

# Breast Mass Classification using the Fundamental Deep Learning Approach: To build the optimal model applying various methods that influence the performance of CNN

Jin Lee<sup>1,2</sup>, Kwang Jong Choi<sup>3</sup>, Seong Jung Kim<sup>2</sup>, Ji Eun Oh<sup>2</sup>, Woong Bae Yoon<sup>2</sup>, Kwang Gi Kim<sup>2,\*</sup>

## Abstract

Deep learning enables machines to have perception and can potentially outperform humans in the medical field. It can save a lot of time and reduce human error by detecting certain patterns from medical images without being trained. The main goal of this paper is to build the optimal model for breast mass classification by applying various methods that influence the performance of Convolutional Neural Network (CNN). Google's newly developed software library Tensorflow was used to build CNN and the mammogram dataset used in this study was obtained from 340 breast cancer cases. The best classification performance we achieved was an accuracy of 0.887, sensitivity of 0.903, and specificity of 0.869 for normal tissue versus malignant mass classification with augmented data, more convolutional filters, and ADAM optimizer. A limitation of this method, however, was that it only considered malignant masses which are relatively easier to classify than benign masses. Therefore, further studies are required in order to properly classify any given data for medical uses.

**Key Words:** Breast mass, Classification, Deep learning, Tensorflow

## I. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer and the leading cause of cancer death among women worldwide, accounting for approximately 1.7 million cases and 521,900 deaths in 2012 [1]. The American Cancer Society, in turn, recommends women over the age of 40 to get breast cancer screening mammograms on a regular basis for the purpose of early detection [2]. However, it is difficult for radiologists to detect and analyze masses due to their variation in shape, size, and boundary as well as their low signal to noise ratio, resulting in unnecessary biopsies or missed masses [3].

A computer-aided diagnosis (CAD) system has been traditionally used in breast mass classification. However,

according to one study on the effectiveness of CAD, it shows no significant improvements in the sensitivity for invasive breast cancer [4]. To solve this problem, convolutional neural networks (CNNs) based on deep learning approaches are being developed by many researchers to be used in clinical practice. Unlike traditional CAD systems that use pre-determined features, CNNs determine the most relevant features from data in order to classify images as normal tissue or malignant masses [5].

A CNN commonly includes the convolutional layers, the pooling layers, followed by fully connected layers. The convolutional layers consist of a set of learnable filters that are convolved with the input image. The pooling layers reduce the size of the input and max-pooling is

---

Manuscript received Sept. 1, 2016; Revised Sept. 21, 2016; Accepted Octo. 10, 2016. (ID No. JMIS-2016-0013)

Corresponding Author (\*): Biomedical Engineering Branch, Precision Medicine and Cancer Informatics, Research Institute, National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do, Korea, Tel: +82-31-920-2241, E-mail: kimkg@ncc.re.kr.

<sup>1</sup> Bachelor of Arts degree in biology from Taylor University, Upland, Indiana, United States.

<sup>2</sup> Biomedical Engineering Branch, Division of Precision Medicine and Cancer Informatics, National Cancer Center, Goyang, Korea.

<sup>3</sup> Korea Christian International School, Goyang, Korea.

commonly used. Fully connected layers have full connections to all activations in the previous layer and calculate the final output with a soft-max function [6].

## II. MATERIALS AND METHODS

### 2.1 Dataset

The mammogram data set used in this study was obtained from 340 breast cancer cases. Each case included both the mediolateral oblique (MLO) and craniocaudal (CC) views of each breast, as well as 400 normal tissue and 319 malignant mass square regions of interest (ROIs). During the process, unclear data, such as images with masses all over the breast and mammograms taken while breastfeeding, were taken out. These images were then split into training, testing and validation data respectively 60%, 20% and 20%.

### 2.2 Data Augmentation

Due to the small size of our dataset, data augmentation was performed by rotating the original images  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  and flipping the same images horizontally and vertically. Since masses do not have a particular orientation [3], rotated and flipped images were recognized as different images from the original ones. Data augmentation resulted in the data size 6 times larger than the original dataset comprising 2400 normal and 1914 malignant masses.

### 2.3 CNN Architecture

A visual representation of the two layered CNN architecture used in this study is shown in Figure 3. It consists of 3 stages of convolutional layers, ReLU (rectified linear unit) activation layers, and max pooling layers, followed by fully connected layers. A dropout layer with dropout factor of 0.75 was added before the fully connected layers to prevent overfitting [7]. Examples of mammogram images at the low-level, mid-level, high-level features, and a convolved image are also shown in Figure 1.

In addition, optimum number of iterations had to be determined since a small number of iterations results in less training and too large a number of iterations results in high error rates. Observing the iterations vs. accuracy graph in Figure 2, it was confirmed that to keep increasing the number of iterations does not further increase the testing accuracies. Therefore, the number of iterations was set to 50,000 where the curve reached a plateau, and the batch size was set to 30 for all datasets.

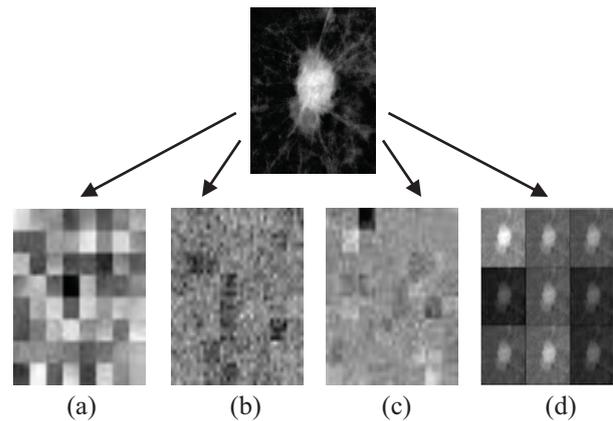


Fig. 1. Examples of mammogram images (a) Low-level, (b) mid-level, (c) high-level features, and (d) convolved image

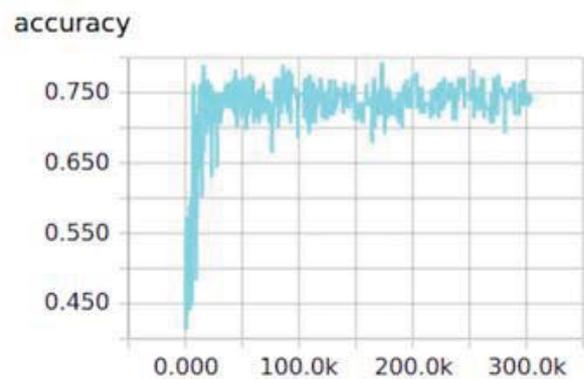


Fig. 2. A sample Iterations vs. Testing accuracies graph

### 2.4 Number of convolutional filters

Each filter or kernel in convolutional layers extracts particular features from the images. Before the number of filters was increased, the model used 32 filters in the first convolutional layer and then 64 filters in the second layer, extracting 2,048 features from one image. The number of convolutional filters was then increased to 64 and 128, to see if it enables the model to extract more features and show better performance.

### 2.5 Image sizes

Since the CNN model used in this study was modified from MNIST classification model, 28 by 28 was the default setting that could be used as input image sizes. To see if increasing image sizes enables the network to extract smaller and more detailed features and ultimately to show better performance in breast mass classification, input image sizes of 64 by 64, 128 by 128, and 256 by 256 were compared.

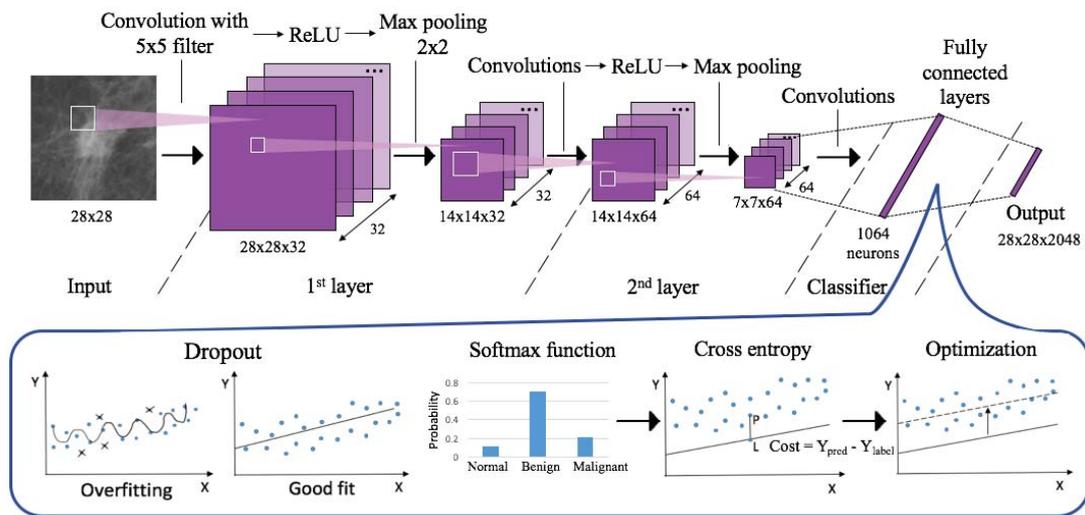


Fig. 3. Visual representation of 2 layered CNN [8]

## 2.6 Optimizer

Adam (Adaptive Moment Estimation) optimizer was applied to the model instead of RMS optimizer to see if it is more suitable for minimizing the loss function. RMS optimizer applies the same learning rate for all parameters while ADAM optimizer computes individual adaptive learning rates for different parameters. Also, it only requires first-order gradients with little memory requirement [9].

## 2.7 Measurement

For each variable, we measured accuracies, sensitivities, and specificities to evaluate the performance of the proposed CNN model. This process was repeated 10 times and the average with standard deviation were calculated for more credibility.

$$\text{Accuracy} = \frac{\# \text{ of total correct}}{\# \text{ of total prediction}}$$

$$\text{Sensitivity} = \frac{\# \text{ of correct mass}}{\# \text{ of total mass}}$$

$$\text{Specificity} = \frac{\# \text{ of correct normal}}{\# \text{ of total normal}}$$

## III. RESULTS

Mass classification was performed using 4 variables

explained above. With 400 normal tissue and 319 malignant mass ROIs, the testing accuracy was 0.78, the sensitivity was 0.88, and the specificity was 0.67. Data augmentation resulted in the data size 6 times larger than the original dataset comprising 2400 normal and 1914 malignant masses. The accuracy was 0.81 with the sensitivity of 0.80, and the specificity of 0.82. Also, a t-test was performed and the p-value was 0.19.

With 32 and 64 convolutional filters, the testing accuracy was 0.81, the sensitivity was 0.80, and the specificity was 0.82. Increasing the number of convolutional filters to 64 and 128 resulted in the accuracy of 0.86, the sensitivity of 0.84, and the specificity of 0.89. For increasing image sizes, the accuracy with 64 by 64 input image sizes was 0.86. When they were increased to 128 by 128 and 256 by 256, the accuracies also increased to 0.88 and 0.89. T-test results between image sizes 64 by 64 and 128 by 128 produced a p-value of 0.11, and a p-value of 0.61 was shown between image sizes 128 by 128 and 256 by 256.

Mass classification with a RMS optimizer resulted in the accuracy of 0.86, sensitivity of 0.84, and specificity of 0.89. When the optimizer was changed to ADAM, the accuracy was 0.89, the sensitivity was 0.90, and the specificity was 0.87. A t-test was performed and the p-value was 0.006.

In conclusion, the best classification performance showed an accuracy of 0.887, sensitivity of 0.903, and specificity of 0.869 for normal tissue versus malignant mass classification with augmented data, more convolutional filters, and ADAM optimizer.

Sample output images with the labels and predictions

are shown in figure 4 (first row: normal, second row: malignant) and the accuracies of each method are summarized in table 1. The standard deviations for the

average accuracies of the measurements were all within 10%. Also, Box plots comparing each method are shown in figure 5.

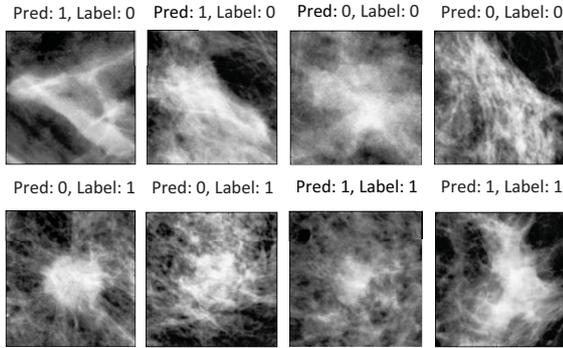


Fig. 4. Sample output images of the proposed model (Normal: 0, Malignant: 1)

Table 1. Summary of accuracies

Data size	# of conv filters	Image sizes	Optimizer	Accuracy
Original	32/64	28x28	RMS	<b>0.78</b>
Aug	<b>32/64</b> <b>64/128</b>	28x28	RMS	<b>0.81</b> <b>0.86</b>
Aug	64/128	<b>64x64</b> <b>128x128</b> <b>256x256</b>	RMS	<b>0.86</b> <b>0.88</b> <b>0.89</b>
Aug	64/128	28x28	<b>RMS</b> <b>ADAM</b>	<b>0.86</b> <b>0.89</b>

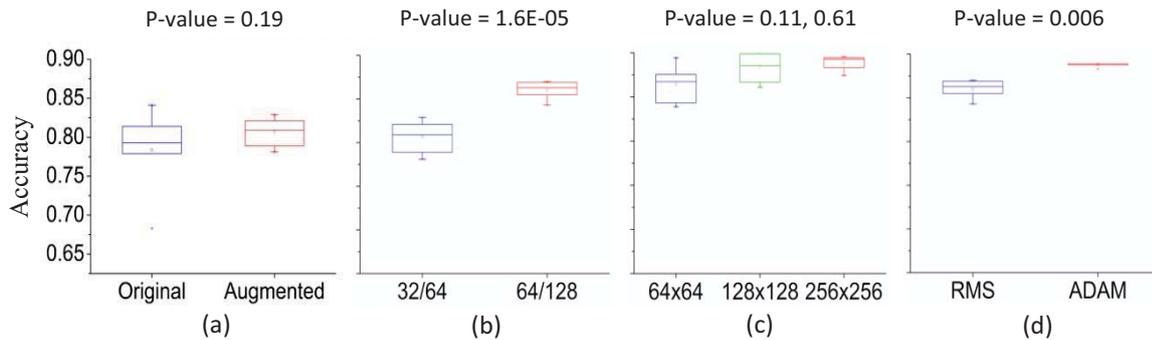


Fig. 5. Comparison of the various methods (a) Data augmentation, (b) # of convolutional filters, (c) image sizes, and (d) Optimizer

#### IV. DISCUSSION

Data Augmentation allowed the model to determine and learn the most relevant features from bigger data, and it actually increased the accuracy and specificity by 3% and 15%. However, the sensitivity dropped by 8% and the p-value of 0.19 indicates that the accuracies of original dataset and augmented data are not significantly different.

Adding more convolutional filters increased the accuracy, sensitivity, and specificity by 5%, 4%, and 7%, verifying that more filters extracted more features from the images. A p-value of 1.6E-05 is also consistent with the result that adding more convolutional filters significantly increased the testing accuracy.

Increasing image sizes raised the accuracy from 0.86 to 0.89 when compared at 30 epochs, verifying that it enabled the model to extract smaller and more detailed features. However, accuracies of different image sizes did not show much difference after they stabilized. Therefore, 28 by 28 image sizes were used to compare the

optimizers for a faster performance.

Changing the optimizer increased the accuracy by 3%, and the sensitivity by 6%, but decreased the specificity by 2%. The p-value of 0.006 also indicates that changing the optimizer from RMS to ADAM significantly increased the testing accuracy. Also, changing the optimizer resulted in a huge increase in the training accuracy (about 15%) as shown in Figure 6, and is expected to make a good training model for further use.

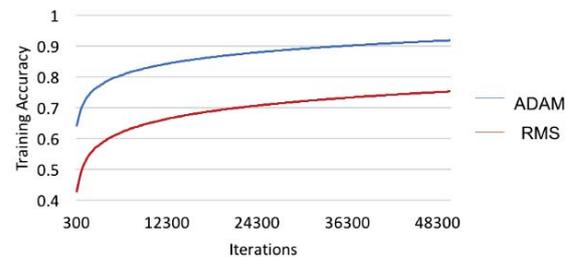


Fig. 6. Iterations vs. Training accuracy comparing RMS and ADAM optimizers

The main goal of this paper was to build the optimal model for breast mass classification by applying various methods that influence the performance of Convolutional Neural Network (CNN). The proposed model achieved the accuracy of 0.887, sensitivity of 0.903, and specificity of 0.869 for normal tissue versus malignant mass classification with augmented data, more convolutional filters, and ADAM optimizer.

Therefore, it is verified that breast mass classification using CNN has potential to be a better assisting tool than a CAD system in providing a consistent second opinion to a radiologist by reducing false-positive and false-negative diagnoses [10]. A limitation of this method, however, was that it only considered malignant masses that are relatively easy to classify than benign masses. Therefore, further studies are required in order to properly classify any given data for medical uses.

#### Acknowledgement

This study was supported by the Biomedical Engineering Branch of the National Cancer Center. I would like to express sincere gratitude to the Biomedical Engineering Branch colleagues who provided assistance and insight.

#### REFERENCES

- [1] Lindsey A. Torre, Freddie Bray, Rebecca L. Siegel, Jacques Ferlay, Joannie Lortet-Tieulent and Ahmedin Jemal, "Global Cancer Statistics," *A Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87-108, 2015.
- [2] "American Cancer Society Recommendations for Early Breast Cancer Detection in Women without Breast Symptoms." *American Cancer Society*. N.p., 20 Oct. 2015. 14 July 2016.
- [3] Arzav Jain and Daniel Levy, "DeepMammo - Breast Mass Classification Using Deep Convolutional Neural Networks," pp. 1-7, 15 July 2016.
- [4] J.J. Fenton, L. Abraham, S. H. Taplin, B.M. Geller, P.A. Carney, C. D'Orsi, J. G. Elmore, W. E. Barlow, "Effectiveness of computer-aided detection in community mammography practice," *J. Natl. Cancer Inst.*, vol. 103, pp. 1152–1161, 2011.
- [5] Thijs Kooi, Albert Gubern-Merida, Jan-Jurre Mordang, Ritse Mann, Ruud Pijnappel, Klaas Schuur, Ard Den Heeten and Nico Karssemeijer, "A Comparison Between a Deep Convolutional Neural Network and Radiologists for Classifying Regions of Interest in Mammography," *Breast Imaging Lecture Notes in Computer Science*, vol. 9699, pp. 51-56, 2016.
- [6] Mordang, Jan-Jurre, Tim Janssen, Alessandro Bria, Thijs Kooi, Albert Gubern-Merida, and Nico Karssemeijer, "Automatic Microcalcification Detection in Multi-vendor Mammography Using Convolutional Neural Networks," *Breast Imaging Lecture Notes in Computer Science*, pp. 35-42, 2016.
- [7] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.* 15, pp. 1929–1958, 2014
- [8] Samer Hijazi, Rishi Kumar and Chris Rowen, "Using Convolutional Neural Networks for Image Recognition," *Candence*, pp.1 – 12, 2015.
- [9] Diederik P. Kingma and Jimmy Lei Ba. "Adam: A Method for Stochastic Optimization," *ICLR 2015*, pp. 1-15, 2015.
- [10] B. Sahiner, Heang-Ping Chan, N. Petrick, Datong Wei, M.A. Helvie, D.D. Adler, and M.M. Goodsitt, "Classification of Mass and Normal Breast Tissue: A Convolution Neural Network Classifier with Spatial Domain and Texture Images," *IEEE Transactions on Medical Imaging*, vol. 15, no.5, pp. 598-610, 1996.

## Authors



**Jin Lee** is currently pursuing Bachelor of Arts degree in biology from Taylor University, Upland, Indiana, United States. She was a student researcher at National Cancer Center Research Institute (Biomedical Engineering Branch) in 2016. Her research interests include image

classification, artificial intelligence, deep learning, 3D printing, genetics, and genomics.



**Kwangjong Choi** is currently a junior in Korea Christian International School. He was a student researcher at National Cancer Center Research Institute (Biomedical Engineering Branch) in 2016. His research interests include image classification, image segmentation, artificial intelligence, and deep learning.



**Seong jung Kim** is at yonsei univ student in Korea and his major is Computer Science now he is in researcher at National Cancer Center Research Institute and he is studying Machine Learning that is applied to Medical Prediction.



**Woong Bae Yoon** is Researcher in National Cancer Center Research Institute (Biomedical Engineering Branch) in 2016. He is now a candidate for a Master's degree in computer and software from Hanyang University, Seoul, Korea. He received Bachelor of Engineering degree in Computer Science from Dankook University. He has interests in the areas of

image processing, augmented reality.



**Ji Eun Oh** received the MS degree and PhD in Radiological Science from Yonsei University, Korea in 2008 and 2013. She is a researcher in National Cancer Center Research Institute (Biomedical Engineering Branch). She has interests in the areas of medical imaging, 3D reconstruction, radiological science, and big data processing.



**Kwang Gi Kim** received the MS degree in Physics from POSTECH, Korea in 1998, and PhD in Biomedical Engineering from Seoul National University, Korea. He is Chief Researcher in National Cancer Center Research Institute (Biomedical Engineering Branch) in 2016. He has interests in the areas of medical imaging, big data processing, medical robot and

medical optics area.