Research Article

LARGEST ROI SEGMENTATION FOR BREAST CANCER CLASSIFICATION USING A VGG16 DEEP LEARNING NETWORK

Thanh-Tam NGUYEN[1] , Thanh-Hai NGUYEN[1,*] , Ba-Viet NGO[1] ,
Thanh-Nghia NGUYEN[1]

[1]Department of Industrial Electronics and Biomedical Engineering, Faculty of Electrical – Electronics
Engineering, HCMC University of Technology and Education,
Ho Chi Minh City, Vietnam

tamnt.ncs@hcmute.edu.vn, nthai@hcmute.edu.vn, vietnb@hcmute.edu.vn, nghiant@hcmute.edu.vn

*Corresponding author: Thanh-Hai NGUYEN; nthai@hcmute.edu.vn

Abstract. *The exact evaluation of breast cancer images for patients is very important, because they can be early treated for lasting their life. This article proposes a classification system for finding breast cancer images, in which each breast lesion image is segmented to produce a largest Region of Interest (ROI) and a VGG16 deep learning network is applied for classification. An Otsu threshold is utilized on two datasets from two sources of CBIS-DDSM and MIAS to create largest ROI with main features. For the classification with high performance, two datasets of the breast lesions were augmented by rotating, flipping, and brightness variation. This article was proposed an algorithm with processing images sets before classification using VGG16. In particular, the results of the largest ROI datasets for four types of breast lesions were represented through segmentation, normalization and enhancement. In addition, the results of classifying four types of breast lesions (BC, BM, MC, MM) were evaluated using confusion matrix, with the high accuracy of around 95%. Another evaluation was that these image sets without ROI/with ROI parts/With the largest ROI only using the Otsu segmentation were compared and the highest accuracy was of the image sets with the largest ROI. The results with the high accuracy demonstrated to illustrate the effectiveness of the proposed method. It means that this method can be developed to classify many stages of breast cancers during diagnosis and treatment.*

Keywords

1. Introduction

Breast cancer in humans is one of the most dangerous diseases and also it is very difficult to treat [1]. The cause of breast cancer is still unclear, so it is very necessary to recognize soon for treatment. However, many researches have suggested that mutations of genes in the BRCA1 or BRCA2 are the main causes [2] such as environmental pollution, radiation, chemicals in food, smoking habits and other issues, which can increase the risk for having breast cancer treatment.

With many different cancers, breast cancer often appears in women, the majority of women are over 40 years old [3]. In particular, 2.3 million women were diagnosed with breast cancer and 685,000 died globally in 2020. About the end of 2020, there were 7.8 million women alive, who had been diagnosed with breast cancer in the past five years [4]. However, if cancers with early diagnosis can be detected, human can have a very high chance of treatment [5]. In particular, if the cancers are early detected in the first stage, the treatment rate is 80%, while second stage is about 60% [6].

An Otsu segmentation algorithm is often employed for determining thresholds to be able to ROI [7]. In particular, the Otsu segmentation is to determine gray level, where a gray level image is calculated to divide pixels into two groups: background pixels, and object pixels. In addition, breast images with many different parts on them are difficult for classifying high performance. Therefore, creating ROI from the breast image using segmentation methods is very necessary [8, 9, 10]. In this research, the largest ROI with features is detected before training VGG16. It is obvious that the obtained results can be better compared to previous works with the same datasets.

Image enhancement using histogram equalization method will make the image clearer and extract features [11]. It means that a histogram equalization technique is often used for enhancing contrast in a variety of different image types such as medical and satellite images. There are many methods of histogram equalization and essentially they are divided into two types of global and local histogram equalization. In particular, the algorithm proposed, in [11], used Gaussian Mixture Model for the model of the gray level distribution of the image. Furthermore, the intersections of the Gaussian Model were used to model the dynamic range of the images into input gray level intervals. Another research is that based on histogram equalization and superpixel segmentation, the method was introduced to find whether an image with grayscale. Thus, this paper can infer whether an image is dark, bright, or a small dynamic range [12].

In recent years, there has been a lot of researches on Artificial Intelligence (AI) in the medical field, particularly applying AI technology in identifying and classifying diseases such as heart [13, 14], human skin [15], brain cancer [16], breast cancer [17] and others. This article proposed an automatic segmentation method based on CNN with a 3x3 kernel. With this method of using a small kernel, it allows the design of a deep learning network architecture and also limits overfitting. The proposed method was applied on the CNN network for the MRI image set of the Brain Tumor Segmentation Challenge 2013 database (BRATS 2013) [16]. In these reseahces, AI has been applied for recognition of breast cancer and evaluating the level of cancers based on mammograms. In addition, using AI to classify and identify cancers, the cancers for early treatment are very important based on images taken from different technologies such as CT scanner, MRI [18], X-ray [19].

In the problem of identifying and classifying breast cancer, largest sets of mammograms with information related to breast lesions are provided by experienced doctors for building a machine learning system. In particular, the mammography data reports such as Breast Imaging_Reporting & Data System_BIRADS with 7

points from BIRADS_0 to BIRADS_6 are used with the machine learning systems for evaluation. With this evaluation, testing can be utilized to help doctors to reach a more exact conclusion related to breast cancer [20].

Currently, there are many researches using Convolutional Neural Networks (CNNs) for classifying breast lesions, in which breast cancer detection is very important [21]. In addition, the CNNs are applied for classifying benign and malignant tumors on digital images with the accuracy of up to 91%. Another CNN, called Convolutional Neural Network Improvement for Breast Cancer Classification (CNNI-BCC), was structured in 2019 and applied for classifying 3 lesion cases such as benign, malignant, and normal with its accuracy of 89.5%, 90.5% and 90.7%, respectively [22].

X-ray images of diseases related to breast lesions are often of inconsistent quality and also contain many unnecessary artifacts. Therefore, in most researches, image preprocessing algorithms are applied to uniformly improve image quality and enhance the effectiveness of classification. Thus, image preprocessing before applying for training advanced deep learning networks such as VGG16 [23], EfficientNet [24], or others can increase the effectiveness of classification performance. In this research, the extract largest ROI image normalization and enhancement will be employed before using training the VGG16 deep learning network for classification of breast lesions.

This article is to evaluate the effectiveness of image preprocessing with largest ROI and enhancement before training a deep learning network of VGG16. Moreover, four types of breast x-ray images (Benign Mass, Benign Calcification, Malignant Calcification, Malignant Mass) were obtained from CBIS-DDSM [25] and MIAS [26], in which data augmentation was suggested in this research so that the classification effectiveness is increased using VGG16. In particular, image datasets have been extracted ROI, enhanced and other processing methods are applied, including rotating, flipping images and their brightness changes. It is obvious that the classification performance is increased and the obtained results are also compared to previous works for evaluation of the proposed algorithm.

2. Materials and methods

In this article, the method of a process for classification of breast lesion is proposed as described in Fig. 1. In particular, datasets are obtained from CBIS-DDSM [25] and MIAS [26] and augmentation image datasets using rotating, flipping and brightness change, in which four types of lesion images (Benign Mass, Benign Calcification, Malignant Mass, Malignant Calcification) are
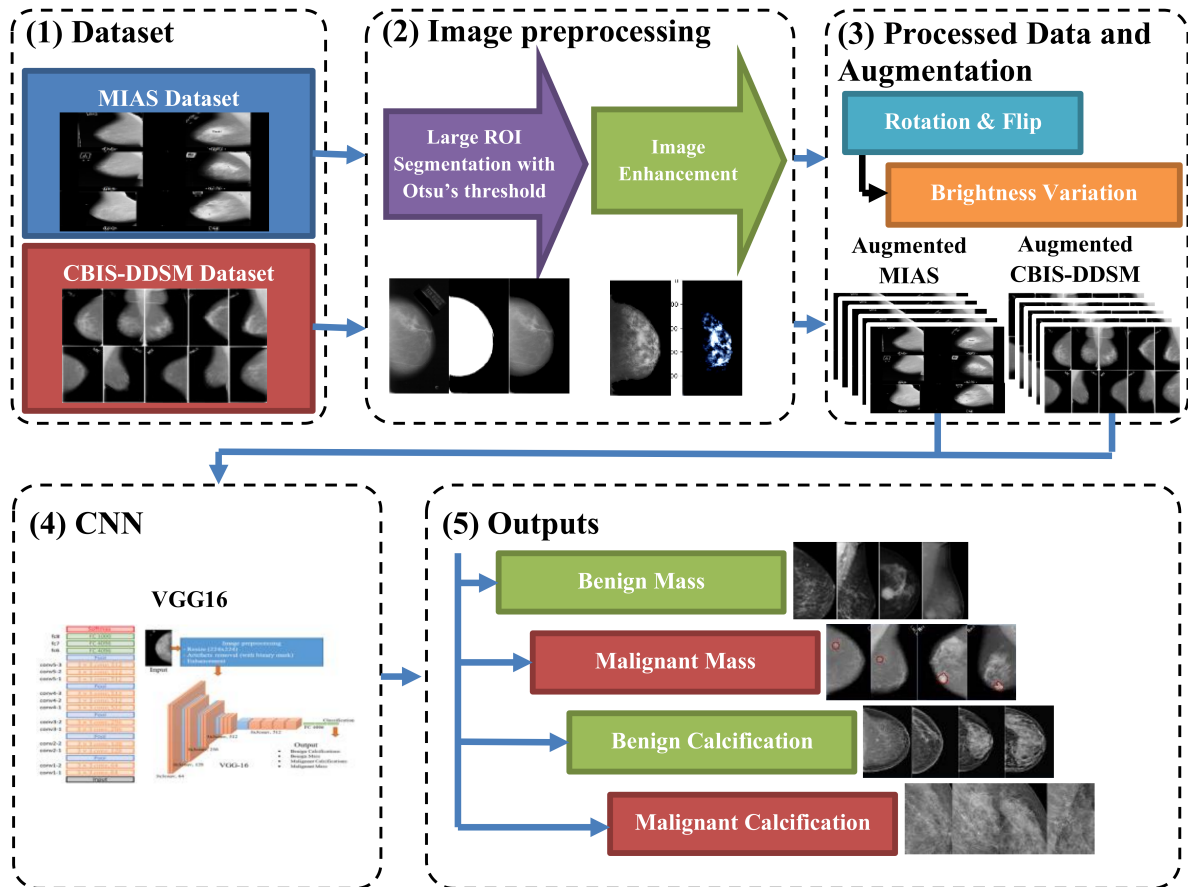
Fig. 1: Block diagram of the classification and process: (1) two datasets with four breast lesion image sets; (2) Images segmented to find the largest ROI and enhanced; (3) Images augmented from the largest ROI images; (4) and (5) the VGG16 with four outputs for breast lesion classification.

used. All image datasets are preprocessed before Otsu segmentation to choose the largest ROI containing features in each lesion image. Therefore, the largest ROI datasets are fed into a VGG16 deep learning network. This network is configured to refine the last two layers and also added appropriate output layers. Finally, the Confusion matrix is applied to evaluate the effectiveness of the proposed classification.

In addition, this article performs three cases of input image sets to the input of the VGG16 network for evaluating accuracy between them. In particular, the image sets are used for three cases such as the image sets without the Otsu segmentation, ones with just the Otsu segmentation, and the sets with the largest ROI using the Otsu segmentation.

2.1. Materials

Mammograms are used in this research from two image sources: CBIS-DDSM and MIAS. Moreover, CBIS-DDSM is an updated version of the DDSM providing easily accessible data and improved ROI segmentation

[25]. The original DDSM is a database of 2,620 scanned film mammography studies, while the CBIS-DDSM has 1,644 images of four types of breast lesions (Benign Mass; Malignant Mass; Benign Calcification; Malignant Calcification) and the image number in each type and their information are shown in Tab. 1. In similarity, the MIAS Mammography dataset includes 1,128 images of 362 women and only 322 images of four types of breast lesions with their information and the image number are described in Tab. 2.

Tab. 1: Information of the CBIS-DDSM dataset.

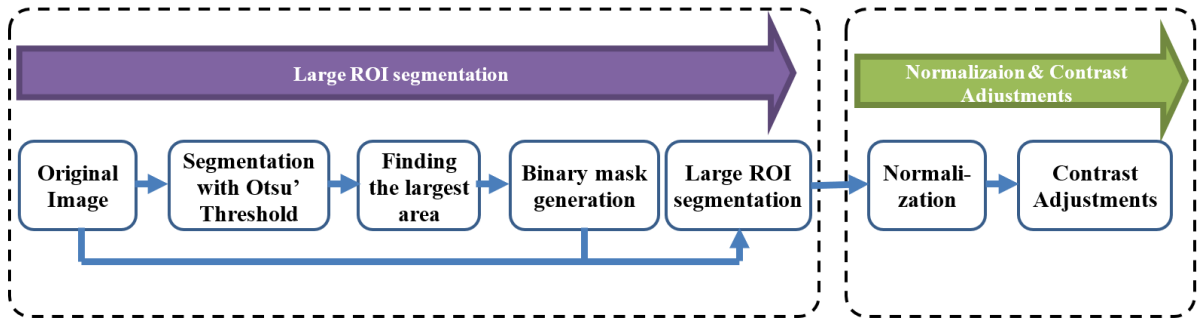| Properties	Value
Total Number of Images	1644
Image Dimension	224 × 224
Color Codec	RGB
Benign Mass	753
Benign Calcification	414
Malignant Calcification	339
Malignant Mass	419

Fig. 2: Description of process for creating the largest ROI with enhancement.

Tab. 2: Information of MIAS dataset.

Properties	Value
Total Number of Images	322
Image Dimension	224×224
Color Codec	RGB
Benign Mass	37
Benign Calcification	12
Malignant Calcification	13
Malignant Mass	19

2.2. Determination of the largest ROI

In this research, all images are processed to determine the largest ROI focused on the breast lesion area before used to train the VGG16 network for classification of breast lesion images. With processing the images to just produce the largest ROI, the accuracy classification is high and also the classification system reduces the model's computation time. In particular, the Otsu algorithm is applied to an original image to determine a threshold for image segmentation to obtain objects in the image. Moreover, the largest ROI with the breast lesion part is determined based on the largest pixel number and the smaller ROI parts are eliminated. This image with the largest ROI is a binary image which is multiplied with the original image to produce the original image with the largest ROI only. Finally, the largest ROI image is normalized and adjusted to produce the wanted ROI image as shown in Fig. 2.

An original mammogram may contain unwanted components which do not contribute for accurate image classification, but also reduce classification accuracy. Furthermore, these components can increase training and recognition times for classifiers. In particular, lines of text or notes on the image or bright areas at the edges of the image can affect the training and classification process. To increase classification accuracy, mammograms need to be processed to remove these areas. Furthermore, the background component of the image will be removed to determine the largest ROI before entering the network for training and classification.

Image segmentation is the process of dividing an image into regions with similar features. In addition, edges and textures are important features which need to obtain for segmentation. In this article, gray mammograms are segmented based on thresholding using the Otsu method. Suppose the threshold $q = 1, 2 \ldots, (L-1)$ is chosen to divide the number of gray levels L into two sets C_0 and C_1 and the Otsu threshold method allows to find q based on the largest variance between the two sets C_0 and C_1 and its equation is expressed as follows:

$$\sigma_B^2 = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \quad (1)$$

in which μ_T is the average gray level of one image.

In order to determine this variance, the weight ω_0 and the average gray level μ_0 of the set C_0, the weight ω_1 and the average gray level μ_1 of the set C_1 are calculated as follows:

$$\omega_0 = \sum\nolimits_{q=0}^{k-1} p_q(r_q) \quad (2)$$

$$\omega_1 = \sum\nolimits_{q=k}^{L-1} p_q(r_q) \quad (3)$$

$$\mu_0 = \sum\nolimits_{q=0}^{k-1} q p_q(r_q) / \omega_0 \quad (4)$$

$$\mu_1 = \sum\nolimits_{q=k}^{L-1} q p_q(r_q) / \omega_1 \quad (5)$$

Therefore, the probability density p_q of the gray level is calculated as follows:

$$p_q(r_q) = \frac{n_q}{M \times N} \quad (6)$$

in which, n_q is the pixel number of the qth gray level and $M \times N$ is the pixel number of one gray image.

The Otsu segmentation method will give many different thresholds, but to choose a suitable threshold, one is based on the variance value of the two sets as in Eq. (1). In particular, the largest variance is the value considered to obtain the best threshold for segmentation. However, a threshold close to the recommended threshold can be chosen for the best segmentation and it is suitable for the typical data.

To effectively remove unwanted components and obtain only the largest ROI, an algorithm to find the largest region based on pixels is applied. In particular, the gray segmented image will be converted to the binary image for determining the largest ROI based on the white pixels, while eliminating smaller regions. Furthermore, to obtain the original image with only the largest ROI, the binary image with the largest ROI will be multiplied with the original image.

For determining the largest ROI, an Otsu threshold is applied to produce objects with the same pixel gray levels and then a largest object, called the largest ROI, will be found. In particular, the algorithm for finding the largest ROI will be determined as follows:

$$\chi_i = \sum_{x=0}^{h_i} \sum_{y=0}^{w_i} O_i(x, y) \tag{7}$$

in which, χ_i is the sum of the white pixels, O_i is the ith ROI with the size of $h_i \times w_i$ $(i = 1, 2, \ldots, n)$.

After completing the largest ROI segmentation, the images are normalized and enhanced using histogram equalization. The enhancement will make all synchronous images with features and brigtness and the classification will be more accurate. The images after enhancement will be fed into the VGG16 network for training and classification.

To normalize images, the method with the values of min and max gray levels is applied in this study. Therefore, the new gray level value is normalized using the following equation:

$$x_{new} = x_{old} \frac{x_{old} - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

in which, x_{new} is the new gray level of pixel. x_{old} is the old gray level of pixel. $x_{\max}$ describes the maximum gray level of image. And $x_{\min}$ describes the minimum gray level of image.

To enhance the image, the histogram method is utilized. In particular, this method will recalculate the gray level values based on the probability density of each gray level, in which it has the relationship between the probability density of the next gray level and that of the previous gray level through sum calculation. Therefore, each gray level is calculated for equlization based on histogram in the image described by the following equation:

$$s_q = (L - 1) \sum_{q=0}^{L=1} p_q(r_q) \tag{9}$$

in which, L is the number of the gray level with $q = 0, 1, 2, \ldots, L - 1$.

Assume that r_q is the gray level value in the original and s_q is the gray level in the equalized image and the probability density p_q of the gray levels of the original image as in Eq. (6). Furthermore, nq is the number of

pixels at the gray level $q = 0, 1, 2, \ldots, L-1$ and $M \times N$ is the number of pixels horizontally and vertically in an image.

2.3. VGG16 network model

VGG16 is a model of a Deep Convolutional Neural Network (DCNN) [19]. This model achieved 92.7% accuracy in the top-5 of testing in the ImageNet dataset [23]. The increased depth of the VGG model can support kernels for extracting more complex features. In this article, VGG16 was used to classify four types of breast lesions (Benign Mass; Malignant Mass; Benign Calcification; Malignant Calcification) of breast X-ray image sets with lesions. In addition, the VGG16 network is fine-tuned so that it can achieve higher accuracy compared to the full network as depicted in Fig. 3. Furthermore, the computer configuration was used to train the VGG16 network includes an Intel Core i9-9980HK processor, NVidia Tesla P4 GPU, 32GB DDR4 Ram Memory, and 1TB SSD for storage. With this configuration, the network performs relatively well and gives more accurate classification results.

In this article, a VGG16 with the total 39 layers and each layer has five blocks. Moreover, each block contains two convolutional layers followed by a Maxpooling layer. Therefore, each convolution layer is multiplied to a two-dimensional array with weights of input data for performing a linear operation. The input layer of the architecture requires the size of the image with $(224 \times 224 \times 3)$ and the input size for the first convolutional layer is similar. In particular, the first block has two convolutional layers with 64 channels with the kernel size (3×3) and the same padding, followed by a (2×2) Maxpooling layer. In similarity, the second block contains two convolutional layers of 128 channels with th 3×3 kernel and they are followed by a Maxpooling layer with the (2×2) size. The last three blocks contain three convolutional layers with the Maxpooling layer. In addition, the channel sizes of the three convolutional layers in blocks 3, 4, and 5 are 256, 512, and 512, respectively and all of them are the (3×3) kernel size. Thus, the original input image is scaled down to half its size within each Maxpooling layer. After stacking the convolutional and Maxpooling layers, the feature output from the final Maxpooling layer has a size of $(7 \times 7 \times 512)$. A flattening layer is added to create a $(1 \times 25,088)$ features vector. Moreover, a dense layer is added to produce four channels for the layers. In this deep learning network, there is a Softmax activation function at the end for normalizing the classification vector obtained from FC. Finally, the VGG16 model is pre-trained on the ImageNet database.

The convolutional layers in the VGG16 model use the ReLU loss function, which is equal to 0 for $x < 0$
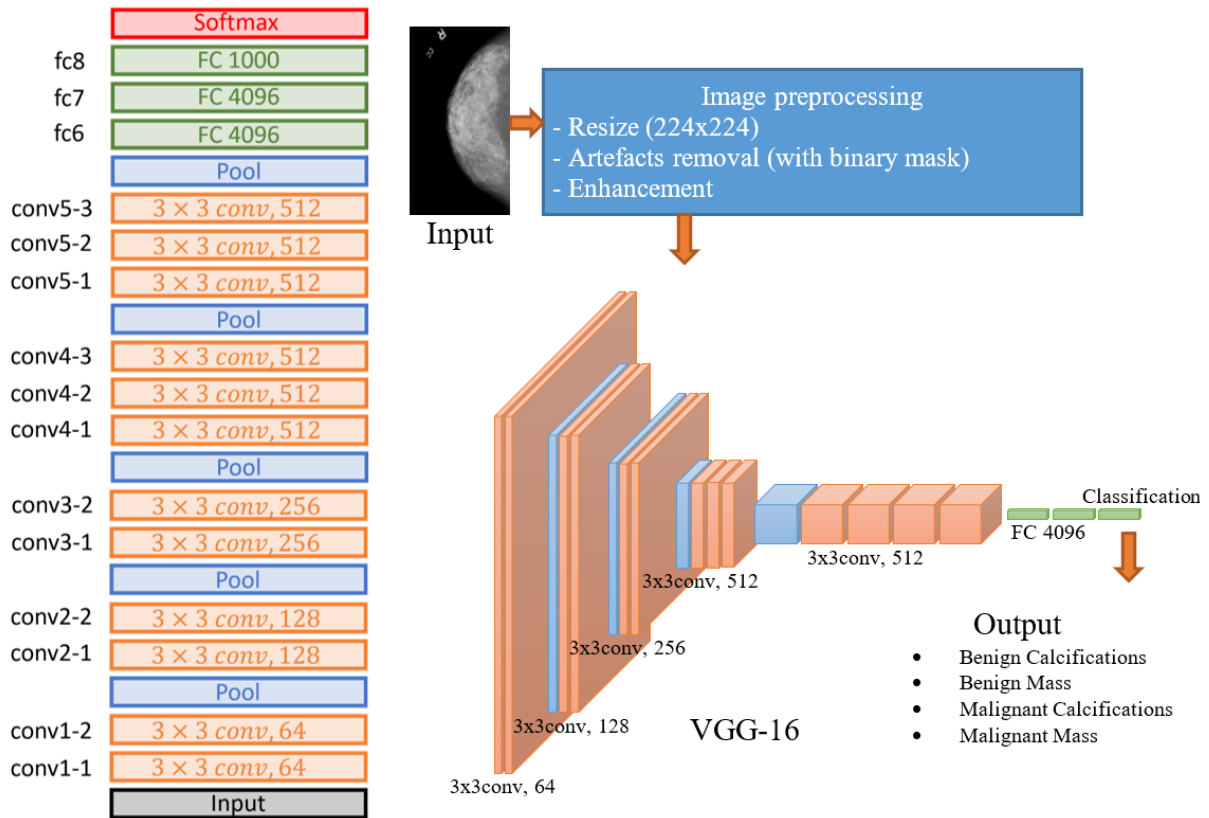
Fig. 3: Architecture of the fine-tuned VGG16 network used in this article.
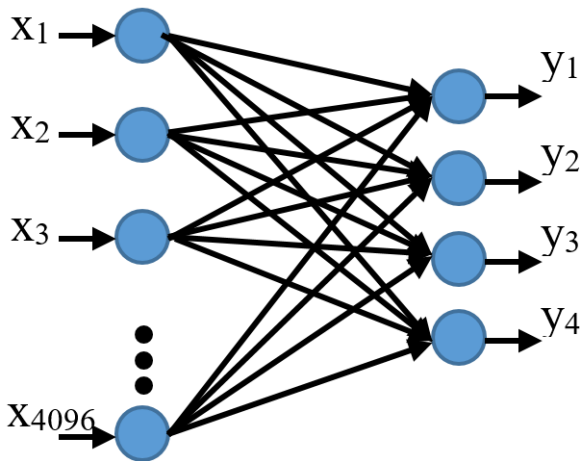

Fig. 4: Description of the last FC layer.

and equal to x for all $x \geq 0$ and it is described as follows:

$$ReLU(x) = \max\{x, 0\} = \begin{cases} x, & if \quad x \geq 0 \\ 0, & if \quad x < 0 \end{cases} \tag{10}$$

In this article, the VGG16 model has two phases: training and fine-tuning. During training, the last two layers of 16 layers in the VGG16 model are fine-tuned, and weights in the remaining layers are kept from the prototype trained on ImageNet. In the output layer of the model, the dense and FC layers of the network model are updated by the new features of the images during the fine-tuning process. Moreover, the final FC layer will be modified for classifying four types of breast lesion images. In particular, to achieve the purpose of classifying 4 breast lesion image layers including Benign Mass, Benign Calcification, Malignant Calcification, and Malignant Mass, the last FC layer is adjusted to have 4 outputs as shown in Fig. 4.

Therefore, the pre-trained weights of layers are transferred by using transfer learning and then the model is retrained using the image datasets. In this model, the 'Softmax' activation function is used to calculate the probability for each class and to create the prediction results. Therefore, Softmax is expressed by the following equation:

$$Z_j = \frac{e^{x_j}}{\sum_{i=1}^{n} e^{x_i}}; \quad j = 1, 2, \ldots, n \tag{11}$$

To train the model in this research, the maximum number of epochs is set to 350 and the batch size to 16. The Adam optimization algorithm is applied with a learning rate of 0.001. In addition, the Adam algorithm is an enhanced version of Stochastic Gradient Descent, which is a combination of the Momentum and RMSP algorithms and is described by the following mathe-

matical equation:

$$W_{T+1} = w_t - \propto m_t \qquad (12)$$

in which,

$$m_t = \beta m_{t-1} + (1 - \beta) \left[\frac{\delta L}{\delta w_t} \right] \qquad (13)$$

and, W_{T+1} is weights at time t+1. w_t is weights at time t. $\propto$ is learning rate. m_t is aggregate of gradients at time t. δL is derivative of Loss Function. δw_t is derivative of weights at time t. And β is moving average parameter.

3.   Results and discussions

This article proposes an algorithm for the classification of breast lesions, in which image sets are processed to extract the largest ROI before inserting them into VGG16. In this algorithm, the results of image processing and classification using VGG16 are demonstrated to illustrate the effectiveness of this algorithm. Finally, the comparison and evaluation of the results are performed to illustrate that the proposed method is effective.

3.1.   Determination of largest ROI

With a breast lesion image, it is very necessary to accurately determine whether it is breast cancer because this can help doctors have more information to decide breast cancer. In this study, image datasets with breast lesions were segmented to extract the largest ROI parts with many features for better lesion classification using deep learning networks. In particular, to remove unwanted components in the images for creating the largest ROI, the image segmentation method using Otsu thresholding was utilized as shown in Fig. 5.

Segmentation with different threshold values can produce segmented images with different ROI parts. Therefore, simulation results in this research showed that adjusting image brightness before segmentation significantly improved the regions segmented with the Otsu algorithm. With a chosen Otsu threshold, extracting the largest ROI is obtained. In particular, the breast lesion image only retains the largest ROI containing many possible features and most of the unwanted components in the original image could be eliminated as shown in Fig. 5.

In Fig. 5, the images of (a4), (b4), (c4), (d4) just have the largest ROI and smaller ROI parts are eliminated compared to the images (a1), (b1), (c1), (d1). Moreover, the images of (a4), (b4), (c4), and (d4) after

eliminating can be seen more clearly lesions, called features, around ROI. The processed images make it easy to see the difference between the four types of breast lesions (BM, MM, BC, MC).

Fig. 6 shows the ROI image normalized and adjusted for contrast using the Histogram equalization algorithm. In particular, the images after correction were clearer with features focusing on the largest ROI parts and this can help the cancer classification system achieve better performance. In addition, the enhanced images of (a3), (b3), (c3), and (d3) produced the largest ROI parts nearly synchronous, clearer compared to the images of (a1), (b1), (c1), (d1). This is important due to the ability to classify with higher accuracy using VGG16 compared to the images without enhancement.

3.2.   Determination of largest ROI

In this article, two main datasets of CBIS-DDSM and MIAS are used for the classification of four breast lesions using the deep learning network. However, they are not enough for training and classifying to produce the desired accuracy. Thus, an augmentation of image sets is necessary for using the deep learning network. The rotation and flip images in these two datasets and also the brightness adjustment applied were performed as shown in Tab. 3 and Tab. 4.

All image sets after the largest ROI extraction and enhancement are used for training the network. In this article, the VGG16 is applied for the classification of four types of breast lesions. To evaluate the performance of training, the datasets were divided into the training set (80%) and the testing set (20%), respectively. In particular, after data augmentation, 15,118 X-ray images of breast lesions were extracted with the largest ROI divided into 12,093 for training (80%) and 3,025 images for testing (20%) as depicted in Fig. 7. In particular, the average class for training is around 3,000 images and similarly it is 750 images for testing in this research.

3.3.   Results for classification of breast lesion images

Fig. 8 shows the training curves for the classification model, including the loss curve and the accuracy curve. It is obvious that the training curve smoothly converges from the first stage to the last stage without collisions. Furthermore, the distance between the validation accuracy curve (red) and the training one (blue) shows that they nearly have no overfitting during the training process. In similarity to the training curve, the loss curve shown in Fig. 8 converges steadily. With these
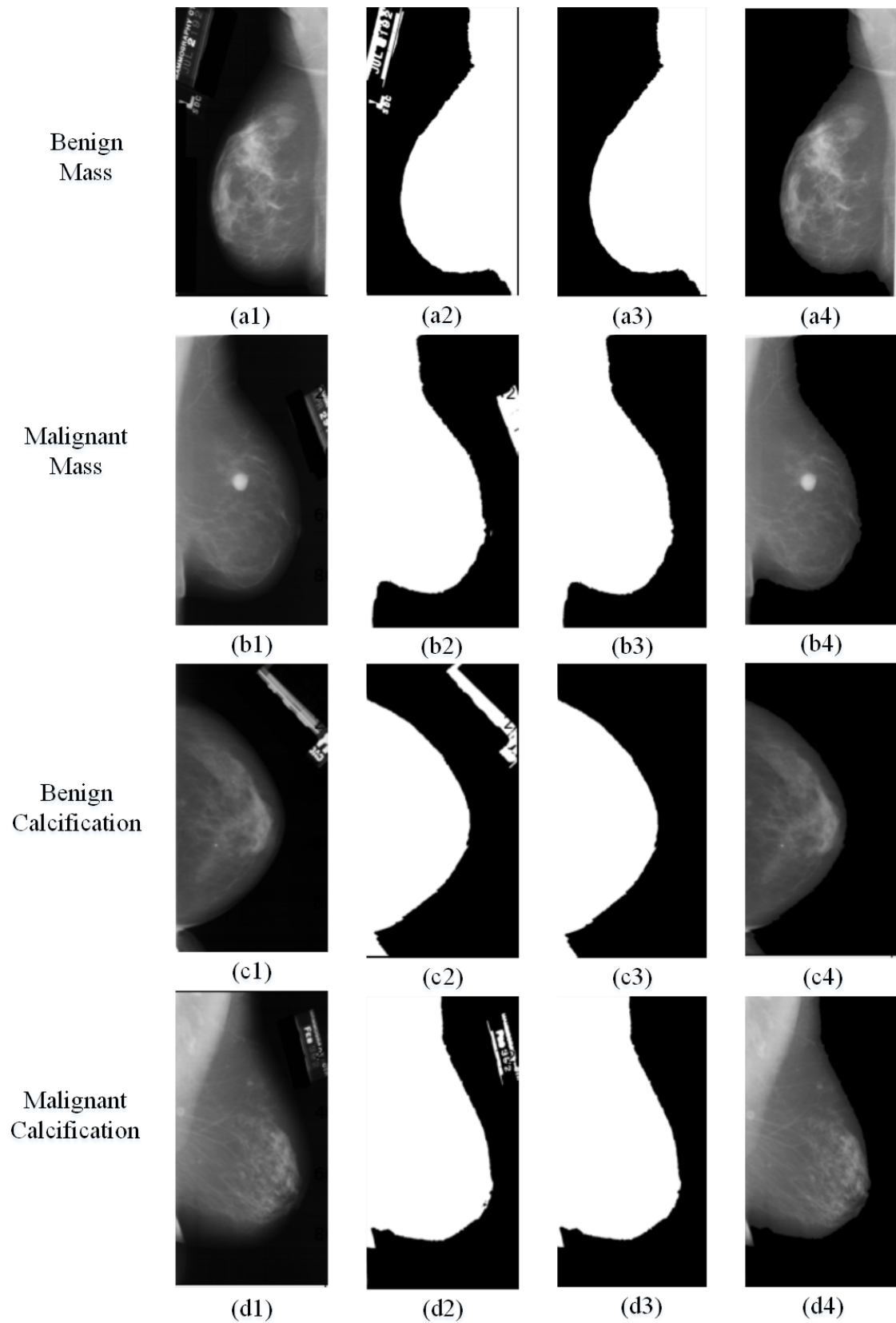
Fig. 5: Description of extracting the largest contour-based ROI and eliminating unwanted components: (a1-a4) Original images of four breast lesions; (b1-b4) Segmented images with the Otsu threshold method (c1-c4) Binary images with the largest ROI eliminated unwanted components; (d1-d4) Original image with the largest ROI without unwanted components.
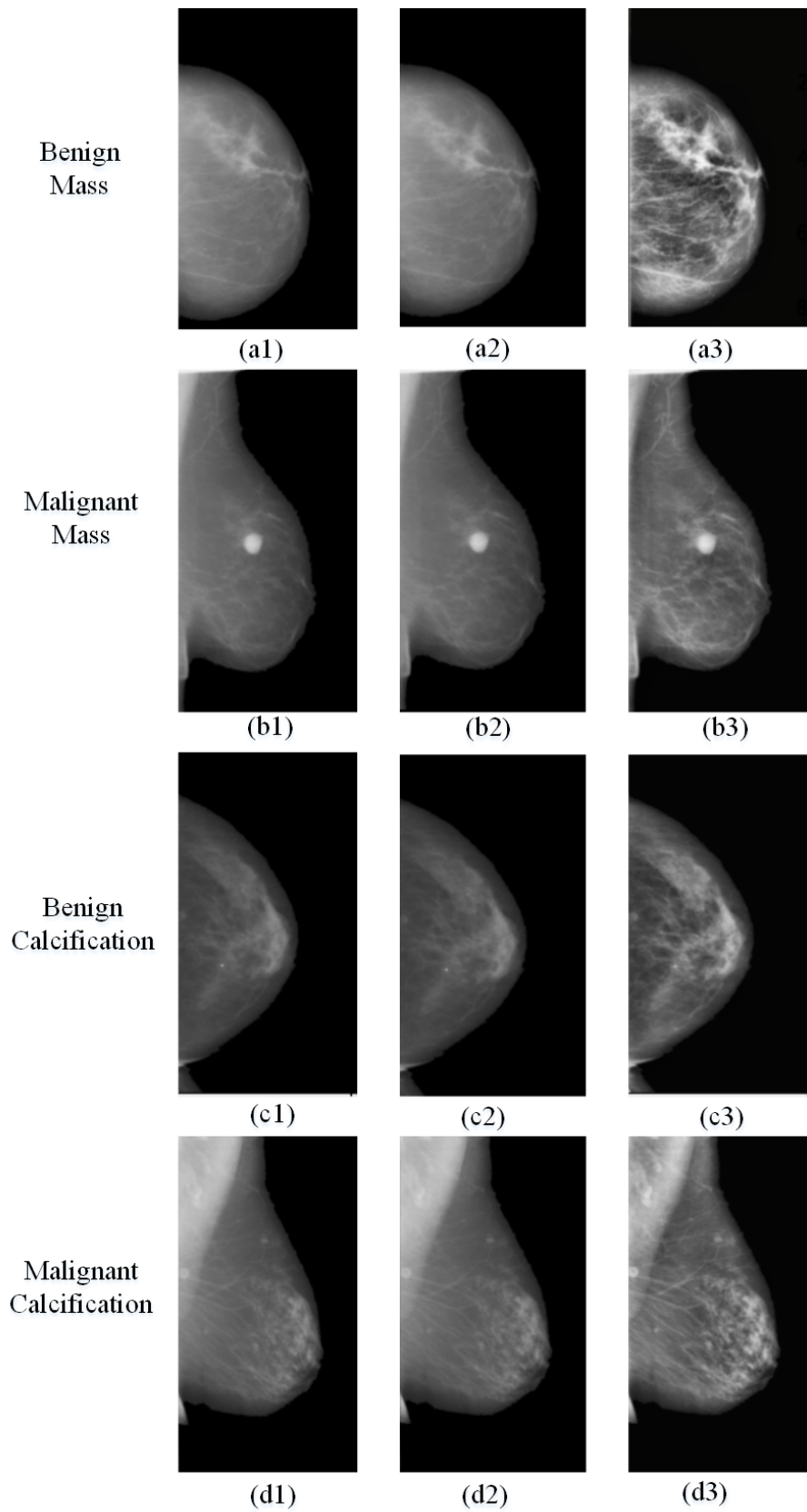
Fig. 6: Description of results for normalizing and the largest ROI image enhancement using histogram equalization: (a1-a3) Original images; (b1-b3) Normalized images; (c) Image with the histogram equalization.

Tab. 3: Description of four types of the augmented image sets of CBIS-DDSM.

	Original images	Augmented images	The augmented sets divided for training and testing	
			Training	Testing
Image No.	1644	12542	10033	2509
Benign Mass	753	3165	2532	633
Benign Calcification	414	2897	2318	579
Malignant Calcification	339	2985	2387	598
Malignant Mass	419	3495	2796	699

Tab. 4: Description of four types of the augmented image sets of CBIS-DDSM.

	Original images	Augmented images	The augmented sets divided for training and testing	
			Training	Testing
Image No.	322	2576	2060	516
Benign Mass	146	1168	934	234
Benign Calcification	48	384	307	77
Malignant Calcification	52	416	333	83
Malignant Mass	419	3495	2796	699
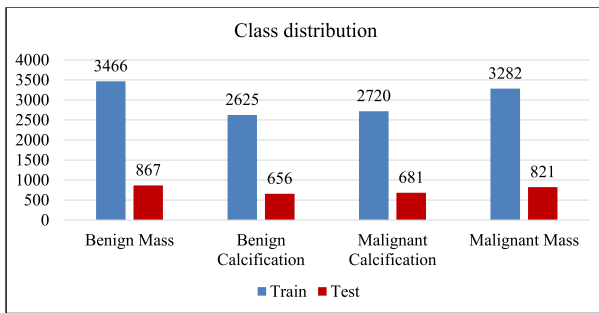
Fig. 7: Distribution of four types of datasets for training (blue) and validation (red).

(a)

(b)

Fig. 8: Description of training process with over 350 epochs for Adam with learning rate 0.0001: (a) accuracy curve for training and validation; (b) loss curve for training and validation.

results of the training and loss curves without overfitting or underfitting, we can see that the proposed model for the classification of four breast lesions is effective. In addition, the accuracy rate over the epochs curve shows that the accuracy increases when the number of iterations increases, in which the blue line is for the training set, and the red line is for the test set.

Fig. 9(a) and Fig. 9(b) show the results of using the confusion matrix for the VGG16 classification model with the highest accuracy. In particular, the row values represent the actual labels of the image sets, BC, BM, MC, and MM, the column values describe the predicted labels of four types of these image sets, and the diagonal lines express the True Positive (TP) values. In addition, the confusion matrix shows that no bias exists for any layer and also, they predict all similar layers. Specifically, with the CBIS-DDSM dataset, the model can classify MM and BM lesions with the highest accuracies of 95.71% and 95.58% respectively. Meanwhile, with the MIAS dataset, BM and BC lesions yield the highest classification accuracies of 93.16% and 92.21%
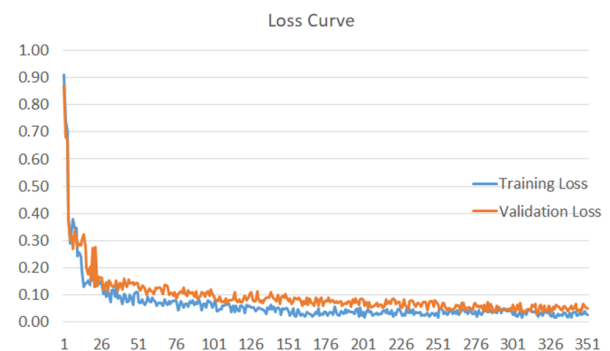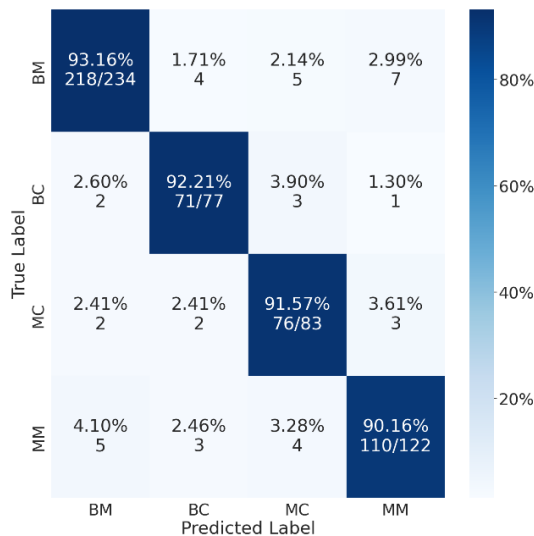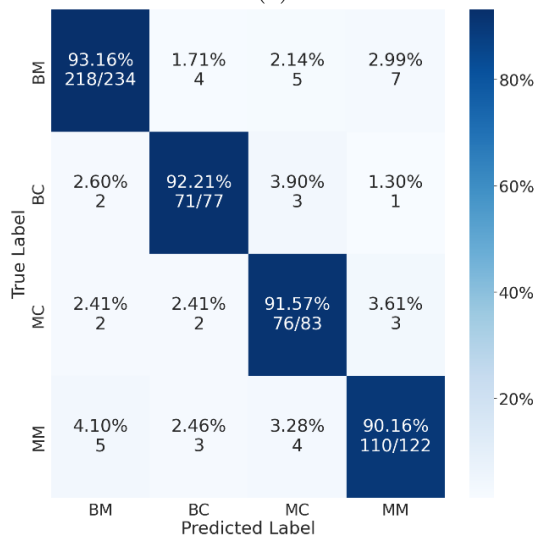
(a)

(b)

Fig. 9: Description of training process with over 350 epochs for Adam with learning rate 0.0001: (a) accuracy curve for training and validation; (b) loss curve for training and validation.

respectively. The average accuracy over the CBIS-DDSM dataset is 95.23%, which is higher than the accuracy achieved when classifying on the MIAS dataset at 91.78%. Notably from Fig. 9, the misclassification rates from BM to MM are high for both test datasets at 1.74% and 2.99% respectively. Additionally, MM is also misclassified as BM with rates of 2.00% and 4.10% on the CBIS-DDSM and MIAS datasets. These classification results demonstrate the effectiveness and potential of the proposed model.

In Fig. 10, the accuracy curves for testing and validating show that the model achieves high accuracy with a low false positive rate for the data sets. In particular, the yellow curve is the result of the training set and the blue curve is the result of the testing
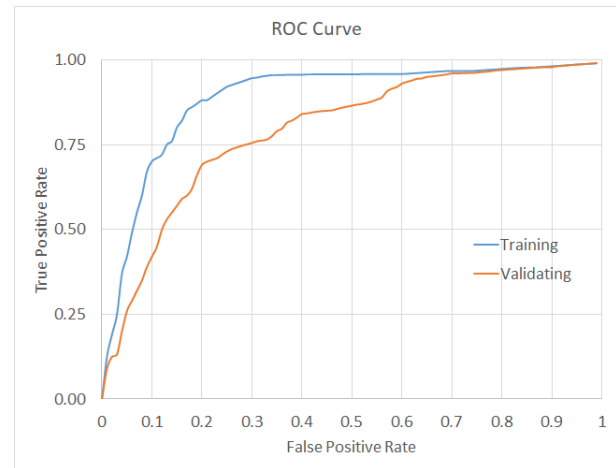

Fig. 10: Accuracy rate curves of training and testing using the proposed model.

set. Thus, the receiving operating curve (ROC) and the area under the curve (AUC) taken from the ROC curve show that the proposed model along with the proposed method is effective for distinguishing different classes. In particular, an AUC value close to 1 indicates that this model can detect most classes, where the ROC curve almost touches the peak of the y-axis, which means the false positive rate is close to 0. The true positive rate is close to 1, which proves that the model's effectiveness is very high.

Extracting the largest ROI and enhancing images has improved the classification performance of the proposed model, particularly, the proposed model in this article achieved 95% accuracy. For the high performance, the image sets were extracted with the largest ROIs to capture many features of breast lesions and then normalized and enhanced. For the effective evaluation of the proposed model, this article shows three cases of data sets as shown in Tab. 5. In particular, the image set is processed following the proposed with the largest ROI producing the highest performance of 95%, while the image set without processing is 82.1% of accuracy smaller than that of just Otsu segmentation. This result proves that the proposed model can classify four types of breast lesions to make it easier to identify breast cancer from mammograms with the highest accuracy. One outstanding point is that applying the large ROI region segmentation based on the Otsu algorithm allows for quick segmentation without using pre-segmented data for training. Furthermore, compared to other deep learning networks, such as U-Net, for segmented data, this method is simpler, and less time-consuming, but still increases the average accuracy of the system.

To evaluate the reliability of the results, with a limited number of data sets, we use the cross-validation method with 5-fold on the same data set. Fig. 11 illus-

Tab. 5: Information of the CBIS-DDSM dataset.

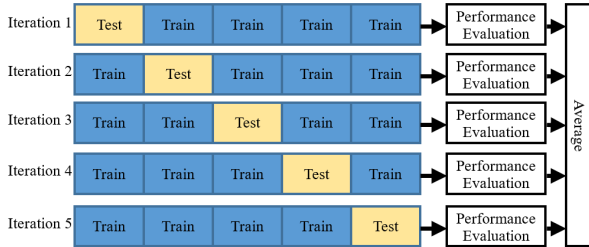| Methods	Average Accuracy
No image processing	82.1%
Otsu segmentation	89.2%
Otsu segmentation with the largest ROI	95%

Fig. 11: 5-fold cross validation.

trates how to divide the data set for use in this evaluation process. In particular, the data set is divided into 5 segments, each iteration uses 1 of the 5 segments in turn as the validation set, while the remaining 4 segments are used for training. The results of 5 iterations are used to compare with each other for evaluating the stability of the model.

With the results of the 5 iterations as shown in Tab. 6, there is no big difference between the iterations and these results are reliable. Moreover, from these results, Confidence Intervals (CIs) are calculated and shown in Fig. 12. In particular, the CI of Accuracy is quite narrow, showing the stability of accuracy. The confidence intervals of Sensitivity, Precision, and F1 score are larger but still do not have too large deviations between iterations. These results show the stability of the model.

Tab. 7 shows the highest accuracy achieved by the proposed model compared to previous works with similar methods, datasets, or models. More than half of
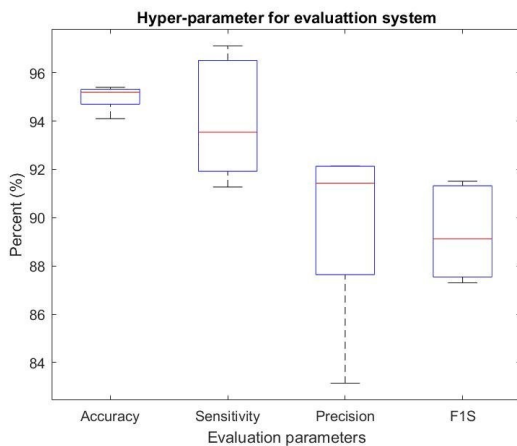

Fig. 12: Mean value and confidence interval of Accuracy, Sensitivity, Precision, and F1s.

the previous studies in Tab. 7 used the VGG16 and similar datasets compared to the study in this article. In [27], the transfer learning method with the MVGG model was proposed for image classification based on the VGG16 model and the result was 94.3% with 15 epochs using the DDSM database.

In a study by Shallu and Mehra [28] using the pre-trained and fine-tuned VGG16, its performance achieved the highest accuracy of 92.60%. With VGG16 and VGG19s using the same dataset, Hameed et al. [29] had the result with the obtained accuracy of 94.71%. Similarly, in the research of Li et al. [30], his article applied several pre-trained models and the highest accuracy was 92.78% with VGGNet. Our fine-tuned VGG16 network with the FC layer achieved 95% accuracy in breast lesion classification using the CBIS-DDSM dataset after Adam optimization. In general, with the above statistics, the models using VGG16 achieved the classification with a very high accuracy.

Two image sets (CBIS-DDSM and MIAS) are very popular for research related to breast lesions. Although different algorithms of classification, the accuracy is very high, around 92%. In our research, these two image sets with four types of breast lesions were used, in which there were the augmented image sets to create the large dataset. In addition, these image sets were processed to extract the largest ROI, which was used for the fine-tuned VGG16 and produced the results with a high accuracy of 94% to 95%. This means that the obtained results demonstrate to illustrate the effectiveness of the proposed algorithm.

4.   Conclusion

This article has proposed a classification model using four breast lesion image sets, including Benign Calcifications (BC), Benign Mass (BM), Malignant Calcifications (MC), and Malignant Mass (MM) from two image sets of CBIS-DDSM and MIAS. All images were extracted with the largest ROI and then normalized and enhanced before applying to train the VGG16 network. Extracting the largest ROI gave a higher classification efficiency of 95% compared to images without extracting the largest ROI. Furthermore, to increase accuracy, these image sets were processed to augment the number of images by rotating and flipping images. In addition, in VGG16, the last two layers are fine-tuned during training and classification to contribute to the classification performance. Therefore, the results demonstrate to illustrate that removing unwanted components and extracting only the largest ROI significantly contributes to improving the proposed model performance. The proposed model with VGG16 shows high performance for classifying four types of breast lesions of CBIS-DDSM and MIAS sets. With the clas-

Tab. 6: Accuracy, Sensitivity, Precision, and F1s over 5 iterations.

Iteration	Accuracy	Sensitivity	Precision	F1s
1	95.30%	0.9631	0.9214	0.8763
2	95.40%	0.9712	0.9213	0.8912
3	94.10%	0.9354	0.8914	0.9126
4	94.90%	0.9127	0.9143	0.8730
5	95.20%	0.9214	0.8314	0.9151
Average	**94.98%**	**0.9408**	**0.8960**	**0.8936**

Tab. 7: Comparison between this research with previous work.

Paper	Model	Dataset	Epoch	Accuracy
Khampari et al. (2021) [27]	Hybrid MVGG16 ImageNet	Digital Database with DDSM having 2620 images.	15	94.3%
Shallu and Mehra, (2018) [28]	VGG16 + LR VGG19 + LR ResNet50 + LR	2042 full-field digital images provided by the 1st Hospital of Shanxi Medical University and 30,630 augmented images	-	92.60% 90.40% 79.40%
Hameed et al. (2020) [29]	Full trained VGG16 + VGG19 Fine-tuned VGG16 + VGG19	544 images provided by Colsanitas Colombia University	200	93.53% 95.29%
Li et al. (2019) [30]	AlexNet VGGNet GoogLeNet DenseNet DenseNet-II	2042 full-field from the First Hospital of Shanxi Medical University and augmentation of 30,630 images	-	92.70% 92.78% 93.54% 93.87% 94.55%
Al-antari et al. (2018) [30]	Fully Integrated CAD	410 Full-field Digital images and 896 augmented images	100	95.64%
Nasir Khan et al. (2019) [31]	MVFF CADx	3568 CBIS-DDSM images and 322 MIAS mammograms	100	93.73%
Zahra Jafari and Ebrahim Karami (2023) [32]	Concat. Model with the NN classifier	54,713 images of RSNA dataset	-	96%
This study	Largest ROI and fine-tuned VGG16	12,093 images of CBIS-DDSM including augmentation and 3,025 augmented images of MIAS	350	95%

sification performance up to 95%, the proposed model may be developed for classification with many more different breast lesion image sets to support in breast cancer diagnosis. In addition, with the results of this research, the proposed method can be developed to be able to apply to other deep learning networks and different datasets.

Acknowledgment

Author Contributions

Thanh-Tam NGUYEN developed the theoretical formalism, performed the analytic calculations, and performed the numerical simulations to produce results.

Thanh-Hai NGUYEN supervised and edited the whole article. Ba-Viet Ngo contributed to the draft version of the manuscript and evaluated the results. Thanh-Nghia NGUYEN contributed to the acquisition of data and augmented and analyzed them.

References

[1] ABAAN, O. D., W. E. CRISS. Gene Therapy in Human Breast Cancer. *Turkish Journal of Medical Sciences*. 2002, vol. 32, pp. 283–291. https://journals.tubitak.gov.tr/medical/vol32/iss4/1/.

[2] FACKENTHAL, J. D., O. I. OLOPADE. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer*. 2007, vol. 7, no. 12, pp. 937–48. DOI: 10.1038/nrc2054.

[3] FACKENTHAL, J. D., O. I. OLOPADE. Telemammography: A Novel Approach for Early Detection of Breast Cancer Through Wavelets Based Image Processing and Machine Learning Techniques. *Advances in Soft Computing and Machine Learning in Image Processing*. 2018, pp. 149—183. DOI: 10.1007/978-3-319-63754-9_8.

[4] ARNOLD, M., et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*. 2022, vol. 66, pp. 15–23. DOI: 10.1016/j.breast.2022.08.010.

[5] ARNOLD, M., et al. Comparison of wire-guided localization and radio-guided occult lesionlocalization in preoperative localization of nonpalpable breast lesions. *Turk J Med Sci*. 2016, vol. 46, no. 6, pp. 1829–1837. DOI: 10.3906/sag-1507-162.

[6] YAN, J., Z. LIU, S. DU, J. LI, L. MA, L. LI. Diagnosis and Treatment of Breast Cancer in the Precision Medicine Era. *Methods Mol Biol*. 2020, vol. 2204, pp. 53–61. DOI: 10.1007/978-1-0716-0904-0_5.

[7] ZHANG, J., J. HU. Image Segmentation Based on 2D Otsu Method with Histogram Analysis. *2008 International Conference on Computer Science and Software Engineering, Wuhan, China*. 2008, pp. 105–108. DOI: 10.1109/CSSE.2008.206.

[8] MUSTRA, M., M. GRGIC, R. M. RANGAYYAN. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. *Medical & Biological Engineering & Computing*. 2016, vol. 54, pp. 1003–1024. DOI: 10.1007/s11517-015-1411-7.

[9] SHI, P., J. ZHONG, A. RAMPUN, H. WANG. A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms. *Computers in Biology and Medicine*. 2018, vol. 96, pp. 178–188. DOI: 10.1016/j.compbiomed.2018.03.011.

[10] KASHYAP, K. L., M. K. BAJPAI, P. KHANNA, G. GIAKOS. Mesh-free based variational level set evolution for breast region segmentation and abnormality detection using mammograms. *International Journal for Numerical Methods in Biomedical Engineering*. 2018, vol. 34, iss. 1. DOI: 10.1002/cnm.2907.

[11] CELIK, T., T. TJAHJADI. Automatic Image Equalization and Contrast Enhancement Using Gaussian Mixture Modeling. *IEEE Transactions on Image Processing*. 2012, vol. 21, pp. 145–156. DOI: 10.1109/TIP.2011.2162419.

[12] YAO, L., S. MUHAMMAD. A novel technique for analysing histogram equalized medical images using superpixels. *Computer Assisted Surgery*. 2019, vol. 24, pp. 53–61. DOI: 10.1080/24699322.2018.1560100.

[13] NGUYEN, T. H., N. T. NGUYEN, M. H. NGUYEN, S. LIVATINO. Wavelet-Based Kernel Construction for Heart Disease Classification. *Advances in Electrical and Electronic Engineering*. 2019, vol. 17, no. 3, pp. 306-319. DOI: 10.15598/aeee.v17i3.3270.

[14] NGUYEN, M.-H., V.-H. TRAN, T.-H. NGUYEN, T.-N. NGUYEN. A Deep Learning Framework for Inter-Patient ECG Classification. *IJCSNS International Journal of Computer Science and Network Security*. 2019, vol. 19, pp. 74–84. DOI: 10.15598/aeee.v17i3.3270.

[15] WEI, L.-S., Q. GAN, T. JI. Skin Disease Recognition Method Based on Image Color and Texture Features. *Comput Math Methods Med*. 2018, vol. 2018. DOI: 10.1155/2018/8145713.

[16] PEREIRA, S., A. PINTO, V. ALVES, C. A. SILVA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*. 2016, vol. 35, no. 5, pp. 1240–1251. DOI: 10.1109/TMI.2016.2538465.

[17] HUSAINI, M. A. S. A., M. H. HABAEBI, S. A. HAMEED, M. R. ISLAM, T. S. GUNAWAN. A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks. *IEEE Access*. 2020, vol. 8, pp. 208922–208937. DOI: 10.1109/ACCESS.2020.3038817.

[18] KESKIN, N. K., et al. Detection of the differences in the apparent diffusion coefficient values in different histopathological types of malignant breast lesions and comparison of cellular region/stroma ratio and histopathological results. *Turkish Journal Of Medical Sciences*. 2018, vol. 48, pp. 817–825. DOI: 10.3906/sag-1801-89.

[19] YÜCEL, A., B. DEĞİRMENCİ, M. ACAR, H. ELLİDOKUZ, R. ALBAYRAK, A. HAKTANIR. Knowledge About Breast Cancer and Mammography in Breast Cancer Screening Among Women Awaiting Mammography. *Turkish Journal of Medical Sciences*. 2005, vol. 35, pp. 35–-42. https://journals.tubitak.gov.tr/medical/vol35/iss1/6/.

[20] DOMINGUES, I., P. H. ABREU, J. SANTOS. Bi-Rads Classification of Breast Cancer: A New Pre-Processing Pipeline for Deep Models Training. *2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece*. 2018, pp. 1378–1382. DOI: 10.1109/ICIP.2018.8451510.

[21] AMALA, R. K., H. -P. CHAN, L. HADJI-ISKI, M. A. HELVIE, C. D. RICHTER, K. H. CHA. Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets. *IEEE Transactions on Medical Imaging*. 2019, vol. 38, no. 3, pp. 686-696. DOI: 10.1109/TMI.2018.2870343.

[22] TING, F. F., Y. J. TAN, K. S. SIM. Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications*. 2019, vol. 120, pp. 103-–115. DOI: 10.1016/j.eswa.2018.11.008.

[23] SIMONYAN, K., A. ZISSERMAN. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*. 2014, pp. arXiv:1409.1556. DOI: 10.48550/arXiv.1409.1556.

[24] TAN, M., Q. V. LE. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *36th International Conference on Machine Learning*. 2019, pp. 6105–6114.

[25] LEE, R., et al. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data*. 2017, vol. 4. DOI: 10.1038/sdata.2017.177.

[26] SUCKLING, J., et al. Mammographic Image Analysis Society (MIAS) database v1.21. *Apollo - University of Cambridge Repository*. 2015. DOI: 10.17863/CAM.105113.

[27] KHAMPARIA, A., et al. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional Systems and Signal Processing*. 2021, vol. 32, pp. 747–765. DOI: 10.1007/s11045-020-00756-7.

[28] SHALLU, R. MEHRA. Breast cancer histology images classification: Training from scratch or transfer learning?. *ICT Express*. 2018, vol. 4, iss. 4, pp. 247–254. DOI: 10.1016/j.icte.2018.10.007.

[29] HAMEED, Z., S. ZAHIA, B. GARCIA-ZAPIRAIN, J. J. AGUIRRE, A. M. VANEGAS. Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models. *Sensors*. 2020, vol. 20, no. 16. DOI: 10.3390/s20164373.

[30] LI, H., S. ZHUANG, D.-A. LI, J. ZHAO, Y. MA. Benign and malignant classification of mammogram images based on deep learning. *Biomedical Signal Processing and Control*. 2019, vol. 51, pp. 347–354. DOI: 10.1016/j.bspc.2019.02.017.

[31] KHAN, H. N., A. R. SHAHID, B. RAZA, A. H. DAR, H. ALQUHAYZ. Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network. *IEEE Access*. 2019, vol. 7, pp. 165724–165733. DOI: 10.1109/ACCESS.2019.2953318.

[32] JAFARI, Z., E. KARAMI. Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. *Information*. 2023, vol. 14, no. 7. DOI: 10.3390/info14070410.