, doi: 10.1016/j.ultrasmedbio.2023.10.004.

[16] J. Ding, H. D. Cheng, J. Huang, J. Liu, and Y. Zhang, "Breast ultrasound image classification based on multiple-instance learning," *Journal of Digital Imaging*, vol. 25, no. 5, pp. 620–627, Oct. 2012, doi: 10.1007/s10278-012-9499-x.

[17] Y. W. Lee, C. S. Huang, C. C. Shih, and R. F. Chang, "Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks," *Computers in Biology and Medicine*, vol. 130, p. 104206, Mar. 2021, doi: 10.1016/j.compbiomed.2020.104206.

[18] P. Crystal, S. D. Strano, S. Shcharynski, and M. J. Koretz, "Using sonography to screen women with mammographically dense breasts," *American Journal of Roentgenology*, vol. 181, no. 1, pp. 177–182, Jul. 2003, doi: 10.2214/ajr.181.1.1810177.

[19] E. Devolli-Disha, S. Manxhuka-Kërliu, H. Ymeri, and A. Kutllovci, "Comparative accuracy of mammography and ultrasound in women with breast symptoms according to age and breast density," *Bosnian Journal of Basic Medical Sciences*, vol. 9, no. 2, pp. 131–136, May 2009, doi: 10.17305/bjbms.2009.2832.

[20] S. Sajed, A. Sanati, J. E. Garcia, H. Rostami, A. Keshavarz, and A. Teixeira, "The effectiveness of deep learning vs. traditional methods for lung disease diagnosis using chest X-ray images: A systematic review," *Applied Soft Computing*, vol. 147, p. 110817, Nov. 2023, doi: 10.1016/j.asoc.2023.110817.

[21] P. Khanna, M. Sahu, and B. K. Singh, "Improving the classification performance of breast ultrasound image using deep learning and optimization algorithm," in *2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES)*, Dec. 2021, pp. 1–6, doi: 10.1109/TRIBES52498.2021.9751677.

[22] J. Tagnamas, H. Ramadan, A. Yahyaouy, and H. Tairi, "Multi-task approach based on combined CNN-transformer for efficient segmentation and classification of breast tumors in ultrasound images," *Visual Computing for Industry, Biomedicine, and Art*, vol. 7, no. 1, p. 2, Jan. 2024, doi: 10.1186/s42492-024-00155-w.

[23] B. Abhisheka, S. K. Biswas, D. Saha, S. Das, and B. Purkayastha, "The efficacy of combining deep and handcrafted features for breast cancer classification using ultrasound images," in *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Dec. 2023, pp. 735–740, doi: 10.1109/UPCON59197.2023.10434836.

[24] M. Song and Y. Kim, "Deep representation for the classification of ultrasound breast tumors," in *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Jan. 2022, pp. 1–6, doi: 10.1109/IMCOM53663.2022.9721796.

[25] A. Shah, "Breast ultrasound images dataset," *Kaggle Web*, 2021. https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset?resource=download (accessed Mar. 15, 2024).

BIOGRAPHIES OF AUTHORS

Lawysen was born in March 15, 2002 in the city of Jakarta, Indonesia. He obtained a Bachelor of Computers degree from Bina Nusantara University in 2024 and is currently studying to obtain a Master of Information Technology degree at Bina Nusantara University. His research interests are intelligent systems/machine learning. He can be contacted at email: lawysen@binus.ac.id.

Gede Putra Kusuma Gede Putra Kusuma received PhD degree in Electrical and Electronic Engineering from Nanyang Technological University (NTU), Singapore, in 2013. He is currently working as a Lecturer and Head of Department of Master of Computer Science, Bina Nusantara University, Indonesia. Before joining Bina Nusantara University, he was working as a Research Scientist in I2R – A*STAR, Singapore. His research interests include computer vision, deep learning, face recognition, appearance-based object recognition, gamification of learning, and indoor positioning system. He can be contacted at email: i.negara@binus.ac.id.