Double-Dilation Non-Pooling Convolutional Neural Network for Breast Mass Mammogram Image Classification

Mei-Ling Huang* Ting-Yu Lin*

ABSTRACT

Background and Objective: Many researchers have examined breast mammogram images by using deep learning methods. Some scholars have used the existing convolutional neural network (CNN) model for classification, whereas other scholars have improved on the CNN model to obtain superior image classification performance.

Method: This study used breast mass mammograms (BMMs) from the INbreast database, MIAS database, and Digital Database for Screening Mammography to establish a new BMM database for testing two developed breast mass classification models that can extract diverse features: (1) a double-dilation non-pooling CNN (DDNPNet) and (2) AlexNet II. We compared the results of these two classification models with those of three other CNN models, namely AlexNet, DenseNet, and ShuffleNet.

Results: For the established BMM database, DenseNet exhibited the highest evaluation indices. The accuracy, specificity, sensitivity, F1 score, and training time of DenseNet were 98.59%, 98.21%, 99.08%, 98.43%, and 6 h 39 min 21 s, respectively. Although DenseNet provided satisfactory results, it required a long training time. For the established BMM dataset, the accuracy, specificity, sensitivity, F1 score, and training time of the DDNPNet model were 95.41%, 95.86%, 94.83%, 94.81%, and 26 min 23 s, respectively. Thus, the DDNPNet model provided similar classification results to the other models but in considerably less time.

Conclusion: The breast mass classification models proposed in this study can assist physicians in the analysis of BMMs.

Keywords: Breast mass mammogram database, Convolutional neural network, DDNPNet, Mass classification

INTRODUCTION

Due to the fast pace of modern life and long-term stress, an individual may be affected by various disease-related issues. Breast cancer is a major disease that women must pay attention to. According to statistics from the American Cancer Society[1], approximately 268,600 women received a diagnosis of breast cancer in the United States in 2019. Moreover, in the same year, approximately 41,760 female patients with breast cancer were on the verge of death. Most patients with breast cancer are diagnosed in the late stage of the disease. According to Siegel et al.[2-7], the number of new diagnoses and mortalities of breast cancer in the United States have increased every year from 2015 to 2020. Thus, the control of breast cancer has become an urgent public issue. The incidence of breast cancer has gradually increased in not only developed countries but also in developing countries. The best method for controlling breast cancer is early detection and diagnosis. The method for early diagnosis of breast cancer involves conducting a biopsy; however, a biopsy imposes a large burden on patients because patients must endure pain when they undergo biopsies[8]. With the advancement of technology, some new methods are now available to facilitate the diagnosis of breast cancer by physicians. As indicated by the American College of Radiology, tools such as mammograms, breast ultrasounds, and breast magnetic resonance imaging (MRI) scans, can be used to diagnose breast cancer. Ardakani et al.[9] stated that mammograms are a common tool to diagnose breast cancer. Moreover, computer-aided diagnosis (CAD) can help physicians recognize from a breast mass image whether a breast mass is benign or malignant. This strategy helps patients avoid undergoing a biopsy and reduces their pain. Moreover, it can improve the accuracy with which physicians can recognize benign or malignant masses in mammogram images. The appearances and shapes of benign and malignant masses of breast cancer are different. Benign tumors have a relatively smooth shape, whereas malignant tumors are irregularly shaped[10].

Many studies have been conducted on mammograms, and mammograms have become important tools for determining whether a patient has breast cancer. In recent years, rapid developments have occurred in deep learning (DL), which is useful in cancer-related research. Many studies on mammogram image recognition have used DL because DL does not require complex feature extraction methods. The purpose of DL is to help physicians perform aided diagnosis and improve the quality of medical care. The following text presents the improvements achieved by studies in the CAD technology used for breast cancer.

In many studies, mammograms have been preprocessed before being input into a convolutional neural network (CNN) model for training. Makandar et al.[11] performed contrast-limited adaptive histogram equalization (CLAHE) on mammograms. They focused on 20 images in the MIAS database and found that CLAHE can effectively denoise mammograms and enhance their contrast. In addition to the CLAHE method, breast mammogram preprocessing methods include image normalization, data augmentation, and image resizing[12,13]. To obtain a high-performance CAD system for breast cancer, Al-Masni et al.[12] adopted the image normalization technique. Li et al.[10] stated that image normalization can deal with the differences in images caused by the use of different equipment.

* Department of Industrial Engineering & Management
National Chin-Yi University of Technology
Taiwan.
E-mail: huangml@ncut.edu.tw

Jiao et al.[14] used a CNN to classify 600 breast mass images from the Digital Database for Screening Mammography (DDSM). Dhungel et al.[15] used the CNN method to extract features from the INbreast database and then used the random forest method to classify breast mass mammograms (BMMs). A classification accuracy of 95.00% was achieved in the aforementioned study. Al-Masni et al.[12] used fully connected neural networks to classify breast mass images from the DDSM and achieved a classification accuracy of 97.00%. Chougrad et al.[16] applied different CNN models, such as VGG16, ResNet50, and InceptionV3, to classify breast mass images from t8he DDSM and the INbreast, BCDR, and MIAS databases. The results indicated that InceptionV3 had the highest classification accuracy for all the databases. A higher accuracy was obtained when combining the DDSM, INbreast database, and BCDR database than when using a single database. Al-antari et al.[17] used DL methods to classify breast mammograms from the INbreast database. They used the transfer learning AlexNet model for mass classification and achieved a classification accuracy of 95.64%. The aforementioned authors also achieved a high training speed (each image required 12.23 s for classification).

The aforementioned studies used the existing CNN model to classify breast mammograms. In addition, some studies have made changes to the CNN model architecture to decrease the classification time and increase the classification accuracy. Cai et al.[18] improved DenseNet by adding the squeeze and excitation DenseBlock (SE-DenseBlock) into the DenseNet. They classified breast mammograms from the DDSM and BCDR database. Furthermore, the aforementioned authors compared SE-DenseNet with VGG16, Inceptionv3, ResNet, and DenseNet and found that SE-DenseNet exhibited the best performance. Li et al.[10] used a self-collected database and added the inception concept to the first convolutional layer in DenseNet to create DenseNet-II. DenseNet-II can classify breast masses as benign or malignant more accurately and quickly than DenseNet can. The aforementioned authors compared five models: DenseNet-II, DenseNet, AlexNet, VGGNet, and GoogLeNet. Their results indicated that DenseNet-II outperformed the other four models for all the model evaluation indicators used. The accuracy, sensitivity, and specificity of DenseNet-II were 94.55%, 95.60%, and 95.36%, respectively.

Sun et al.[19] developed a multiview CNN to classify masses from the DDSM and MIAS database. They also set different dilation factors for the convolutional layer in the CNN to extract features at different scales. The aforementioned authors compared the multiview CNN with other popular CNN models (such as ShuffleNet and InceptionV4) and found that the multiview CNN exhibited the highest classification accuracy (82.02%). Agnes et al.[20] improved the CNN model and proposed a multiscale all CNN (MA-CNN). They used different dilation factors for the convolutional layer and removed all the pooling layers from the model. Only the large-stride convolutional layer remained in the model.

The aforementioned authors used four CNN methods, namely the (1) original CNN, (2) all convolutional CNN, (3) multiscale CNN, and (4) MA-CNN, to classify normal tissues, benign masses, and malignant masses in images from the MIAS database. They found that the MA-CNN (accuracy of 96.47%) outperformed the original CNN model (accuracy of 81.13%), all convolutional CNN (accuracy of 89.80%), and multiscale CNN model (accuracy of 90.70%). The MA-CNN can use different dilation factors to extract features at different scales; thus, it provides higher classification accuracy than other methods do.

As indicated by the aforementioned studies, the study of breast mammograms through DL methods has become a major research trend. Studies have indicated that the dilation CNN can extract considerable feature information. Moreover, high-accuracy classification results can be obtained using a large-stride convolutional layer to replace the pooling layer. Therefore, this study proposed using AlexNet II and the double-dilation non-pooling CNN (DDNPNet), which are based on the aforementioned improved models, for breast mammogram classification. The INbreast database, MIAS database, and DDSM are common breast mammogram databases used in breast cancer research. The breast mammograms used in this study were obtained from the INbreast database, MIAS database, and DDSM. We also combined these three databases to establish a new BMM database. Data augmentation must be performed to avoid overfitting in DL. This study first preprocessed the breast images by using four methods: image resizing, CLAHE, image normalization, and data augmentation. Then, the preprocessed images were input into the AlexNet II and DDNPNet models to classify them. The classification performance of the proposed methods was compared with that of three existing CNN models, namely AlexNet, DenseNet, and ShuffleNet.

The remainder of this paper is structured as follows. Section 2 introduces the databases used in this research and the models proposed in this study. Section 3 presents the results of this study. Section 4 provides a comparison between existing CNN methods and the proposed models for breast mammogram classification. Finally, Section 5 presents the conclusion of this study.

MATERIALS AND METHODS

The study flowchart is depicted in Figure 1. We collected breast mammograms from the INbreast database, MIAS database, and DDSM and then performed image preprocessing. The image preprocessing methods used in this study were image resizing, CLAHE, image normalization, and data augmentation. A CNN was used to classify the breast mammograms from the DDSM and the INbreast, MIAS, and BMM databases.

We proposed two models, namely AlexNet II and DDNPNet, for breast mammogram classification. Compared with other classification models, the aforementioned two models can extract more diverse features and require a shorter training time. Four model evaluation indicators, namely accuracy, specificity, sensitivity, and F1 score, were used to assess the performance of the two proposed models. Finally, the performance of the proposed models was compared with that of the AlexNet, DenseNet, and ShuffleNet models.

Data Extraction: Breast mammograms from the INbreast database[21] were collected from Centro Hospitalar de S. Joao, Breast Centre, Porto. The aforementioned dataset contains 410 breast mammogram images, of which we collected 106 images displaying breast masses. Among the 106 images, 35 images displayed benign masses and 71 images displayed malignant masses. The format of the images was DICOM. The DICOM format was converted to the PNG format by using MATLAB software. The INbreast database contains breast mammograms captured from two perspectives: the craniocaudal (CC) and mediolateral oblique (MLO) perspectives[22].

The MIAS database (MiniMammographic Database)[23] contains 322 breast mammograms in the PGM format. Of these images, 53 images display breast masses. Of the aforementioned 53 images, 33 depict benign masses and 20 depict malignant masses. The images of the MIAS database were converted from the PGM format to the PNG format by using MATLAB software. The database website clearly indicates whether the mammograms display benign or malignant tumors. The MIAS database contains mammograms with 8- and 32-bit depth. We used MATLAB software to normalize the bit depth to 8.

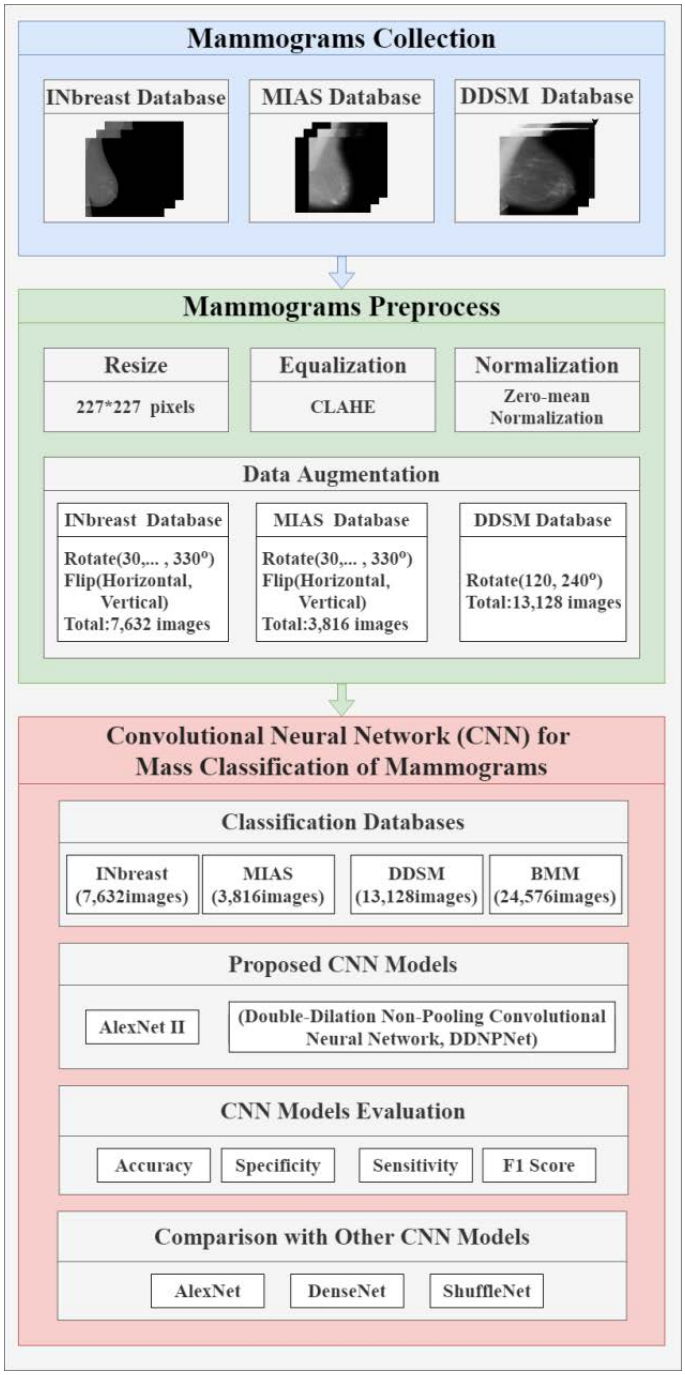


Figure 1: Flowchart

The aforementioned database only includes images captured from the MLO perspective[24]. The MIAS database contains image perspective information, which may affect model classification. Therefore, we cropped the mammograms from the MIAS database and only retained the breast region.

The DDSM[25] contains mammograms in the LJPEG format. LJPEG is a specialize format that can maintain the clarity of the original image. Thus, we used Ruby language in Cygwin to convert the DDSM images from the LJPEG format to the PNG format. The bit depth of the DDSM images is inconsistent; therefore, MATLAB software was used to normalize the bit depth to 8. The original DDSM comprises 10,480 images. After conducting sorting, 2,188 images with breast masses were included in this study. Of these 2,188 images, 995 depicted benign masses and 1,193 depicted malignant masses. Moreover, the DDSM contains images captured from the CC and MLO perspectives[26,27]. Because the DDSM images contain image perspective information, we cropped the images and only retained the breast region.

The MIAS database and DDSM do not follow the BI-RADS seven-level standard. The websites of these databases indicate whether the mass displayed in an image is benign or malignant. The INbreast database follows the BI-RABS standard. If a BMM from the INbreast database has a BI-RADS score of 1-3, the corresponding breast mass is categorized as a benign mass. Moreover, if a BMM from the INbreast database has a BI-RADS score of 4-6, the corresponding breast mass is classified as a malignant mass[17]. This study adopted the aforementioned rules for classifying images from the INbreast database.

Image Preprocessing: Four image preprocessing methods were used in this study: image resizing, CLAHE, image normalization, and data augmentation.

First, all the collected images were resized to 227 × 227 pixels to be input into the adopted CNN classification model. Then, CLAHE was conducted for image preprocessing. CLAHE is an extension of histogram equalization (HE) and adaptive HE (AHE). Although HE can make the target area more prominent, problems related to background noise still exist. To solve the issue of background noise in HE, AHE was proposed. AHE can make the target area more prominent than HE can; however, the drawback of background noise still occurs. Therefore, Zuiderveld[28] proposed the CLAHE method, which not only makes the target area clearly visible but also prevents background noise from interfering with the image. In this study, the original images and the images obtained after CLAHE were included in the BMM image classification model. Thus, 212 images were obtained for the INbreast database after CLAHE. Of these images, 70 displayed benign masses and 142 displayed malignant masses. A total of 106 images were obtained for the MIAS database after CLAHE. Of these images, 66 images displayed benign masses and 40 displayed malignant mass. A total of 4,376 images were obtained for the DDSM after CLAHE. Of these images, 1,990 depicted benign masses and 2,386 depicted malignant masses. Finally, 4,694 images were obtained for the BMM database after CLAHE. Of these images, 2,126 depicted benign masses and 2,568 depicted malignant masses. Table 2 presents the number of images obtained after CLAHE for each database.

As displayed in Figure 2, the position of the breast mass was clearer in the image obtained after CLAHE than in the original image. Thus, conducting CLAHE on breast mass images can enhance the ability of a classification model to learn the characteristics of the breast mass and the classification performance of the model.

Studies have indicated that image normalization can not only solve the problem of uneven lighting in an image but also increase the robustness of a classification model[10]. Because the INbreast and MIAS databases contained a small number of images, the number of images was augmented by performing multiangle rotation (θ = 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330°) as well as horizontal and vertical flipping. The DDSM initially contained a large number of images; therefore, only two-angle rotation (θ = 120° and 240°) was performed to augment the number of images in the DDSM. Data augmentation can not only increase the number of samples but also prevent the model from overfitting due to a small number of images. Table 3 presents the number of images obtained after data augmentation for each database.

Proposed CNN Models: Different patients have different sizes and shapes of breast masses. This study integrated three breast mammogram

Table 1: Comparison of the INbreast database, MIAS database, and DDSM[24]

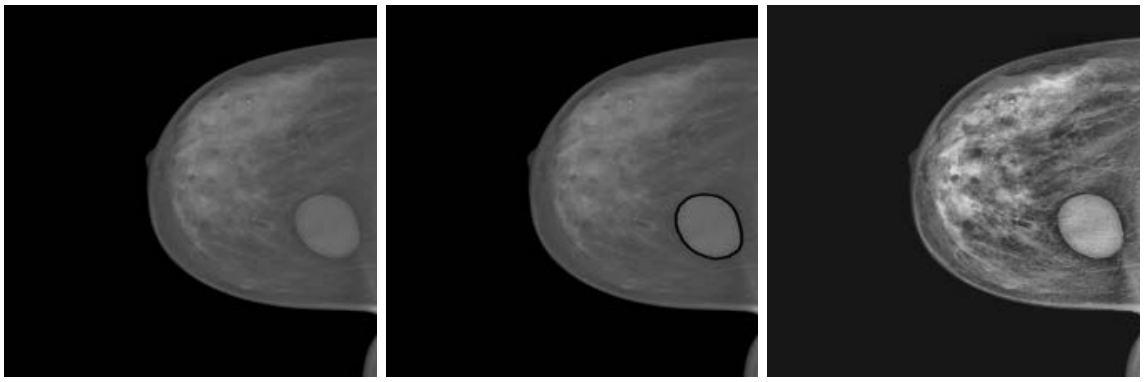
	INbreast database	MIAS database	DDSM database
Original image format	DICOM	PGM	LJPEG
Original image size	2560*3328 pixel 3328*4084 pixel	1024*1024 pixel	Different sizes (pixels)
View	CC & MLO	MLO	CC & MLO
Total number of images (pictures)	410	322	10,480
Mass categories	According to BI-RADS standard	Benign/ Malignant	Benign/ Malignant
Number of masses	106	53	2,188

Table 2: Number of images obtained after CLAHE for each database

	INbreast database	MIAS database	DDSM database	BMM database
Benign	70	66	1,990	2,126
Malignant	142	40	2,386	2,568
Total	**212**	**106**	**4,376**	**4,694**

Table 3: Number of images after data augmentation for each database

	INbreast database	MIAS database	DDSM database	BMM database
Benign	2,520	2,376	5,970	10,866
Malignant	5,112	1,440	7,158	13,710
Total	7,632	3,816	13,128	24,576



(a) (b) (c)

Figure 2: Comparison between the images obtained before and after CLAHE: (a) original image, (b) location of the breast mass marked by a black circle, and (c) image obtained after CLAHE

databases. The background and capturing methods for each database are marginally different; therefore, to extract diverse features that can facilitate image classification, this study used a dilation convolutional (DC) layer to improve the existing AlexNet model and developed two models, namely AlexNet II and DDNPNet. Fig. 3 displays examples of different dilation factors. In Figure 3, d denotes the dilation factors. For example, when d = 1, simple convolution is performed for feature extraction; when d = 2, feature extraction is performed in two intervals; and when d = 3, feature extraction is performed in three intervals.

As displayed in Figure 3 and Figure 4, double-dilation convolution can increase the number of features extracted by a model. Moreover, features in different dimensions can be extracted using different dilation factors. Convolution involves multiplying two matrices to obtain a feature map. A simple convolution can be performed using Eq. (1)[20]. Dilation convolution also involves the multiplication of two matrices to form a feature map. By using different dilation factors, feature extraction can be performed in an extensive manner for different intervals in the original image. The calculation formula for different dilation factors is presented in Eq. (2)[20].

$$C[x,y] = I[x,y] \times H[x,y] \quad (1)$$

$$C[x,y] = I[x,y] \times_d H[x,y] \quad (2)$$

where C [x,y] denotes the coordinate axis position of the feature map extracted through convolution, I [x,y] denotes the coordinate axis position of the original image, H [x,y] denotes the coordinate axis position of the convolution kernel, and d denotes the dilation factor.

In a simple CNN, a series of convolution and pooling procedures must be performed to reduce the image size gradually. However, the models proposed in this study do not contain pooling layers but only contain the convolutional layer. To compensate for the absence of pooling layers, this study enlarged the stride of the convolutional layer. In general, the size of the image is affected by many factors, such as the input image size, stride, kernel size, and padding. These factors can determine the image size after convolution or pooling. The formula for calculating the image size after convolution and pooling is presented in Eq. (3)[20]. In this study, the image size was constantly reduced by increasing the stride of the convolutional layer.

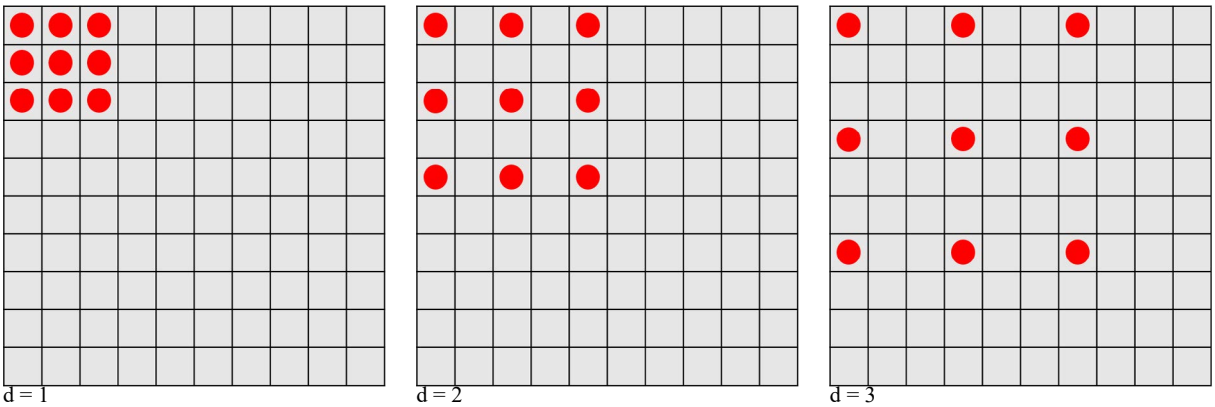


Figure 3: Examples of different dilation factors

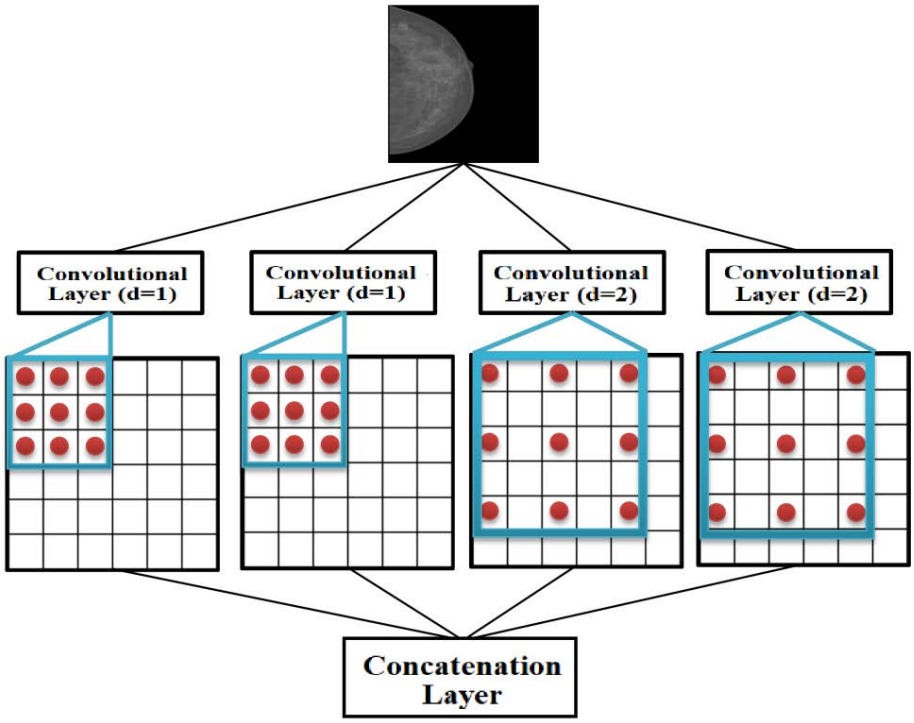


Figure 4: Concept of double-dilation convolution

$$dimension[I]_{subsampled} = \frac{dimension[I]_{input} + 2 \times padding - kernel\ size}{stride} + 1 \quad (3)$$

AlexNet II Model: In this study, the AlexNet model was improved to form the AlexNet II model. The first convolutional layer of the AlexNet model was replaced with a DC layer so that more diverse features could be extracted. In addition, all the pooling layers of the model were removed and replaced with a convolutional layer having a large stride. The number of fully connected layers was changed from 3 to 1. The AlexNet II model operates automatically from the image input step to the classification result output step. It does not require manual feature extraction. Figure 5 displays the architecture of the AlexNet II model.

After an image is input into AlexNet II, it first enters the double DC layers. AlexNet II can obtain features having diverse receptive fields through the convolution operation with different dilation factors (d = 1 and 2). It transfers the rich features extracted from the previous layer to the next convolutional layer through the concatenation layer. In AlexNet II, the concatenation layer and convolutional layer 1 are connected to a rectified linear unit (ReLU) and cross-channel-normalized layer. The remaining convolutional layers are only connected to one ReLU. A ReLU is a common activation function in CNN that can solve the problem of vanishing gradients.

Because double-dilation convolution is an important feature extraction tool for AlexNet II, a stride of 4 was used in double-dilation convolution to extract detailed feature information. Because the aforementioned model lacks a pooling layer, the stride must be increased. Therefore, the strides of the group convolutional layer and convolutional layer set were set as 5 (in the original AlexNet model, the stride is set as 1 or 2). In addition, an increased stride can marginally reduce the model training time.

DDNPNet Model: Figure 6 presents the architecture of the DDNPNet model. In the DDNPNet model, different dilation factors and no pooling layers are used. The difference between DDNPNet and

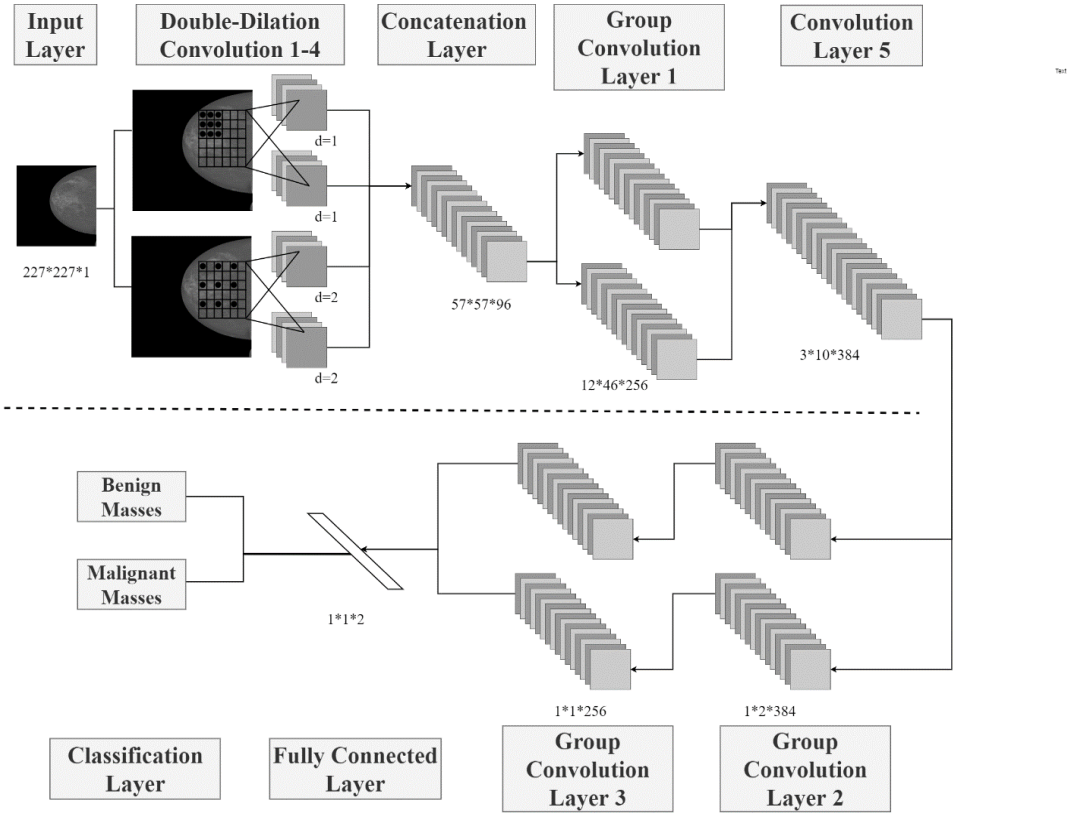


Figure 5: Architecture of the AlexNet II model

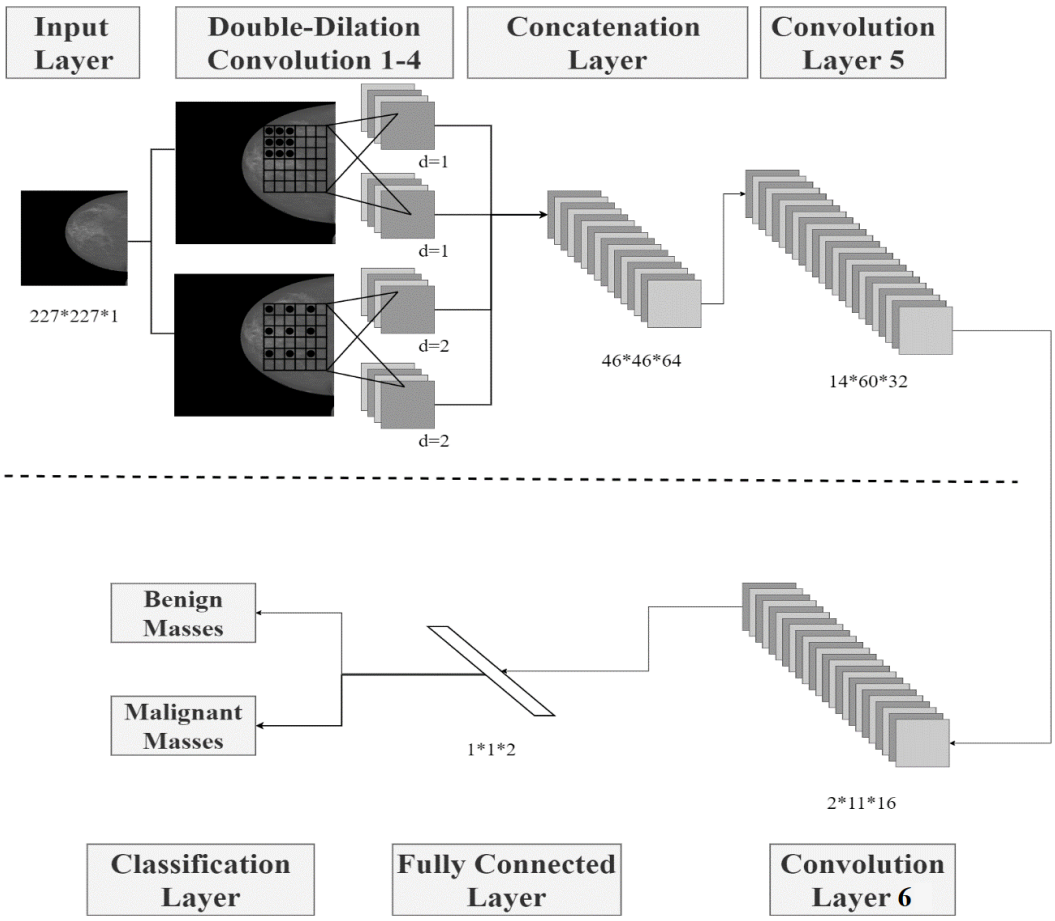


Figure 6: Architecture of the DDNPNet model

AlexNet II is that DDNPNet contains a simple convolutional layer, whereas AlexNet II contains a group convolutional layer. DDNPNet reduces one convolutional layer and thus requires a shorter training time. Furthermore, cross-channel normalization, rather than batch normalization, is used in DDNPNet.

Similar to AlexNet II, DDNPNet is a fully automatic classification model. Images are passed through double DC layers and then enter the concatenation layer. The concatenation layer passes a large number of features extracted from the previous layer to the next convolutional layer. The strides of the convolutional layers of DDNPNet are set as 5 to compensate for the absence of pooling layers. After each convolutional layer, a ReLU layer and batch normalization layer are added. The advantage of batch normalization is that it can speed up the training process and improve the model performance.

Model Evaluation: As presented in Table 4, the model classification results can be categorized into four results: true positives (TPs), false negatives (FNs), false positives (FPs), and true negatives (TNs). Four model evaluation indicators were used in this study: accuracy, specificity, sensitivity, and F1 score.

RESULTS

The experiments in this study were conducted using MATLAB R2019a. The relative parameters were set as follows: initial learning rate = 0.0003; minibatch size = 10; max epochs of AlexNet, AlexNet II, and DDNPNet = 25; and max epochs of DenseNet and ShuffleNet = 5. We used 80% of the input images as the training set and 20% of the input images as the testing set. Moreover, we applied fivefold cross validation. Table 5 presents the number of training and testing image sets. All the experiments in this study were performed on a PC with the following specifications: Intel(R) Core (TM) i7-8700 CPU @ 3.20GHz with 32 GB RAM.

Breast Mass Classification Results for the Inbreast Database:

As presented in Table 6 and Figure 7, for the INbreast database, the accuracy, specificity, sensitivity, F1 score, and training time of AlexNet were 90.75%, 93.05%, 86.08%, 86.07%, and 10 min 44 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of DenseNet were 98.95%, 98.83%, 99.21%, 98.44%, and 2 h 8 min 57 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of ShuffleNet were 94.06%, 95.73%, 90.67%, 90.97%, and 17 min 2 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of AlexNet II were 95.39%, 97.95%, 90.20%, 92.81%, and 9 min 23 s, respectively. Finally, the accuracy, specificity, sensitivity, F1 score, and training time of DDNPNet were 95.77%, 97.83%, 91.59%, 93.45%, and 8 min 23 s, respectively.

Breast mass classification results for the MIAS database

As presented in Table 7 and Figure 8, for the MIAS database, the accuracy, specificity, sensitivity, F1 score, and training time of AlexNet were 93.03%, 86.83%, 96.55%, 94.65%, and 5 min 19 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of DenseNet were 98.03%, 97.43%, 98.40%, 98.42%, and 1 h 1 min 25 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of ShuffleNet were 94.71%, 91.81%, 96.47%, 95.75%, and 8 min 39 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of AlexNet II were 91.93%, 85.84%, 95.62%, 93.65%, and 4 min 39 s, respectively. Finally, the accuracy, specificity, sensitivity, F1 score, and training time of DDNPNet were 93.61%, 87.64%, 97.22%, 94.99%, and 3 min 51 s, respectively

Breast mass classification results for the DDSM database

As presented in Table 8 and Figure 9, for the DDSM, the accuracy, specificity, sensitivity, F1 score, and training time of AlexNet were 99.92%, 99.90%, 99.93%, 99.91%, and 18 min 31 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of DenseNet were 99.93%, 99.90%, 99.97%, 99.92%, and 3 h 32 min 41 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of ShuffleNet were 99.86%, 99.82%, 99.92%, 99.85%, and 29 min 3 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of AlexNet II were 99.12%, 99.16%, 99.08%, 99.04%, and 18 min 21 s, respectively. Finally, the accuracy, specificity, sensitivity, F1 score, and training time of DDNPNet were 99.63%, 99.62%, 99.65%, 99.60%, and 13 min 41 s, respectively.

Breast mass classification results for the BMM database

As presented in Table 9 and Figure 10, for the BMM database, the accuracy, specificity, sensitivity, F1 score, and training time of AlexNet were 94.83%, 94.55%, 95.19%, 94.21%, and 38 min 31 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of DenseNet were 98.59%, 98.21%, 99.08%, 98.43%, and 6 h 39 min 21 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of ShuffleNet were 95.04%, 95.30%, 94.71%, 94.40%, and 53 min 44 s, respectively. The accuracy, specificity, sensitivity, F1 score, and training time of AlexNet II were 93.35%, 93.51%, 93.18%, 92.71%, and 31 min 1 s, respectively. Finally, the accuracy, specificity, sensitivity, F1 score, and training time of DDNPNet were 95.41%, 95.86%, 94.83%, 94.81%, and 26 min 23 s, respectively.

DISCUSSION

Table 10 presents the literature on breast mammogram classification. Ribli et al.[29] used the VGG16 model to classify breast masses and obtained an AUC of 0.95. Al-antaria et al.[17] used the transfer learning AlexNet model for mass classification and achieved a classification accuracy of 95.64%. In this study, the AlexNet model yielded an accuracy of 90.75% for the classification of the images from the INbreast database. However, the AlexNet II model achieved a higher classification accuracy (95.39%) and required a lower training time than did the AlexNet model for classifying the images in the aforementioned dataset. Sun et al.[19] used the multiview and multidilation CNN (MVMDCNN-Loss) to classify breast mass images from the DDSM and the MIAS database. They achieved classification accuracies of 82.02% and 63.06% for the DDSM and the MIAS database, respectively. The DDSM and the MIAS database were also used in this study. AlexNet II, in which dilation convolution is performed, achieved classification accuracies of 99.12% and 91.93% for the DDSM and the MIAS database, respectively. Li et al.[10] improved DenseNet by using a self-collected database to classify breast masses and obtained a classification accuracy of 94.55%. In this study, the original DenseNet model was used for classifying mammograms from four databases. The aforementioned model provided accuracies of more than 98% for all the databases. This result is similar to that of Li[10]. However, DenseNet required a long training time. Agnes et al.[20] used four dilation factors (d = 1, 2, 3, and 4) in the first convolutional layer and removed all the pooling layers from the classification model. The pooling layers were replaced with large-stride convolutional layers to classfy images from the MIAS database. Agnes et al.[20] achieved a classification accuracy of 96.47% with their model. This study used the DDNPNet model for automatic feature extraction and classification and implemented double-dilation convolution in the first convolutional layer. Furthermore, we also removed all the pooling layers from the model. We used DDNPNet to classify breast mass mammograms obtained by integrating three databases (into the BMM database). A

Table 4: Confusion matrix of mass classification

		Prediction	
		Malignant mass (Positive)	**Benign mass (Negative)**
Actual	**Malignant mass (Positive)**	The mass is actually a malignant mass, and the model also classifies the mass as a malignant mass. (TP)	The mass is actually a malignant mass, but the model classifies the mass as a benign mass. (FP)
	Benign mass (Negative)	The mass is actually a benign mass, but the model classifies the mass as a malignant mass. (FN)	The mass is actually a benign mass, and the model also classifies the mass as a benign mass (TN)

Table 5: Number of training and testing image sets

CV	**Training /Testing**	**INbreast**		**MIAS**		**DDSM**		**BMM**	
		Benign	**Malignant**	**Benign**	**Malignant**	**Benign**	**Malignant**	**Benign**	**Malignant**
1	**Training**	2,016	4,089	1,990	1,152	4,776	5,726	8,782	14,023
	Testing	504	1,023	476	288	1,194	1,432	2,174	3,485
2	**Training**	2,017	4,090	1,901	1,153	4,777	5,726	8,695	13,938
	Testing	503	1,022	475	287	1,193	1,432	2,171	3,480
3	**Training**	2,017	4,089	1,901	1,153	4,777	5,727	8,695	13,937
	Testing	503	1,023	475	287	1,193	1,431	2,171	3,481
4	**Training**	2,017	4,090	1,901	1,153	4,777	5,726	8,695	13,938
	Testing	503	1,022	475	287	1,193	1,432	2,171	3,480
5	**Training**	2,013	4,090	1,901	1,149	4,773	5,727	8,687	13,926
	Testing	507	1,022	475	291	1,197	1,431	2,179	3,492

Table 6: Breast mass classification results for the INbreast database

Model	**INbreast Database**				
	Accuracy	**Specificity**	**Sensitivity**	**F1 Score**	**Time**
AlexNet	90.75%	93.05%	86.08%	86.07%	00:10:44
DenseNet	98.95%	98.83%	99.21%	98.44%	02:08:57
ShuffleNet	94.06%	95.73%	90.67%	90.97%	00:17:02
AlexNet II	95.39%	97.95%	90.20%	92.81%	00:09:23
DDNPNet	95.77%	97.83%	91.59%	93.45%	00:08:23

Table 7: Breast mass classification results for the MIAS database

Model	**MIAS Database**				
	Accuracy	**Specificity**	**Sensitivity**	**F1 Score**	**Time**
AlexNet	93.03%	86.83%	96.55%	94.65%	00:05:19
DenseNet	98.03%	97.43%	98.40%	98.42%	01:01:25
ShuffleNet	94.71%	91.81%	96.47%	95.75%	00:08:39
AlexNet II	91.93%	85.84%	95.62%	93.65%	00:04:39
DDNPNet	93.61%	87.64%	97.22%	94.99%	00:03:51

Table 8: Breast mass classification results for the DDSM

Model	**DDSM Database**				
	Accuracy	**Specificity**	**Sensitivity**	**F1 Score**	**Time**
AlexNet	99.92%	99.90%	99.93%	99.91%	00:18:31
DenseNet	99.93%	99.90%	99.97%	99.92%	03:32:41
ShuffleNet	99.86%	99.82%	99.92%	99.85%	00:29:03
AlexNet II	99.12%	99.16%	99.08%	99.04%	00:18:21
DDNPNet	99.63%	99.62%	99.65%	99.60%	00:13:41

Table 9: Breast mass classification for the BMM database

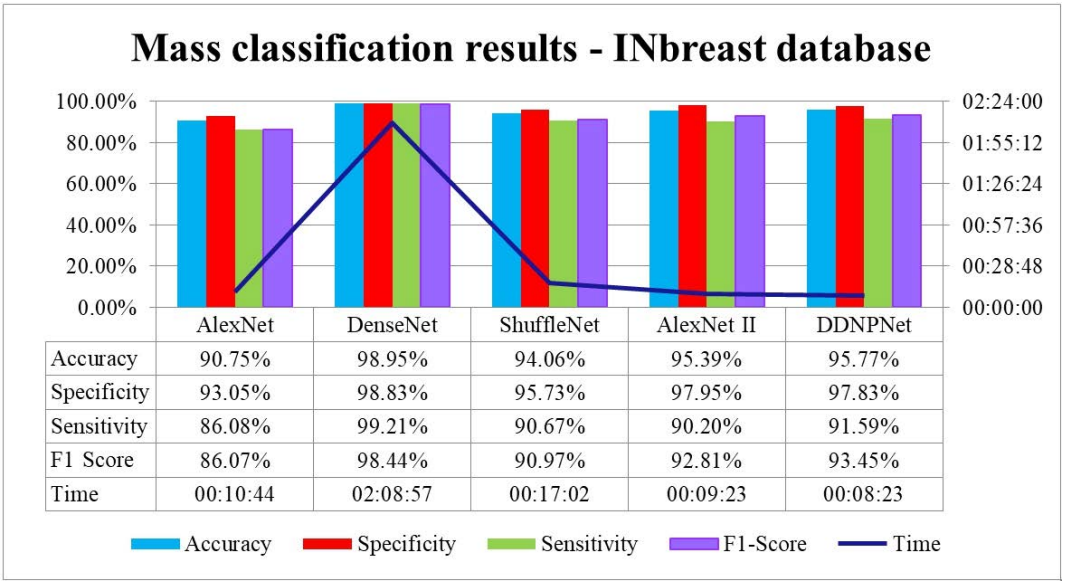
Model	**BMM Database**				
	Accuracy	**Specificity**	**Sensitivity**	**F1 Score**	**Time**
AlexNet	94.83%	94.55%	95.19%	94.21%	00:38:31
DenseNet	98.59%	98.21%	99.08%	98.43%	06:39:21
ShuffleNet	95.04%	95.30%	94.71%	94.40%	00:53:44
AlexNet-II	93.35%	93.51%	93.18%	92.71%	00:31:01
DDNPNet	95.41%	95.86%	94.83%	94.81%	00:26:23

Table 10: Related literature

Literature	**Database**	**Image Preprocessing**	**Classification Model**	**Classification Categories**	**Results (Acc.)**
Ribli et al.[29]	DDSM Self-Collected INbreast (2909 patients)		VGG16	Benign/ Malignant Masses (Binary)	0.95 (AUC)
Al-Masni et al.[12]	DDSM (2400 images)	Multi-Threshold Peripheral Equalization Otsu thresholding 2D Gaussian Low Pass Filter Normalization Data Augmentation (Rotate)	YOLO-based CAD system	Benign/ Malignant Masses (Binary)	97.00%
Al-antaria et al.[17]	INbreast (896 images)	ROI extraction Data Augmentation (Rotate)	Transfer learning-AlexNet	Benign/ Malignant Masses (Binary)	95.64%
Chougrad et al.[16]	INbreast BCDR DDSM Total: 6116	ROI extraction Normalization Data Augmentation (Rotate, Shear, Zoom..)	VGG16 ResNet50 Iinceptionv3	Benign/ Malignant Masses (Binary)	98.94%
Li et al.[10]	Self-Collected (2402 patients)	Zero-mean normalization Data Augmentation (Rotate, Zoom, Mirror)	DenseNet-II	Benign/ Malignant Masses (Binary)	94.55%
Sun et al.[19]	DDSM (1445 images)		MVMDCNN-Loss	Benign/ Malignant Masses (Binary)	82.02%
	MIAS (644 images)				63.06%
Agnes et al.[20]	MIAS (4500 images)	Medium-Filter Remove Pectoral Muscles Data Augmentation (Random Rotate, Flip)	MA-CNN	Benign/ Malignant Masses and Normal (Triple)	96.47%
Proposed Method	INbreast (7693 images)	Normalization CLAHE Data Augmentation (Rotate, Flip)	AlexNet	Benign/ Malignant Masses (Binary)	90.75%
			DenseNet		98.95%
			ShuffleNet		94.06%
			AlexNet II		95.39%
			DDNPNet		95.77%
Proposed Method	MIAS (3816 images)	Normalization CLAHE Data Augmentation (Rotate, Flip)	AlexNet	Benign/ Malignant Masses (Binary)	93.03%
			DenseNet		98.03%
			ShuffleNet		94.71%
			AlexNet II		91.93%
			DDNPNet		93.61%
Proposed Method	DDSM (13,128 images)	Normalization CLAHE Data Augmentation (Rotate, Flip)	AlexNet	Benign/ Malignant Masses (Binary)	99.92%
			DenseNet		99.93%
			ShuffleNet		99.86%
			AlexNet II		99.12%
			DDNPNet		99.63%
Proposed Method	BMM (24,576 images)	Normalization CLAHE Data Augmentation (Rotate, Flip)	AlexNet	Benign/ Malignant Masses (Binary)	94.83%
			DenseNet		98.59%
			ShuffleNet		95.04%
			AlexNet II		93.35%
			DDNPNet		95.41%

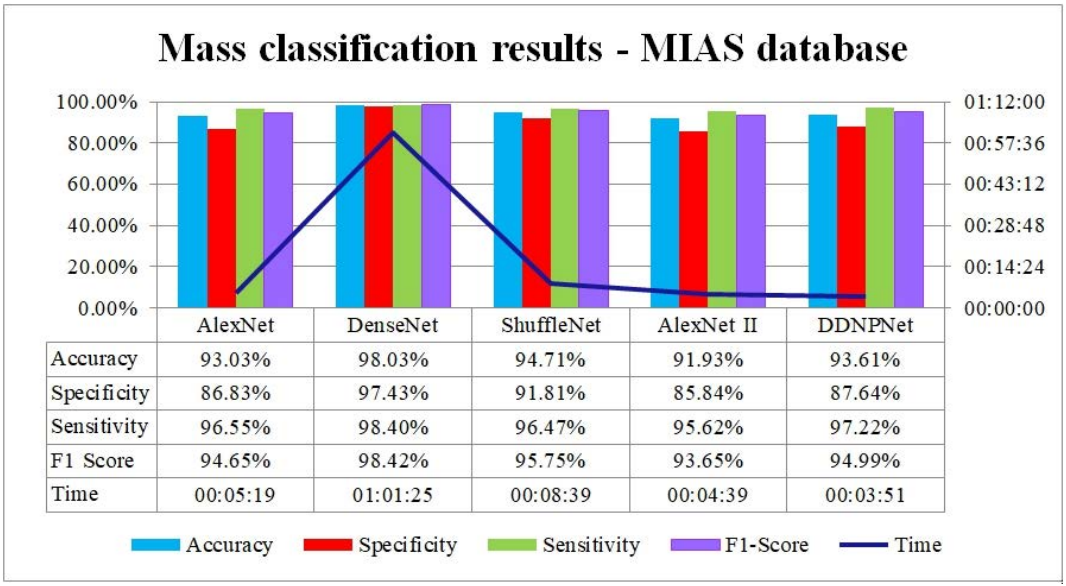
Table 11: Comparison of the adopted classification models

Model	**Characteristic**	**Advantage**	**Disadvantage**
AlexNet	The beginning of the CNN model.	Model training time is shorter than DenseNet and ShuffleNet, while still maintaining well accuracy in the MIAS database.	Model training time still takes longer than AlexNet II and DDNPNet.
DenseNet	Layer by layer, maintaining rich feature information.	Obtained the highest accuracy.	Consume a lot of computing resources, and the training time is longer.
ShuffleNet	Shuffle between channels and exchange feature information with others.	Can maintained well accuracy in MIAS database.	Model training time longer than AlexNet, AlexNet II, and DDNPNet.
AlexNet II	Using different dilation factors and removing all pooling layers.	Model training time is shorter than AlexNet II and DDNPNet.	The accuracy is lower in MIAS database.
DDNPNet	Using different dilation factors and removing all pooling layers, to extract more features.	The model training time is the shortest.	The accuracy is lower in MIAS database.



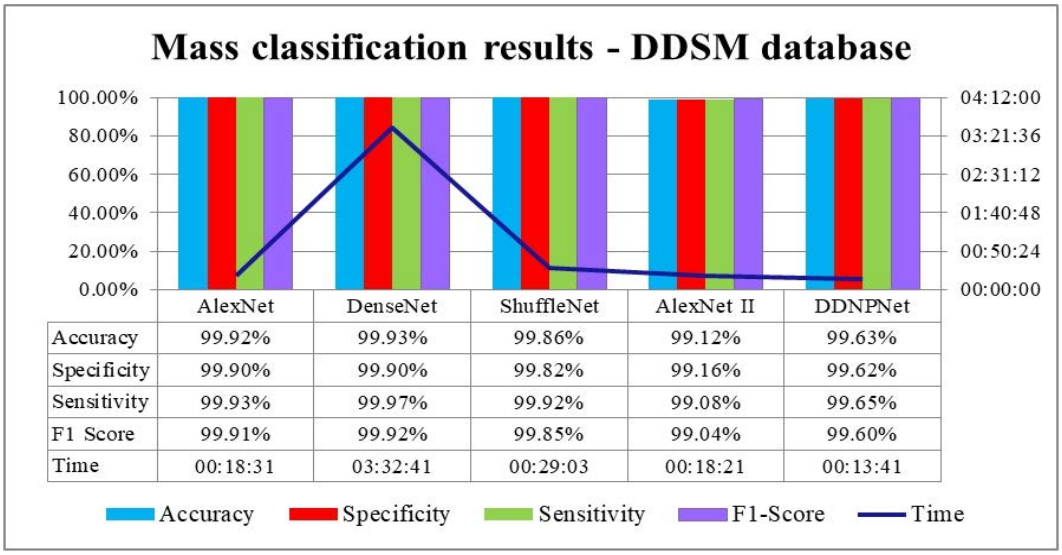
	AlexNet	DenseNet	ShuffleNet	AlexNet II	DDNPNet
Accuracy	90.75%	98.95%	94.06%	95.39%	95.77%
Specificity	93.05%	98.83%	95.73%	97.95%	97.83%
Sensitivity	86.08%	99.21%	90.67%	90.20%	91.59%
F1 Score	86.07%	98.44%	90.97%	92.81%	93.45%
Time	00:10:44	02:08:57	00:17:02	00:09:23	00:08:23

Figure 7: Breast mass classification results for the INbreast database



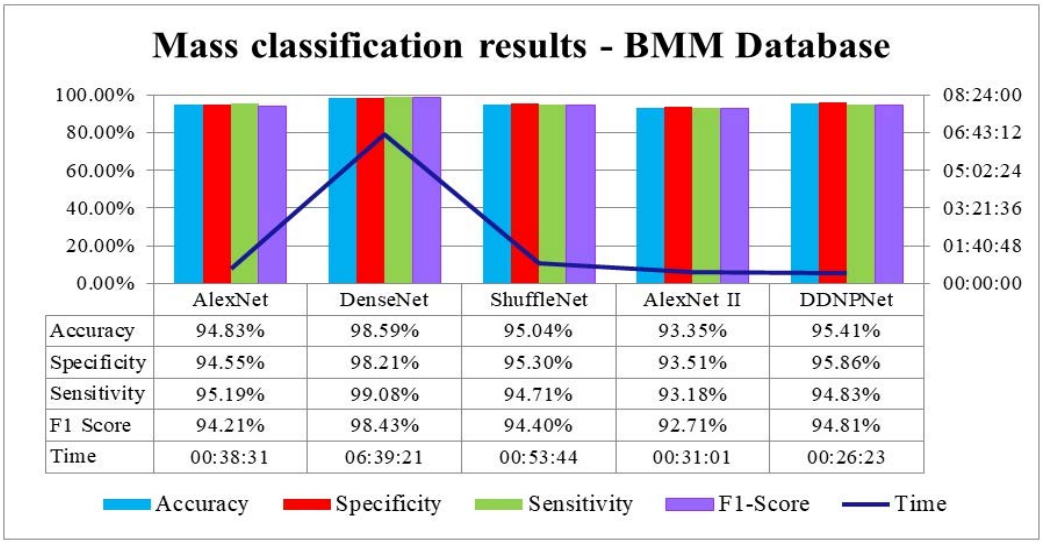
	AlexNet	DenseNet	ShuffleNet	AlexNet II	DDNPNet
Accuracy	93.03%	98.03%	94.71%	91.93%	93.61%
Specificity	86.83%	97.43%	91.81%	85.84%	87.64%
Sensitivity	96.55%	98.40%	96.47%	95.62%	97.22%
F1 Score	94.65%	98.42%	95.75%	93.65%	94.99%
Time	00:05:19	01:01:25	00:08:39	00:04:39	00:03:51

Figure 8: Breast mass classification results for the MIAS database



	AlexNet	DenseNet	ShuffleNet	AlexNet II	DDNPNet
Accuracy	99.92%	99.93%	99.86%	99.12%	99.63%
Specificity	99.90%	99.90%	99.82%	99.16%	99.62%
Sensitivity	99.93%	99.97%	99.92%	99.08%	99.65%
F1 Score	99.91%	99.92%	99.85%	99.04%	99.60%
Time	00:18:31	03:32:41	00:29:03	00:18:21	00:13:41

Figure 9: Breast mass classification results for the DDSM



	AlexNet	DenseNet	ShuffleNet	AlexNet II	DDNPNet
Accuracy	94.83%	98.59%	95.04%	93.35%	95.41%
Specificity	94.55%	98.21%	95.30%	93.51%	95.86%
Sensitivity	95.19%	99.08%	94.71%	93.18%	94.83%
F1 Score	94.21%	98.43%	94.40%	92.71%	94.81%
Time	00:38:31	06:39:21	00:53:44	00:31:01	00:26:23

Figure 10: Breast mass classification results for the BMM database

classification accuracy of 95.41% was obtained with DDNPNet for the BMM model. This result was similar to that of Agnes[20]. Moreover, DDNPNet provided a marginally lower classifcation accuracy but required a considerably shorter training time than DenseNet did (classification accuracy of 98.59% for DenseNet and 95.41% for DDNPNet; training time of 6 h 39 min 21 s for DenseNet and 26 min 23 s for DDNPNet).

ome studies segmented the pectoral muscles in the preprocessing methods or only extracted the surrounding area of the breast mass into the classification model because the position of the pectoral muscles affects the accuracy of the classification model. Because breast masses are sometimes located at the pectoral muscles, we did not segment the pectoral muscles in this study. Although pectoral muscle segmentation was not performed, high classification accuracies were obtained in this study.

Different models have different advantages and disadvantages. For example, although DenseNet provided the highest classification accuracy, it required considerable computing resources and a long training time (the training time of DenseNet for the BMM database was 6 h 39 min 21 s) due to its layer-by-layer connection structure. However, this layer-by-layer structure prevents the loss of important features due to layer transmission, and the last layer can receive the effective feature information extracted from the previous layers.

Many studies on breast mammogram classification have not integrated the INbreast database, MIAS database, and DDSM for related research on breast mass classification. However, this study found that after integrating the aforementioned databases, the accuracies of the adopted classification methods did not decrease. The DDNPNet model had a higher accuracy than the AlexNet and ShuffleNet models and required less training time than the AlexNet, ShuffleNet, and DenseNet models. Thus, DDNPNet exhibited a satisfactory performance in the classification of benign and malignant breast masses. Table 11 presents a comparison of the characteristics, advantages, and disadvantages of the adopted classification models.

Additional experiments must be conducted in the future to determine whether the proposed DDNPNet model is suitable for other databases. Future studies can also apply DDNPNet to other databases to obtain comprehensive results. In addition, the proposed DDNPNet model can act as a diagnosis reference for physicians and assist busy physicians in aided diagnosis. The advantages and limitations of this study are described in the following text.

The advantages of the models proposed in this study and the study overall are as follows:

The models used in this study do not require complicated manual feature extraction.

The proposed DDNPNet model operates automatically from the image input step to the classification output step.

The DDNPNet model provides high classification accuracy for the database obtained after integrating three databases (i.e., BMM database).

DDNPNet requires the shortest training time among the compared models.

This study classified images from three common breast mammogram databases.

The limitations of this study are as follows:

The images of each category in DDNPNet were unbalanced.

The collected breast mammograms were obtained from Western European and American databases. Future studies can also apply the proposed models for the classification of the breast mass mammograms of Eastern individuals.

CONCLUSION

The AlexNet II and DDNPNet models proposed in this study exhibited high classification accuracies for the DDSM and the INbreast, MIAS, and BMM databases. The AlexNet II and DDNPNet models can extract numerous diverse features from images; therefore, these models can provide accuracies comparable to those of existing CNN methods.

This study makes two contributions. First, we used three common breast mammogram databases, namely the INbreast database, MIAS database, and DDSM, for research. We also integrated the

aforementioned three databases into the BMM database for breast mass classification. Second, in addition to using the existing CNN model, this study constructed the AlexNet II and DDNPNet models by using different dilation factors and replacing all the pooling layers with large-stride convolutional layers.

This study classified the breast mammograms of Western patients; however, differences exist in the body shapes of Western and Eastern individuals. Therefore, the proposed models may not be suitable to other races. Future studies can apply the proposed classification models to different breast cancer detection tools, such as breast ultrasound images, breast photomicrograph images, breast computed tomography images, and breast MRI images. Moreover, studies can continue to improve the architecture of the CNN model by using dilation factors and removing all the pooling layers to obtain superior classification accuracies.

Authorship Contribution: Mei-Ling Huang: Conceptualization, Methodology, Writing- Original draft preparation, Investigation, Supervision, Writing-Reviewing and Editing, Funding acquisition. Ting-Yu Lin: Software, Formal Analysis, Data curation

Potential Conflict of Interest: None

Competing Interest: None

Acceptance Date: 14 June 2022

Sponsorship: The research was funded by the Ministry of Science and Technology of Taiwan, R.O.C. (Research Grant Project number MOST 109-2221-E-167-024-MY2).

REFERENCES

1. American Cancer Society, Breast Cancer Facts and Figures 2019-2020, Atlanta Am Cancer Soc 2019. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, CA. Cancer J Clin 2015;65(1):5-29.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, CA. Cancer J Clin 2016;66(1):7-30.
4. Siegel RL, Miller KD, Jemal A. Cancer Statistics, CA. Cancer J Clin 2017;67(1):7-30.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, CA. Cancer J Clin 2018;68(1):7-30.
6. Siegel RL, Miller KD, Jemal A. Cancer statistics, CA. Cancer J Clin 2019;69(1):7-34.
7. Siegel RL, Miller KD, Jemal A. Cancer statistics, CA. Cancer J Clin 2020;70(1):7-30.
8. Ardakani AA, Gharbali A, Mohammadi A. Classification of breast tumors using sonographic texture analysis. J Ultrasound Med 2015;34(2):225-31.
9. American College of Radiology, Breast Cancer Early Detection and Diagnosis. 2019. https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/mammograms.html.
10. Li H, Zhuang S, Li D, et al. Benign and malignant classification of mammogram images based on deep learning. Biomed Signal Process Control 2019;51:347-54.
11. Makandar A, Halalli B. Breast cancer image enhancement using median filter and CLAHE. Int J Sci Eng Res 2015;6:462-5.
12. Al-masni MA, Al-antari MA, Park JM, et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Programs Biomed 2018;157:85-94.
13. Jung H, Kim B, Lee I, et al. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PLoS One 2018;13(9):e0203355.
14. Jiao Z, Gao X, Wang Y, et al. A deep feature-based framework for breast masses classification. Neurocomputing 2016;197:221-31.
15. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal 2017;37:114-28.
16. Chougrad H, Zouaki H, Alheyane O. Deep Convolutional Neural Networks for breast cancer screening. Comput Methods Programs Biomed 2018;157:19-30.
17. Al-Antari MA, Al-Masni MA, Choi MT, et al. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Inform 2018;117:44-54.
18. Cai Q, Liu X, Guo Z. Identifying Architectural Distortion in Mammogram Images Via a SE-DenseNet Model and Twice Transfer Learning, in: 2018 11th Int Congr Image Signal Process. Biomed Eng Informatics 2018;1-6.
19. Sun L, Wang J, Hu Z, et al. Multi-View Convolutional Neural Networks for Mammographic Image Classification. IEEE Access 2019;7:126273-82.
20. Agnes SA, Anitha J, Pandian SIA, et al. Classification of mammogram images using multiscale all convolutional neural network (MA-CNN). J Med Syst 2020;44(1):30.
21. INbreast database. http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database.
22. Moreira IC, Amaral I, Domingues I, et al. Inbreast: toward a full-field digital mammographic database. Acad Radiol 2012;19(2):236-48.
23. MIAS database. http://peipa.essex.ac.uk/info/mias.html.
24. Suckling P. The mammographic image analysis society digital mammogram database. Digit Mammo 1994;375-86.
25. DDSM database. http://www.eng.usf.edu/cvprg/Mammography/Database.html.
26. Heath M, Bowyer K, Kopans D, et al. Current status of the digital database for screening mammography. Digit Mammogr Springer 1998;457-60.
27. Bowyer K, Kopans D, Kegelmeyer WP, et al. The digital database for screening mammography. Int Work Digit Mammogr 1996;27.
28. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization, in: Graph. Gems IV, Academic Press Professional USA. Graph Gems 1994;474-85.
29. Ribli D, Horváth A, Unger Z, et al. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep 2018;18(1):4165.