Automatic Detection of Breast Calcification in Ultrasound Imaging with Convolutional Neural Network

P D Karunia¹, P Prajitno and D S Soejoko
Department of Physics, Universitas Indonesia
Kampus UI, Depok 16424, West Java, Indonesia
Email: ¹petrananda.dea@ui.ac.id

Abstract. Breast cancer is a common type of cancer that leading death causes of female in the worldwide. Breast calcification can be one of indicator that can be used to detect the breast cancer early. One of the preferred methods used by radiologist to detect breast cancer is ultrasound imaging. Ultrasound imaging is much safer than mammography that followed by radiological effect. However, ultrasound imaging contaminated with speckle noise that looks similar to breast calcification. It can be the cause of the long time diagnosis process. It encourages so many methods of computed aided diagnosis (CADx) that can detect abnormalities automatically. One of them is Convolutional Neural Network (CNN). CNN can be used to classify the normal breast and breast with abnormalities. In this paper, CNN has been proposed for the classification of the ultrasound images into normal breasts and breasts with calcification. Experimental results classification accuracy was 76% and a sensitivity of 84.61%.

1. Introduction
Breast cancer is a common cancer that cause of death for women in the world. However, with early detection, a small period will provide a greater chance of survival. Based on data from the American Cancer Society, within 5 years after diagnosis, patients with early stage cancer diagnosis have a 100% chance of survival. Whereas in stage two has a chance of survival up to 80-90%, stage three has a chance of survival up to 60% -70%, and for the final stage the chance of survival is only about 15% - 20%. Breast cancer detection can be done by mammography or ultrasound imaging [1].

Ultrasound imaging is one of the most frequently used diagnostic tools to detect and classify abnormalities of the breast, because of its low cost, nonionizing radiation, and real-time capability. However, ultrasound image formation has a weakness in spatial resolution and the presence of speckle noise [2]. Breast abnormalities usually begin with microcalcification. Microcalcification is an early indicator of breast cancer that is not palpable. With breast screening, microcalcification can be categorized by looking at its density, morphological appearance, and the pattern or shape of its distribution [3].

Breast calcifications are a calcium deposits that can be majorly categorized as two types which differ in chemical composition. The first is composed of calcium oxalate, which is mainly found in benign breast lesions. The other is hydroxyapatite, a kind of calcium phosphate, which can be found in both benign and malignant [4,5]. The number, size, morphology and distribution of breast calcification gives important information about the malignancy or benignity of breast lesions. Amorphous calcifications with round, thick rod, diffuse and non-catheter distribution are indicate low risk of malignancy calcifications (Figures 1a, b). Large and heterogeneous calcifications, and amorphous or isolated calcifications are more associated with indeterminate calcifications (Figure 1c, d). While, calcifications
that indicates highly suspicious malignant show shows with fine and pleomorphic, or fine linear branching distribution calcifications (Figure 1e, f).

![Figure 1](image)

**Figure 1.** Typical distribution of breast calcification (a) Diffuse, (b) Regional, (c) Isolated, (d) Large and heterogeneous, (e) Linear branching, (f) Clustered [5].

Breast calcification detection is carried out with a human reader, and usually performed by radiologists or oncologists. It takes a long time and may add costs for adding data from other modalities. Another factor that can be a potential error is increasing the length of time for reading and adding information with other modalities is due to noise, so that the reading is not optimal [1]. Noise reduction on ultrasound images still is a serious issue. The error detection can be caused by system that detect the speckle noise as an abnormality. Hence, the filtering of speckles is of prime importance, and in recent years, several studies have been undertaken to address this serious issue [6-7].

Computer-assisted diagnosis (CADx) systems have been developed to reduce costs and increase the radiologist's ability to interpret medical images and differentiate between benign and malignant tissue. One of the method used for CADx is Convolutional Neural Network (CNN). Convolutional Neural Network (CNN) is a deep neural network that is devoted to image recognition. This technique shows how significant the change of the deep layer (node layer) is for image processing. CNN mimics the workings of our brains in processing and recognizing an object [8]. The CNN system consists of three layers of processes, namely the convolution, pooling and fully connected (FC) layers. The convolution and pooling layers are part of the system whose function is to extract image properties, which will then be mapped to the fully connected layer to become output, such as the results of object classification. The output results are evaluated with a loss function of the loss value which is used to update the kernel factor and weight. This process is called training. After the training process, the classification system in the CNN system has been formed and is expected to be able to be used to classify new images accurately [9]. The CNN system includes at least one convolutional layer. Figure 2 shows a 2-dimensional (2D) CNN used to recognize images [10].

2. Literature Review

The conventional approach to analyzing 2D ultrasound images of the breast is likely to be a time-consuming and operator-dependent procedure [11]. Various digital image processing has been developed to obtain a more efficient image analysis method [12-13]. In recent years, several machine learning-based methodologies have been developed, one of which is Convolutional Neural Networks (CNNs). The CNN has also been included in medical image analysis to aid in patient diagnosis [14-15].
This system performs predictions for each pixel value and assign a class label to each pixel in an image [13].

![2D CNN](image)

**Figure 2.** 2D CNN [8]

Recently, CNN has developed due to its high accuracy, great power and flexibility. This feature has been applied to the classification of microcalcification clusters or breast lesions. Becker et al. [16] demonstrated that the cutting-edge network for general image analysis could detect cancer in mammography with an accuracy similar to that of radiologists. Breast cancer detection and breast calcification based on mammography images have also been developed in several studies [17, 18]. Chai at all [17] obtained detection results with a precision of 89.32% and a sensitivity of 86.89%. Meanwhile, analysis based on ultrasound images is carried out by several methods, for example, in [3], analysis is carried out by image segmentation. In this paper, we will use CNN to detect breast calcification on ultrasound imaging based on Latif 2020 [19], with 2 convolutional layers and 2 pooling layers with differences in sample size.

Noise reduction in this paper obtained by mean filter. Mean filter is a linear filtering technique that is also known as an averaging filter, which replaces each value of the pixel in an image of the average the gray levels of the neighbourhood. The technique does not eliminate speckle noise as a whole, but it decreases to some extent. The filter has an effect of blurring and smoothing the image [20].

### 3. Materials

#### 3.1. Data Set

The data used in this study are ultrasound image data obtained from RSUD Dr. Moewardi. The data taken are breast images between August 2019 and August 2020. The images used are limited to patients aged between 20-60 years, with an average age of 46 years. The ultrasound examination process is carried out by an experienced radiologist. The data used in this experiment are ultrasound images with a diagnosis of breast cancer with calcification components and normal breasts. Calcifications are observed as echogenic dots within a mass or lesion or may be distributed as we can see in Figure 1. In this study, the breast images used included 97 images which contained about 48 normal breast images (negative sample), and 49 calcified breast images (positive sample). The data will be used as sample data for the data training, data validation and data testing. The positive and negative sample showed at Figure 3.

For testing purposes, 25 sample data were taken, which included 15 positive samples and 10 negative samples. And other data will be used as sample data in the CNN training and validation process. Due to the limited amount of data, for the training and validation process, the data is multiplied by data augmentation. 38 negative samples and 34 positive samples were multiplied by applying 4 kinds of image transformations, which is rotated (90° and 180°) and flipped the image up down and left right. The total data obtained from the CNN training process is 360 images. Furthermore, for the training and validation processes, the data will be split, 80% in the training process and 20% for the validation.
3.2. Preprocessing

The first step in this research is the preprocessing. This step has been done before the image was processed in the training process. The data obtained are intact data captured on the ultrasound examination. In one chapter, an image can consist of 2 to 4 Region of interest (ROI), so each image has a different size. Therefore, it is necessary to equalize the size of the data and crop it to focusing on positive samples. For positive sample data, cropping was carried out which focused on the calcification area. Furthermore, it will be a positive sample in the CNN training process.

The same treatment was carried out for the negative samples. Negative samples were taken randomly from the image with a normal breast diagnosis. The same size was also applied to the negative sample. This size uniformity was done because the CNN system can only accept samples of the same size. The initial image which is in Dicom format, after the cropping and resizing process was saved in thief to keep a good quality. Next, the image was corrected by normalizing and filtering with mean filter. The purpose of this stage is to reduce speckle noise that can reduce the performance of the CNN. Mean filter used has a 3x3 matrix size. And the result of this process will be evaluated with MSE and PSNR parameters.

\[
MSE = \frac{1}{MN} \sum_{j=0}^{m} \sum_{j=0}^{n} [I(i,j) - I'(i,j)]^2
\]

(1)

With MN is total pixel, matrix size mxn, I (I, j) am the original image, and I’(I, j) is filtered image. And the PSNR was described in Equation (2).

\[
PSNR = 20 \log_{10} \left( \frac{255}{\sqrt{MSE}} \right)
\]

(2)

3.3. Training CNN

At this stage, the CNN model is developed to classify the breast ultrasound image automatically, to classify the image into normal or calcification. The CNN used in this study refers to Latif, 2020, with differences in the size of the input image. The CNN developed is composed of an input layer, two convolutional layers, two pooling layers and a fully connected layer. The input layer has a size of 64x64, the two convolutional layers have a window size of 3x3, and the pooling layer was 2x2. The convolutional layers have several map features and makes use of variable weights. They are responsible for the extraction of various features of each part of the input image. To perform the same feature detection in all possible locations of the input image, all units in one feature map must have the same weight and bias. As a result, each feature map attempts to perform a different local feature detection.

The role of pooling layers is to present information on convolutional layer outputs in a concise and simplified form. With such a pooling layer setting (2x2), the output data size will be half times the input size. The connected output layer is entirely responsible for generating the results of the classification procedure. Training scheme (optimizer) used to update the kernel weight values is Stochastic Gradient
Descent (SGD). SGD is a scheme that is commonly used in the CNN training process. The training process using the SGD scheme is faster than the batch scheme because in the SGD scheme the kernel weight values are updated every time one training data is evaluated. The weights are updated by using an optimization scheme based on stochastic gradient descent, during the training phase [22].

$$W_k = W_{k-1} - \eta \frac{\partial E^m(w)}{\partial W}$$  \hspace{1cm} (3)

Where E (W) is the cost function, m is the sample size and η is the step size. In order to deal with the issue of overfitting, a dropping out scheme has been used [28] which tends to ignore the randomly selected neurons while training. This technique leads to a generalization of CNNs, by aiding in learning multiple internal representations. Furthermore, to counter the issue relating to overflow, a minor weight penalty is incorporated during training.

3.4. Evaluation

After the training process, the output of the system has been tested with a new image. The CNN system that has been compiled is used to classify new images with a normal or calcified diagnosis. The data used at this stage consisted of 25 images consist of 10 normal data and 15 images of calcified breasts. The system will automatically classify images that are included in the category of normal or calcified images, according to the CNN system that has been formed. We evaluate the number of true positives, false negatives, false positives and true negatives. True positive (TP) is positive data that are clarified as true by the system, in this case is a calcification sample that is classified correctly by the system. False negative (FN) is the number of negative data, but incorrectly clarified by the system. FN in this study is the amount of positive data (calcification) which is classified as a normal image. False positive (FP) is negative data that are correctly detected, meaning that the normal image data is classified correctly. While true negative (TN) is negative data that is correctly detected by the system, normal images are classified correctly.

Based on the obtained TP, FN, FP and TN values, the CNN system accuracy and sensitivity were calculated. Accurately represents the percentage of normal and abnormal structures that can be classified correctly. While the sensitivity is the percentage of abnormalities that can be classified by the system.

The amount of sensitivity, accuracy that can be described in Eq. (4) And (5).

$$Accuracy = \frac{TP + TN}{TP + FP + FN} \times 100\%$$ \hspace{1cm} (4)

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$ \hspace{1cm} (5)

4. Experimental Result

Positive and negative samples obtained at the preprocessing stage of the image were uniformed with a size of 64x64. This size was chosen because it is the optimum size based on observations that have been made. After the image sizes are equalized, the images are stored in two different folders, according to the classification based on the doctor's diagnosis. The naming of this folder will later become the image label in the CNN training process. The CNN system will classify each image according to the label. The CNN system hierarchy formed includes an input layer (64x64 size), convolutional layer I (32x64x64 size), pooling layer-I (32x32x32 size), convolutional layer II (64x32x32 size), pooling layer II (64x16x16 size) and the output layer respectively.

The research was conducted using unfiltered image data (original data) and filtered data. The averaging filter was chosen because this filter is the simplest function and ethics applied in 2D grayscale images can provide an angle effect so that a clearer image is obtained. With better sample image quality, it is hoped that the performance of the CNN system that will be formed will also be better.

In the filtration process, the parameters used to analyze the similarity of the image are Mean Square Error (MSE) and Peak-Signal to Noise Ratio (PSNR). The processed image can be said to be similar to the original image if the MSE value is close to 0, and the PSNR value is more than 30 dB. In this study,
the smallest MSE was 0.65 and the highest PSNR was 39.5 dB. The results of this filtration are then used as a CNN training sample. Table 1 shows the results of CNN training for each datum, both filtered and original data. The CNN system classifies images into normal breast images and calcified breast images. For both data, experiments were carried out with epochs of 25, 50, 75, 100, 125 and 150. The epochs will be stopped if the error is constant or increases with increasing epochs. Based on the result, it can be seen that the smallest loss value is achieved at epoch 125 and starts to be constant for epoch 125 and 150. Therefore, the model at these epochs will be stored and used in the testing process.

For unfiltered data, the test validation accuracy achieved for the epoch 125 is 76.04, and 91.67 for the filtered data. The CNN will be tested with new data, to testing the performance of the CNN for classifying ultrasound breast images into normal and calcification. For the unfiltered data sample, classification testing accuracy was 68% and a sensitivity of 73.33%. While, for the filtered data sample, classification testing accuracy was 76 % and 84.61% for the sensitivity.

| Data Result | Test Validation Result |
|-------------|------------------------|
|             | Epoch | Loss  | Accuracy |
| Unfiltered  | 25    | 0.7722| 33.44     |
| Sample      | 50    | 0.5174| 46.79     |
|             | 75    | 0.5097| 76.04     |
|             | 100   | 0.5095| 76.04     |
|             | 125   | 0.5094| 76.04     |
|             | 150   | 0.5094| 76.04     |
| Filtered    | 25    | 0.6724| 46.88     |
| Sample      | 50    | 0.4856| 75.79     |
|             | 75    | 0.4688| 91.67     |
|             | 100   | 0.4682| 91.67     |
|             | 125   | 0.4681| 91.67     |
|             | 150   | 0.4681| 91.67     |

From the experimental result, for the filtered data sample using a mean filter, accuracy increased to 76% for accuracy, and to 84.61% for sensitivity, which is increased for about 9% of the unfiltered data sample. This shows that the filtering process provides better results and effective to improve the performance of the CNN classification system. The better image sample quality, the better performance of CNN classification system.

5. Conclusion
In this research, A CNN models have been proposed for classifying of the Breast Ultrasound Images into either normal or calcification breast. For the filtered and unfiltered data samples, a 76% classification accuracy and 84.61% classification, sensitivity has been achieved with the proposed model with filtered data samples using a mean filter. In addition, the segmentation of breast calcification will be investigated in future studies.

References
[1] Yousef, W.A., Abouelkahire, A.A., Almahallawi, D., Marzouk, S.O., Mohamed, S.K., Mustafa, W.A., Osama, O.M., Saleh, A.A., Abdelrazek, N.M. arXiv:1908.10661v1 [eess. IV] 26 Aug 2019. Do: //arxiv.org/abs/1908.10661.
[2] Hiremath, P. S., Prema T. Akkasaligar, and Sharan Badiger. In Advancements and breakthroughs in ultrasound imaging. Intechopen, 2013.
[3] Chang RF, Hou YL, Huang CS, Chen JH, Chang JM, Moon WK. Automatic detection of microcalcifications in breast ultrasound. Med Phys. 2013 Oct;40(10):102901.
[4] Liu, Shengfeng, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang. Engineering (2019).
[5] Ouyang, Y. L., Zhou, Z. H., Wu, W. W., Tian, J., Xu, F., Wu, S. C. et al. A review of ultrasound detection methods for breast microcalcification. Math. Biosci. Eng. 16, 1761–1785 (2019)

[6] B. Peng, Y. Xian and J. Jiang, 2018 IEEE International Ultrasonics Symposium (IUS), Kobe, 2018, pp. 206-212

[7] Jeng, Geng-Shi, Maria Zontak, Nripesh Parajuli, Allen Lu, Kevinminh Ta, Albert J. Sinusas, James S. Duncan, and Matthew O’Donnell. IEEE Access 6 (2018): 17415-17428

[8] Chiang, Tsung-Chen, Yao-Sian Huang, Rong-Tai Chen, Chiu-Sheng Huang, and Ruey-Feng Chang. IEEE transactions on medical imaging 38, no. 1 (2018): 240-249.

[9] Yamashita, R., Nishio, M., Kinh, R., Do, G., dan Togashi, K. (2018): Convolutional neural networks : an overview and application in radiology, Insights into Imaging, 9, 611–629

[10] Kim, P. (2017). MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence.

[11] T. M. Kolb, J. Lichy, J. H. Newhouse, Radiology, vol. 225, no. 1, pp. 165-175, Oct. 2002.

[12] Anderson ME, Soo MS, Bentley RC, Trahey GE. The detection of breast microcalcifications with medical ultrasound. J Acoust Soc Am. 1997 Jan;101(1):29-39. doi: 10.1121/1.417973. PMID: 9000730.

[13] Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. AJR Am J Roentgenol. 2015 Feb;204(2):234-40. doi: 10.2214/AJR.13.12072. PMID: 25615743

[14] Latif, G., Iskandar, D. A., Alghazo, J., Butt, M., & Khan, A. H. (2018). Deep CNN based MR image denoising for tumor segmentation using watershed transform. International Journal of Engineering & Technology, 7(2.3), 37-42.

[15] Pinheiro, P. and Collobert, R., (2014), In International Conference on Machine Learning, pp. 82-90.

[16] A. S. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, Investigative Radiology, vol. 52, no. 7, pp. 434–440, 2017

[17] Hongmin Cai, Qinjian Huang, Wentao Rong, Yan Song, Jiao Li, Jinhua Wang, Jiazhou Chen, Li Li, Computational and Mathematical Methods in Medicine, vol. 2019, Article ID 2717454, 10 pages, 2019. https://doi.org/10.1155/2019/2717454

[18] A. Papadopoulos, D.I. Fotiadis, A. Likas,An automatic microcalcification detection system based on a hybrid neural network classifier,Artificial Intelligence in Medicine, Volume 25, Issue 2, 2002, Pages 149-167,ISSN 0933-3657, https://doi.org/10.1016/S0933-3657(02)00013-1.

[19] G. Latif, M.O. Butt, F.Yousif Al Anezi and J.Alghazo. 2020. RIVF International Conference on Computing and Comunication Tecnologies (RIVF), 2020, pp.1-5, doi:10.1109/RIVF48685.2020.9140767.

[20] Ahmed S.Bafaraj. 2019. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 6 (2019) pp. 1340-1351 © Research India Publications. http://www.ripublication.com

[21] X`, Ghazanfar, Jaafar Alghazo, Loay Alzubaidi, M. Muzzamal Naseer, and Yazan Alghazo. In 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), pp. 90-95. IEEE, 2018.

[22] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958