Research Article

Is Intensity Inhomogeneity Correction Useful for Classification of Breast Cancer in Sonograms Using Deep Neural Network?

Chia-Yen Lee [ID],[1] **Guan-Lin Chen,**[1] **Zhong-Xuan Zhang,**[1] **Yi-Hong Chou,**[2] **and Chih-Chung Hsu** [ID][3]

[1]*Department of Electrical Engineering, National United University, Miaoli, Taiwan*
[2]*Department of Radiology, Taipei Veterans General Hospital and National Yang Ming University, Taipei, Taiwan*
[3]*Department of Management Information Systems, National Pingtung University of Science and Technology, Neipu, Taiwan*

Correspondence should be addressed to Chih-Chung Hsu; cchsu@mail.npust.edu.tw

The sonogram is currently an effective cancer screening and diagnosis way due to the convenience and harmlessness in humans. Traditionally, lesion boundary segmentation is first adopted and then classification is conducted, to reach the judgment of benign or malignant tumor. In addition, sonograms often contain much speckle noise and intensity inhomogeneity. This study proposes a novel benign or malignant tumor classification system, which comprises intensity inhomogeneity correction and stacked denoising autoencoder (SDAE), and it is suitable for small-size dataset. A classifier is established by extracting features in the multilayer training of SDAE; automatic analysis of imaging features by the deep learning algorithm is applied on image classification, thus allowing the system to have high efficiency and robust distinguishing. In this study, two kinds of dataset (private data and public data) are used for deep learning models training. For each dataset, two groups of test images are compared: the original images and the images after intensity inhomogeneity correction, respectively. The results show that when deep learning algorithm is applied on the sonograms after intensity inhomogeneity correction, there is a significant increase of the tumor distinguishing accuracy. This study demonstrated that it is important to use preprocessing to highlight the image features and further give these features for deep learning models. In this way, the classification accuracy will be better to just use the original images for deep learning.

1. Introduction

In recent years, with dietary habit and the change of lifestyle, a quick rise has been seen in female's breast cancer. In the developed and developing countries, breast cancer has, respectively, become the first and second causes of women's cancer death. According to the article published in International Journal of Cancer, 1993, one in every 8 women in European and American countries has breast cancer [1]. American Cancer Society pointed out as early as in 1999 that breast cancer is the first cause of women's death in the developed countries [2]. In the various breast cancer diagnosis procedures, ultrasound diagnosis is regarded as a kind of "highly accepted" tool because of low price, convenience, universality, and nonradiation. Ultrasound has become a necessary diagnosis tool in all medical centers and a good tool for doctors to initially diagnose the breast cancer clinically.

However, ultrasonic images often contain lots of speckles, noises, and textures so that it is unable to clearly discriminate the detail changes of tissues, e.g., the tumor size and boundary. Therefore, many studies in the past have put forward ultrasound tumor segmentation algorithm, including the well-known deformation model [3, 4], clustering [5], overzero [6], thresholding [7], watershed method [8], and level set method [9, 10]. The above methods can be roughly divided into two categories, respectively, based on pixel or region information [11]. However, in order to make the segmentation results well [12–14], many experiment steps and parameter adjustments are involved. It may have

been affected by the developer's subjective ideas and habits so that these results may be not sufficiently objective, affecting the accuracy in doctor's judgment with naked eyes, and it is likely to cause misdiagnosis. Therefore, the traditional computer-aided diagnosis tool is limited.

Recently, deep learning has been widely used in various applications such as facial recognition [15, 16], object detection, and car identification. In medical imaging, deep learning still offers excellent performance on different fields [17, 18]. With the development of deep learning, the system architecture of computer-aided diagnosis has been changed, and the advantages unavailable in traditional computer-aided diagnosis system have been added. In the deep learning method, it is only needed to specify the training data, and then, the features in the images can be automatically extracted, and more abstract feature descriptions can be extracted according to the nerve cell depth of each layer, including the features from point to edge, contour, and even a higher level so as to gradually reduce the nerve cells at the same time of improving the features. In the deep learning method, selection and extraction of features, as well as data classification, are established under the same structure, which has a higher accuracy than traditional feature extraction.

To address the problem of the breast sonogram detection, there are several approaches that can be used such as a convolutional neural network (CNN) [19–21], deep autoencoder (DAE), or stack denoised autoencoder (SDAE). Zhou et al. demonstrated the importance of feature extraction and selection for tumor classification and confirmed that the classification effect based on the Shearlet-based texture feature was better than those of the other four feature extraction methods by AdaBoost and SVM. In the classification of 200 benign and malignant tumors, the accuracy is up to 90%, indicating the effectiveness of using machine learning for breast tumor classification [22]. In 2017, Moon added adaptive filtering to the CAD system (A-CAD) to emphasize the characteristics of tumor size. CAD was more robust when classifying tumors larger than or equal to 1 cm. The results showed that the accuracy increased from 73.1% to 81.4% after adding the adaptive filter. It is suggested that the tumor classification effect is better after using adaptive filtering [23]. Due to the speckle noise of ultrasonic images, Raha used four kinds of filters (median, high boost, Sobel, and average filter) for image preprocessing, and watershed was used for segmentation. Finally, the k-nearest neighbors algorithm was performed for the classification of benign and malignant tumors with a high accuracy of 96.4% [24]. Abdel-Nasser extracted a high-resolution (HR) image from a set of low-resolution (LR) images using super-resolution to overcome the speckle noise and artifacts problems. The 31 benign and 28 malignant tumors were divided into training and test data by leave-one-out cross validation (LOOCV) and classified by random forests [25]. The classification results with high accuracy were obtained. It was also emphasized that if the texture features were able to accurately represent the tumor information, it is very important to perform preprocessing. Cheng et al. conducted research on deep learning-based CAD. To avoid using inaccurate image processing results to achieve tumor classification, the team used SDAE for US breast cancer, pulmonary nodules in CT scans classification. Relying on the advantages of SDAE well equipped with the automatic feature exploration mechanism and noise tolerance advantage, the classification result can reach 82.4% accuracy. Cheng et al. reported that the research is the first CAD study based on deep learning, which can further study other image features for more accurate classification results [26]. Overall, it is important to do image preprocessing to highlight the image features and further give these features for deep learning models.

In general, CNN is an end-to-end architecture and suitable for various image classification tasks. However, it is relatively hard to find the useful and meaningful features in a constrained-scale training set. In breast sonogram classification task, the training set is usually smaller than the size of the general training images. For example, the number of training images in ILSVRC12 [19] is 1.2 million, while that of our training set is 170. Instead, the autoencoder families deal with feature extraction task as a data reconstruction task. By minimizing the reconstruction error, the primary feature of the training images should be detectable. In this study, therefore, we tend to adopt the autoencoder network to extract the useful features and follow by concatenating a classifier to distinguish the type of breast sonogram. SDAE [27] is one of the state-of-the-art autoencoder architectures. Xing et al. adopted SDAE for the extraction of tumor features and classification of benign and malignant tumors [27]. SDAE is about to find out many typical patterns in the input training data, to deal with the high change problem of tumor margin, and moreover, it is provided with multilayer training automatic feature extraction and noise reduction ability so that the problems likely to occur under traditional methods may be avoided.

The accuracy of classifier is highly related to the quality of test images. As for ultrasonic image, due to the imaging principle and the properties of object to be tested, the research scholars are often encountered with two major problems, much noise and intensity inhomogeneity. Noise-resistant function has been added to SDAE when it is designed, to reduce the noise effect. In terms of intensity inhomogeneity, studies have found that after the preprocess of intensity inhomogeneity correction [28–31] of medical images, the imaging quality is effectively improved so as to reach a better image processing result. There are several intensity inhomogeneity correction methods. For example, the filter method [32, 33] is to take intensity inhomogeneity as low frequency signal and eliminate the intensity inhomogeneity with the filter method, but such a method has the elimination risk of important low-frequency information; the surface fitting method [34] is generally used to establish a curve with polynomial or spline, and the characteristics in the image are fit on such a curve to serve as image intensity inhomogeneity; the histogram method is to take the observed image as the convolution of original image and intensity inhomogeneity, and the original image can be obtained only through deconvolution; the image segmentation method [35, 36] is able to get the segmentation result

and estimated intensity inhomogeneity; or the method combining with many concepts above is adopted [37]. In this study, intensity inhomogeneity correction is conducted before the input of ultrasonic testing images. Then, the corrected images are input into the classifier system established on the basis of SDAE, hoping to overcome the effect of intensity inhomogeneity in ultrasonic imaging for improving the system resolution and quality.

Main aims of the paper are as follows:

(i) A classification algorithm in breast sonogram without subjective influences is proposed

(ii) Up to 85% accuracy can be obtained when deep learning was applied to the intensity inhomogeneity-corrected image

(iii) The study overcomes drawbacks of sonogram and has reliable discrimination ability

(iv) It has the potential to obtain the better classification results after image preprocessing

2. Materials and Methods

2.1. Input Data. This study was approved by the Institutional Review Boards of the Taipei Veterans General Hospital. All the experimental methods were carried out in accordance with the approved guidelines. In this study, we used two kinds of dataset. One is the data collected by our team (a database of 96 malignant and 74 benign images) and the other is the public dataset on the website, Rodrigues, Paulo Sergio (2017), "Breast Ultrasound Image," Mendeley Data, v1 (a database of 150 malignant and 100 benign images) [38].

2.2. SDAE. Artificial neural network is a kind of machine learning concept to simulate the learning of human's brain. As for each nerve cell, the nerve cell signals connected to it will influence the output of such nerve cell. The transmission of nerve signals is to add the received signals and then conduct nonlinear transformation, to get new output. The mutual connection in neural network decides the calculation method to make the actual output approach the expected output as much as possible through neural network adjustment.

The training of one neural network is to limit the output value to be equal to the input value, and this indicates that the output layer has the same quantity of nerve elements as the input layer, and the error between the two layers is used to adjust the weight of each layer. The training of autoencoder (AE) is unsupervised, so no label information is needed. Denoising autoencoder (DAE) is improved on the basis of AE. It is assumed that the input data include noise, so DAE is suitable to learn the features from the data including noise. SDAE is stacked by DAE, to obtain higher level features. The network training of SDAE is layer-wise, for it is of independent training between each DAE. As for the SDAE network after training, the decoding layer is eliminated, while the encoding layer generating features is reserved. To classify the data, logistic regression (LR) layer is added to be taken as output layer, and LR is of supervised

type, adding the expected output volume label information through backpropagation algorithm, and based on the error between actual output and expected output, the network weight between layers is fine-tuned. Therefore, the feature learning of system is the result of combination between SDAE pretraining and LR adjustment.

It is supposed that the input data are x and DAE anti-noise method is used to firstly set some input as 0 or add Gaussian noise to generate $\tilde{x}$ as shown in Figure 1. The way of thinking is to make the input not perfect in the very beginning so as to get better features after training. Therefore, a good result still can be achieved after the input is polluted. DAE encode is obtained according to one nonlinear transformation equation:

$$Z = f_\theta(w\tilde{x} + b), \tag{1}$$

where Z is the output of the hidden layer, called as feature description, w is the weight of input and hidden layers, b is the offset, and f_θ is the activation function of the hidden layer. DAE decoding and reconstruction are realized through the mapping function:

$$x' = g_{\theta'}(w' y + b'), \tag{2}$$

where x' denotes the transpose operation, x the output of DAE (the reconstruction), and w the corresponding weight. DAE may generate a depth network with many hidden layers through stacking. A standard SDAE structure includes two encoding layers and two decoding layers. In the encoding stage, the output of the first encoding layer is taken as the input of the second encoding layer. It is supposed that there are L hidden layers in the encoding stage, and we can obtain the activation function of number k encoding layer as follows:

$$Z^{(k+1)} = f_\theta(w^{k+1}Z^k + b^{k+1}), \quad k = 0, \ldots, L-1, \tag{3}$$

where Z^0 is the original input and Z^L is the feature description of the highest layer. The decoding stage is the same as the encoding stage. The output of the first decoding layer is taken as the input of the second decoding layer, and in this way, the activation function of number k encoding layer can be obtained as follows:

$$x'^{(k+1)} = g_{\theta'}\left(w^{(L-k)T} x'^k + b'^{(k+1)}\right), \quad k = 0, \ldots, L-1, \tag{4}$$

where x'^0 is the output of the last layer in the encoding stage and x'^L is the reconstruction of x. The weight of SDAE after pretraining will be the initial weight of adjustment and classification stages. Once the import and discriminative features have learned by SDAE, the next step is to find a suitable classifier to correctly detect the type of the extracted feature of the input image. Since there are many different classifiers such as linear classifier, perceptron, SVM, random forest, and so on, it is necessary to select one of them for the best performance purpose. Most of them are different from SDAE, while the perceptron is also based on the neural network. Inspired by perceptron learning, we
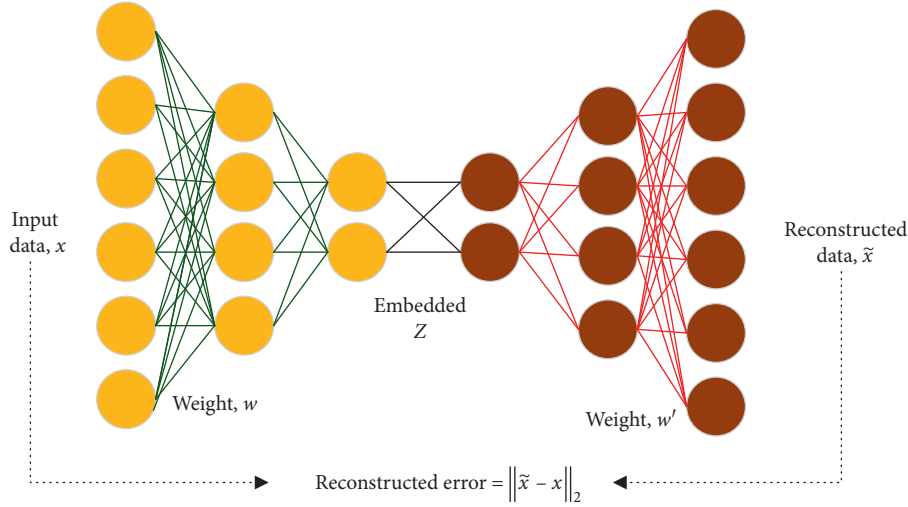
FIGURE 1: The flowchart of the reconstructed error of the DAE used in this paper.

concatenate a sigmoid layer to SDAE and formulate it as a cross-entropy loss function to keep it as an end-to-end architecture.

In the LR stage, the expected output is added, and the weight after pretraining is fine-tuned through the supervised learning method, and the activation function of LR layer is S (sigmoid):

$$S(t) = \frac{1}{1 + \exp(-t)}, \tag{5}$$

where t is the output Z^L of the final encoding layer and also the depth feature of pretraining of the SDAE method, and the output of S is the classification result, with the value ranging from 0 to 1.

The backpropagation algorithm is one kind of gradient descent methods. After the gradient direction is worked out, the weight of the classifier system will advance towards the direction with the quickest gradient descent. In addition, it is a kind of the greedy method, for it always advances towards the steepest direction, seeking for the biggest decent extent.

To realize the slight adjustment of weight, the error calculation method shall be defined. Different error calculation methods have different weight upgrading rules. Here, the common square error will be adopted:

$$E = \frac{1}{2} \sum_{d \in D} (t - y)^2, \tag{6}$$

where D indicates all input data, d indicates one of the input data, t is the expected output, and y is the actual output.

The target is to find out a set of weights, to realize the smallest calculation error:

$$w = w + \Delta w,$$
$$\Delta w = -\alpha \frac{\partial E}{\partial w}, \tag{7}$$

where α indicates the learning rate. The weight upgraded mode is the batch mode. After all inputs are executed, the connecting weight will be changed.

2.3. Intensity Inhomogeneity Correction. The algorithm was composed of two parts, constrained fuzzy cell-based bi-partition and the intensity inhomogeneity modelling [39]. When correcting the intensity inhomogeneity, the polynomial surface fitting method is adopted to estimate the image intensity inhomogeneity. Curve fitting is a kind of concept to represent the existing data through the mathematical method. Fitting is to obtain the discrete information through sampling and testing in engineering. Based on the data obtained in this way, it is hoped to obtain a continuous function (polynomial) or the gathering (spline) of many discrete equations to be identical to the data. Surface fitting is the popularization of curve fitting. It is supposed that the region of interest (ROI) image is composed of foreground (F) and background (B), and each area is regarded as the homogeneous area.

The observed gray level of image can be expressed as follows:

$$O_i = \mu_\psi + P_i + n_i, \tag{8}$$

where O_i is the gray level of number i pixel, μ_ψ is the average value of area where the pixel is located, $\psi \varepsilon \{F, B\}$ represents the area where the pixel is located, and the latter two items indicate intensity inhomogeneity and noise. Here, intensity inhomogeneity is supposed as the normal distribution of spatial changes, which is composed of a fixed variance and the average value P changing with the space, and it is represented with polynomial surface form. P_i represents the value of polynomial in number i pixel, that is, the average value of intensity inhomogeneity of number i pixel in the image, and n_i is the composition of change and noise in the polynomial surface indicating intensity inhomogeneity.

Adopt the least square method to minimize the cost function and estimate the polynomial surface as follows:

$$\varepsilon^2 = \frac{1}{N} \sum_{\forall i} \left(O_i - \mu_\psi - P_i \right)^2. \tag{9}$$

When the pixel is in the foreground area, $\mu_\psi = \mu_F$, and when the pixel is in the background area, $\mu_\psi = \mu_B$. $P_i = P(x_i, y_i)$, in which P is N time polynomial. After the polynomial surface is obtained, the image intensity will be adjusted based on results, to complete the image correction.

The image training and classification steps are shown as follows:

(1) Input training image and initialize the neural network weight by use of SDAE

(2) Remove decoding part and add LR structure, to establish SDAE-LR system

(3) Add expected input volume label and slightly adjust the network weight, to complete the classifier

We use the original images of the two different databases, i.e., private dataset and BUSIS dataset, and conduct intensity inhomogeneity correction for the original images. Therefore, each database will generate two groups of images (the original image and the corrected image) and their labels as training data. These materials are trained in the deep learning model and then performed testing to predict the benign/malignant lesions in the image. In this study, we use five different models to compare the results. Figure 2 shows the flowchart for deep learning.

3. Results and Discussion

Intensity inhomogeneity often causes unclear contour of tumors in sonogram so that it is uneasy to judge the type of tumors clinically. The correction method is proposed for improving image quality and increasing the accuracy in classification. Figure 3 shows the comparison images of one malignant tumor with/without intensity inhomogeneity. Before correction, it is uneasy to find the whole edge of the tumor. There is a missing boundary part at the left bottom of the tumor so that a wrong judgment may be caused. However, after correction, the missing part can be clearly seen, so it could be correctly judged as a malignant tumor.

Certainly, intensity inhomogeneity correction also works well in benign tumor images. Figure 4 shows the comparison images of benign tumor with/without intensity inhomogeneity. The original image with intensity inhomogeneity has unclear boundary at the right and left of the tumor that would cause wrong classification result. After correction, the unclear boundary part can be obviously distinguished so as to classify the type of the tumor correctly.

In this study, we used two kinds of dataset. One is the data collected by our team, named private dataset (96 malignant and 74 benign images), and the other is the public dataset on the website, named BUSIS dataset (150 malignant and 100 benign images). In the private dataset, 143 images and 27 images are selected randomly as training data and test data, respectively. In the BUSIS dataset, 210 images and 40 images are selected randomly as training data and test data, respectively.

First, we feed ultrasound tumor ROI images into SDAE network sized of 28×28. In the private dataset, 143 images and 27 images are selected randomly as training data and test data, respectively. In the BUSIS dataset, 210 images and 40 images are selected randomly as training data and test data, respectively. Second, a backpropagation algorithm is used on SDAE to learn the weights and the feature representation. Third, the images are labeled manually and conducted the second supervised training to classify the malignant/benign tumor images. The experiments of benign/malignant classification are grouped into two groups: (1) original images and (2) contrast-enhanced images by intensity inhomogeneity correction, which are shown in Figures 5–8, respectively.

For the private dataset, the classification accuracy values of testing results of two groups of images are, respectively, 63% and 82%. For the BUSIS dataset, the classification accuracy values of testing results of two groups of images are, respectively, 75% and 83%. The results indicate the improvement of image quality (contrast-enhanced images) actually achieve better classification performance. The results also show that because the image complexity in the private database is higher, the accuracy increases noticeably after intensity inhomogeneity correction.

As we stated previously, CNN may fail when the number of the training images is relatively small. To verify this, we also conduct an experiment of CNN for detecting benign and malignant tumors. The labeled training images are directly fed into AlexNet [19] to learn the prediction results (i.e., benign or malignant tumors). Similarly, the backpropagation algorithm is used to train the weights of AlexNet. Finally, the classification accuracy values of testing results of two groups of private dataset images are 44% and 37%, while the training accuracy values are 99.8% and 99.4%, respectively. Apparently, such CNN architecture is not suitable for such small-scale training images, leading to a serious overfitting problem. It also verified that the proposed SDAE is a proper choice for analyzing malignant/benign tumor images classification. In addition, three models (Inception v3 [40], ResNet [41], and DenseNet [42]) are also used for comparison. Table 1 shows the detection results of the original images (without correction) and the corrected images (with correction) of the private database, in which the displayed value means with/without intensity inhomogeneity correction, respectively. Table 2 shows the results of the BUSIS dataset. To sum up, because the complexity of the images in the BUSIS database is lower, the classification accuracy is higher than that of the private database before or after intensity inhomogeneity correction.

It is clear that the DenseNet achieves the best performance. However, it is also shown that the performance of the different CNN methods cannot achieve good performance, compared with SDAE approach. Although the number of the parameters of AlexNet is relatively low, the total number of parameters is still greatly large than that of SDAE. By fair comparison, we have chosen Resnet with 54 layers and Inception v3 with 34 layers to learn the classifiers with the default settings suggested by their approach. However, such lot parameters in these two networks lead to serious
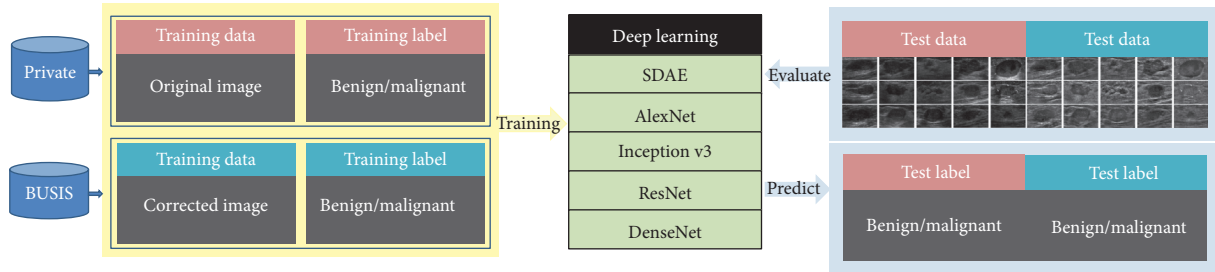
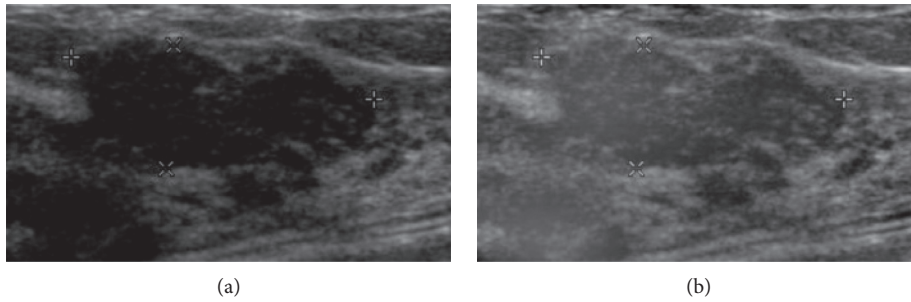Figure 2: The flowchart for deep learning.

(a)   (b)

Figure 3: Malignant: (a) original image and (b) image after correction.

(a)   (b)
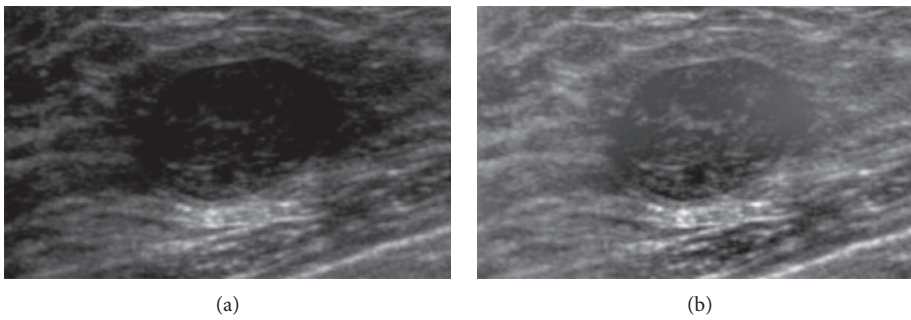
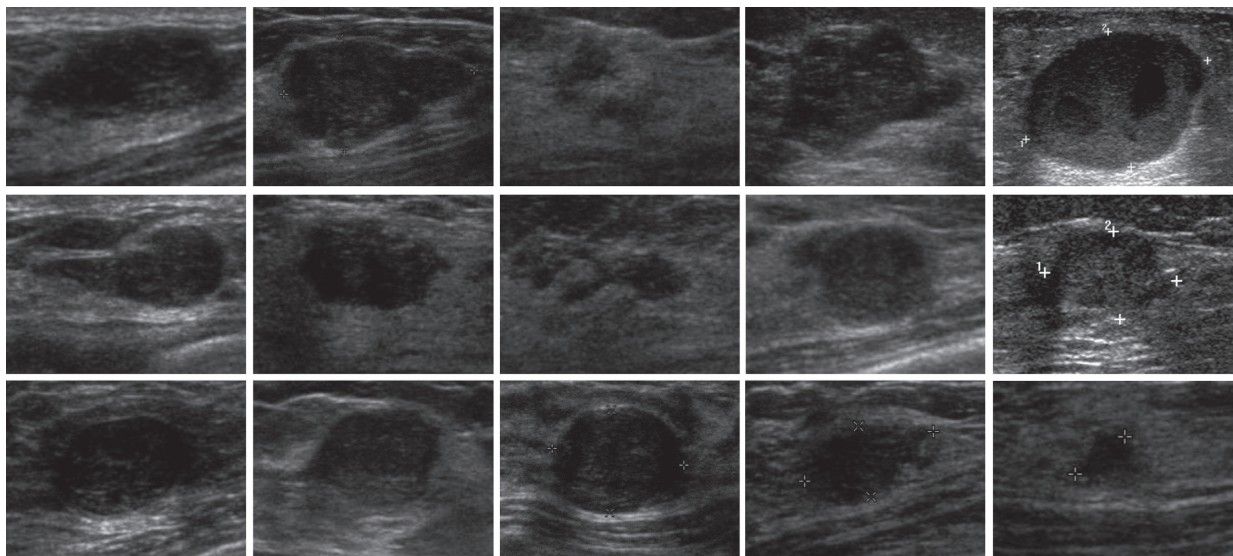Figure 4: Benign: (a) original image and (b) image after correction.

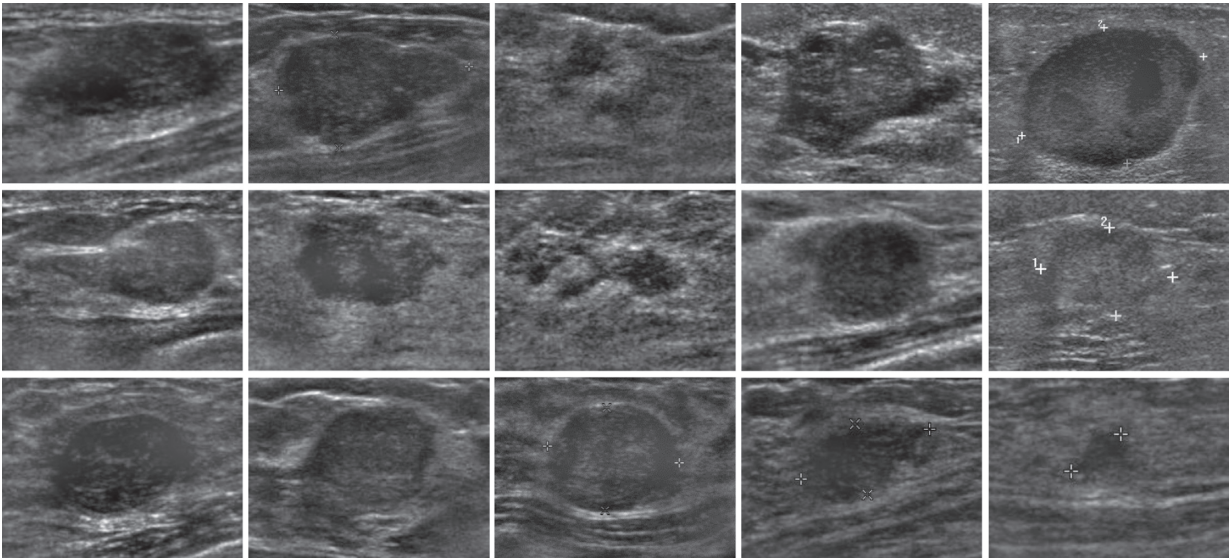Figure 5: The examples of original test images in the private dataset.

FIGURE 6: The examples of test images after correction in the private dataset.
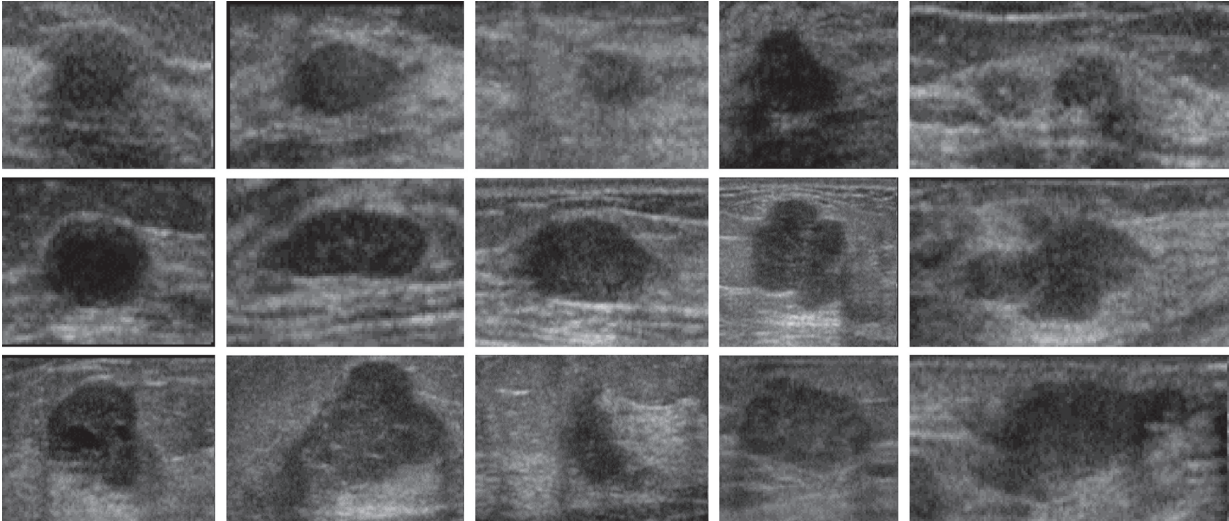

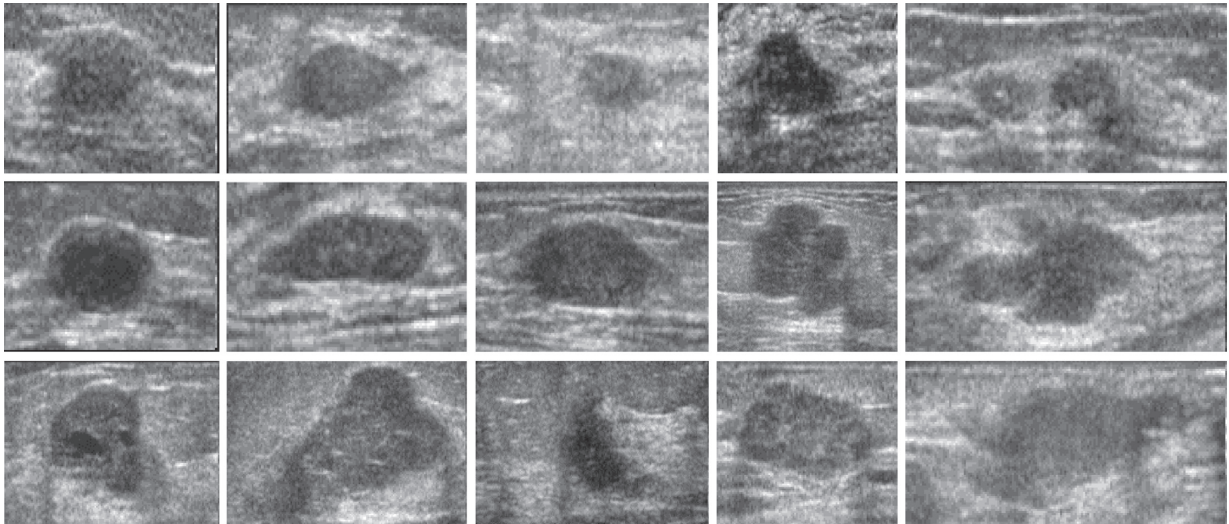FIGURE 7: The examples of original test images in the BUSIS dataset.

FIGURE 8: The examples of test images after correction in the BUSIS dataset.

TABLE 1: Classification results of the private database (the displayed value: original/corrected).

	SDAE	AlexNet	Inception v3	ResNet	DenseNet
TP	8/11	5/6	6/7	7/7	6/6
TN	9/11	7/4	6/5	6/7	8/7
FP	4/2	7/10	8/9	8/7	6/7
FN	6/3	8/7	7/6	6/6	7/7
Precision	0.67/0.85	0.42/0.38	0.43/0.44	0.5/0.5	0.5/0.46
Recall	0.57/0.79	0.38/0.46	0.46/0.54	0.46/0.54	0.46/0.46
Specificity	0.69/0.85	0.5/0.28	0.43/0.36	0.57/0.5	0.57/0.5
Accuracy	0.63/0.82	0.44/0.37	0.44/0.44	0.51/0.52	0.52/0.48
F-measure	0.62/0.85	0.4/0.41	0.44/0.48	0.48/0.52	0.48/0.46

TABLE 2: Classification results of the BUSIS database (the displayed value: original/corrected).

	SDAE	AlexNet	Inception v3	ResNet	DenseNet
TP	17/18	10/16	13/15	16/16	15/17
TN	13/15	12/14	14/16	13/10	14/13
FP	7/5	8/6	6/4	7/10	6/7
FN	3/2	10/4	7/5	4/4	5/3
Precision	0.71/0/78	0.56/0.73	0.68/0.79	0.69/0.61	0.71/0.71
Recall	0.85/0.9	0.5/0.8	0.65/0.75	0.8/0.8	0.75/0.85
Specificity	0.65/0.75	0.6/0.7	0.7/0.8	0.65/0.5	0.7/0.65
Accuracy	0.75/0.83	0.55/0.75	0.68/0.78	0.73/0.65	0.73/0.75
F-measure	0.77/0.83	0.53/0.76	0.67/0.77	0.74/0.70	0.73/0.77

overfitting problem and bad test performance. On the contrary, the most advanced network architecture, DenseNet, achieves good performance among four CNN methods; it still fails to obtain promising performance at all.

In [43], this study discriminates benign cysts from malignant masses in breast ultrasound by transferred deep neural network, Inception V3, with an accuracy of 89.44%. However, this study uses a total of 2058 breast ultrasound masses, comprising 1370 benign and 688 malignant lesions, so better results can be obtained. In [44], Han et al. use GoogLeNet with preprocessing for supporting the classification of breast lesions in ultrasound images. The networks showed an accuracy of about 90%. This study uses large data, 7408 images (4254 benign and 3154 malignant lesions), to get a good classification result. Apparently, CNN architecture is not suitable for such small-scale training images, leading to a serious overfitting problem. It also verified that the proposed SDAE is a proper choice for analyzing malignant/benign tumor images classification. In addition, Han et al. also proposed that the preprocessing can get better classification results.

4. Conclusion

This study reports that the proposed algorithm overcomes the problem of intensity inhomogeneity in the sonogram, and in combination with the deep learning method for tumor classification. This study compared five deep learning models, and SDAE achieved the best identification accuracy. Since this is a small amount of data, SDAE is a suitable choice for analyzing the small dataset.

For the original image group and the corrected image group in the private database, the accuracy values of identifying the benign and malignant tumors are 63% and 82%, respectively. For the same group in the public database, the accuracy values are 75% and 83%, respectively. The results refer that increasing the image quality will help improve the accuracy of classification. The image complexity in the private database is higher. From the results, it can be inferred that for images with high complexity, the classification accuracy increases noticeably after image preprocessing.

This work is different from the traditional algorithm to classify the type of lesions, i.e., benign or malignant in the sonogram. In the traditional way, the tumor contours must be delineated first by the segmentation technique before they can be classified. However, many segmentation algorithms need to adjust many parameters or depend on the developer's subjective ideas. In this way, the result of the segmentation in turn affects the classification result. The proposed method reduced the impact of segmentation processing steps to lead an objective classification result via deep learning. Moreover, this paper combines an intensity inhomogeneity correction method to make the classification result better. This way is advantageous, and it has the potential to obtain the better classification results after image preprocessing.

Abbreviations

AE: Autoencoder
DAE: Denoising autoencoder
SDAE: Stack denoising autoencoder
LR: Logistic regression
ROI: Region of interest.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Chia-Yen Lee and Chih-Chung Hsu conceptualized the study. Chia-Yen Lee was responsible for funding acquisition. Chia-Yen Lee, Chih-Chung Hsu, and Guan-Lin Chen performed the methodology. Chia-Yen Lee and Yi-Hong Chou administered the project. Yi-Hong Chou collected the

resources; Guan-Lin Chen performed software simulation. Chia-Yen Lee and Chih-Chung Hsu supervised the study. Chia-Yen Lee wrote the original draft. Chih-Chung Hsu reviewed and edited the manuscript.

Acknowledgments

References

[1] D. M. Parkin, P. Pisani, and J. Ferlay, "Estimates of the worldwide incidence of eighteen major cancers in 1985," *International Journal of Cancer*, vol. 54, no. 4, pp. 594–606, 1993.

[2] A. C. Society, "Breast cancer facts & figures 1997-1998," 1999, https://www.nrc.gov/docs/ml0716/ml071640135.pdf.

[3] V. Chalana, D. T. Linker, D. R. Haynor, and Y. Kim, "A multiple active contour model for cardiac boundary detection on echocardiographic sequences," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 290–298, 1996.

[4] D. R. Chen, R. F. Chang, W. J. Wu, W. K. Moon, and W. L. Wu, "3-D breast ultrasound segmentation using active contour model," *Ultrasound in Medicine & Biology*, vol. 29, no. 7, pp. 1017–1026, 2003.

[5] B. Cigale and D. Zazula, "Segmentation of ovarian ultrasound images using cellular neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 4, pp. 563–581, 2004.

[6] J. J. Clark, "Authenticating edges produced by zero-crossing algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 43–57, 1989.

[7] M. S. Drew, J. Wei, and Z. N. Li, "Illumination-invariant image retrieval and video segmentation," *Pattern Recognition*, vol. 32, no. 8, pp. 1369–1388, 1999.

[8] P. R. Hill, C. N. Canagarajah, and D. R. Bull, "Texture gradient based watershed segmentation," in *Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, p. IV-3381, IEEE, Orlando, FL, USA, May 2002.

[9] S. H. Lee and J. K. Seo, "Level set-based bimodal segmentation with stationary global minimum," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2843–2852, 2006.

[10] T. Brox and J. Weickert, "Level set segmentation with multiple regions," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3213–3218, 2005.

[11] J. F. Haddon and J. F. Boyce, "Image segmentation by unifying region and boundary information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 929–948, 1990.

[12] S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1292–1300, 2004.

[13] J. Z. Cheng, Y. H. Chou, C. S. Huang et al., "ACCOMP: augmented cell competition algorithm for breast lesion demarcation in sonography," *Medical Physics*, vol. 37, no. 12, pp. 6240–6252, 2010.

[14] R. F. Chang, W. J. Wu, W. K. Moon, and D. R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," *Breast Cancer Research and Treatment*, vol. 89, no. 2, pp. 179–185, 2005.

[15] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8595–8598, IEEE, Vancouver, Canada, May 2013.

[16] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2518–2525, IEEE, Providence, RI, USA, June 2012.

[17] H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1930–1943, 2012.

[18] Y. Xu, T. Mo, Q. Feng et al., "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1626–1630, IEEE, Florence, Italy, May 2014.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, MIT Press, Cambridge, MA, USA, 2012.

[20] M. H. Yap, G. Pons, J. Martí et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.

[21] X. Xie, F. Shi, J. Niu, and X. Tang, "Breast ultrasound image classification and segmentation using convolutional neural networks," in *Proceedings of Pacific Rim Conference on Multimedia*, pp. 200–211, Springer, Hefei, China, September 2018.

[22] S. Zhou, J. Shi, J. Zhu, Y. Cai, and R. Wang, "Shearlet-based texture feature extraction for classification of breast tumor in ultrasound image," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 688–696, 2013.

[23] W. K. Moon, I. L. Chen, J. M. Chang, S. U. Shin, C. M. Lo, and R. F. Chang, "The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound," *Ultrasonics*, vol. 76, pp. 70–77, 2017.

[24] P. Raha, R. V. Menon, and I. Chakrabarti, "Fully automated computer aided diagnosis system for classification of breast mass from ultrasound images," in *Proceedings of 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 48–51, IEEE, Chennai, India, March 2017.

[25] M. Abdel-Nasser, J. Melendez, A. Moreno, O. A. Omer, and D. Puig, "Breast tumor classification in ultrasound images using texture analysis and super-resolution methods," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 84–92, 2017.

[26] J. Z. Cheng, D. Ni, Y. H. Chou et al., "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans," *Scientific Reports*, vol. 6, no. 1, p. 24454, 2016.

[27] C. Xing, L. Ma, and X. Yang, "Stacked denoise autoencoder based feature extraction and classification for hyperspectral images," *Journal of Sensors*, vol. 2016, Article ID 3632943, 10 pages, 2016.

[28] B. M. Dawant, A. P. Zijdenbos, and R. A. Margolin, "Correction of intensity variations in MR images for computer-aided tissue classification," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 770–781, 1993.

[29] P. Vemuri, E. G. Kholmovski, D. L. Parker, and B. E. Chapman, "Coil sensitivity estimation for optimal SNR reconstruction and intensity inhomogeneity correction in phased array MR imaging," in *Proceedings of Biennial International Conference on Information Processing in Medical Imaging*, pp. 603–614, Springer, Glenwood Springs, CO, USA, July 2005.

[30] G. Gerig, M. Prastawa, W. Lin, and J. Gilmore, "Assessing early brain development in neonates by segmentation of high-resolution 3T MRI," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 979-980, Springer, Montréal, Canada, November 2003.

[31] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.

[32] E. Ardizzone, R. Pirrone, and O. Gambino, "Exponential entropy driven HUM on knee MR images," in *Proceedings of 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005*, pp. 1769–1772, IEEE, Shanghai, China, September 2005.

[33] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.

[34] J. A. Hernández, M. L. Mora, E. Schiavi, and P. Toharia, "RF inhomogeneity correction algorithm in magnetic resonance imaging," in *Proceedings of International Symposium on Biological and Medical Data Analysis*, pp. 1–8, Springer, Barcelona, Spain, November 2004.

[35] A. W. C. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Transactions on Medical Imaging*, vol. 22, no. 9, pp. 1063–1075, 2003.

[36] C. Li, R. Huang, Z. Ding, C. Gatenby, D. Metaxas, and J. Gore, "A variational level set approach to segmentation and bias correction of images with intensity inhomogeneity," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 1083–1091, Springer, New York, NY, USA, September 2008.

[37] J. Milles, Y. M. Zhu, G. Gimenez, C. R. Guttmann, and I. E. Magnin, "MRI intensity nonuniformity correction using simultaneously spatial and gray-level histogram information," *Computerized Medical Imaging and Graphics*, vol. 31, no. 2, pp. 81–90, 2007.

[38] M. Xian, Y. Zhang, H. D. Cheng et al., "A benchmark for breast ultrasound image segmentation (BUSIS)," 2018, http://arxiv.org/abs/1801.03182.

[39] C. Y. Lee, Y. H. Chou, C. S. Huang, Y. C. Chang, C. M. Tiu, and C. M. Chen, "Intensity inhomogeneity correction for the breast sonogram: constrained fuzzy cell-based bipartitioning and polynomial surface modeling," *Medical Physics*, vol. 37, no. 11, pp. 5645–5654, 2010.

[40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[42] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, no. 2, p. 3, Honolulu, HI, USA, July 2017.

[43] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, "Comparison of transferred deep neural networks in ultrasonic breast masses discrimination," *BioMed Research International*, vol. 2018, Article ID 4605191, 9 pages, 2018.

[44] S. Han, H. K. Kang, J. Y. Jeong et al., "A deep learning framework for supporting the classification of breast lesions in ultrasound images," *Physics in Medicine & Biology*, vol. 62, no. 19, pp. 7714–7728, 2017.