Breast Cancer Lesion Detection and Classification in mammograms using Deep Neural

A R J Silalahi

Department of Medical Physics, Matana University, Tangerang, Indonesia

*corresponding author: alexander.silalahi@matanauniversity.ac.id

Abstract. A method to automatically detect and classify a lesion into either malignant or non-malignant is presented in this work. The dataset used is obtained from INbreast database and are in format of full-field digital mammography (FFDM). Some of the key challenges in detecting cancerous lesion in mammography are the low contrast between cancerous lesion and its surrounding tissues, apparent contrast similarities between lesion and pectoral muscle, presence of calcifications that may disrupt the detection process, and some level of morphological similarities between the lesion and some normal tissues. The work here consists of two main parts. The first part is the image processing section that aims to sample the lesion with intended lesion-to-surrounding ratio (0.4-0.6) and to avoid sampling from unintended regions such as pectoral muscle. Another key challenge is that the database is relatively small while machine learning requires a relatively large dataset. To improve size of samples, eighty fixed-size images (250 pixels x 250 pixels) are randomly cropped out of each of the previously processed image. The second part is to build the machine learning application based on deep neural network framework to classify samples into two classes, malignant and non-malignant. Our calculations show that both models can detect a single lesion with more than 90% accuracy and area under ROC curve >0.94.

1. Introduction

Breast cancer is the most prevalent cancer among women and among the top of all population. According to USA statistics one in every 8 women will develop some type of breast cancer in their life [1]. Mammography is the gold standard used in medical field to detect and diagnose breast cancer in various stages and conditions. In developed countries, early mammography is conducted at age of 40 and is continued with forward diagnose time interval adjusted to the findings. Mammography has shown to reduce the breast cancer mortality in the range of 25-30% [2]-[3]. However, the trial findings have met some concerns regarding the validity of the results [4]-[7]. In the typical mammography examination, the X-ray images are captured from two different angles of breast. The images are then further analyzed by two radiologists and raised malignant suspicion is followed up with further diagnostic evaluation.

The currently implemented screening evaluation of mammogram images require radiologists to read and analyze the entire image area, which is a tiring and monotonous job, and thus are prone to errors. Computer-aided detection (CAD) programs were developed to identify suspicious locations within the mammograms, which are then continued with further readings by the radiologists. These CAD programs [8] have shown to be useful and helpful for radiologists, in which the use of CAD has similar performance to double reading. However, it also is still controversial because its usage has not shown a performance improvement of radiologists in the last decade in USA. It also is rather expensive with current cost around USD 400 million every year [9]. The current state-of-art CAD program uses handcrafted features of an X-ray images which are then sent to machine learning for identification and classification.

The manual process in selecting the most relevant features can result in wrongly chosen features and results in wrong representation of the mammogram. The exciting and rapid development in the field of computer vision brings the deep neural network, a technique based on artificial neural network, capable
of detecting, classifying and localizing objects in images with very high accuracy, without hand crafting the relevant features. The algorithm simply learns the “correct” features from the mammogram samples. There have been studies using machine learning classifiers [10]-[11] and deep neural networks [12]-[15] on mammography images. The present work is an attempt build a deep neural network model and to improve its predicting power over its predecessors.

The paper is organized into four sections. The required dataset and methods are presented in Section 2 this includes the formulation of deep neural net framework used in the present work. Section 3 contains the results of simulations which mostly includes the performance of deep neural net model in terms of accuracy, convergency, and the attempt to improve the performance by using resnet. Finally, Section 4 discusses the future works to further improve the performance of such CNN method.

2. Materials and Methods

The use of machine learning (ML) to detect and classify malignant breast cancer is not new subject of research and has been applied with various types ranging from rather traditional type of artificial neural network to convolutional neural network. The scope of detection also ranges from patch detection to whole slide image of mammogram. In the current work we use two CNN-based machine learning frameworks with two different schemes: VGGnet and Resnet. These two methods have very good predicting power with relatively simple networks.

2.1. Whole slide image processing

The mammograms were obtained from INbreast database [16] which provides FFDM of 116 patients. According to American College of Radiology [17], breast cancer can be classified using BI-RADS into seven categories (see Table 1).

| BI-RADS index | Interpretation                     |
|---------------|-----------------------------------|
| 0             | Insufficient study                |
| 1             | Negative                          |
| 2             | Benign findings                   |
| 3             | Likely benign findings            |
| 4             | Suspicious findings               |
| 5             | Highly likely of malignancy       |
| 6             | Biopsy proven malignancy          |

In the present work we only try to classify the patches of mammograms to either fully non-malignant or to partially contain malignant lesion. One of the key challenges with using electronic medical record (EMR) is the unavailability of large public database. In the present work, we try to improve the number of samples by randomly choosing quite large number of patches within a mammogram image. The randomness is imposed to make sure that there is no correlation between patches of the same mammogram slide. The algorithm is also tuned to avoid region of pectoral muscle, because previous researches have shown degradation of detection accuracy if it is included. We also know from previous findings that the contrast between tumor region and its surrounding tissue is one of the most important features in defining a tumor. Because of this reason, random patches that include tumor region must contain both tumor area and enough surrounding area. In this work we define ratio (R) as tumor area divided by surrounding area.

\[ R = 0.4 - 0.6 \] (1)
Figure 1. Cropping lesions from malignant and non-malignant areas (excluding pectorial muscle region).

2.2. ConvNet Parameterization

In the present work we use two deep neural net framework: plain and resnet. The two schemes utilize ConvNet blocks that include modules such as: nonlinear activation function, max pooling, batch normalization, dropout, and classification layer. The resnet general structure is similar to VGGNet but is modified to include a shortcut link that connects two previously identified modules. One of the main reasons for this modification is because in general, deep neural network suffers from vanishing gradient issue. The shortcut link ensure that the deeper neural network always performs better than its shallower version.

The activation functions used is the rectified linear unit (ReLU) with form:

\[(x) = \max(0, x)\]  

(2)

Each of the convolutional layer is then processed with a batch normalization layer. A batch normalization does similar thing to hidden layer weights to what normalization does to input data. This process reduces how much hidden values shift around and helps each layer to learn by itself a little bit more independently. Dropout layer is used to avoid overfitting by probabilistically removing numbers of input to a layer. In present work 0.25 fraction of inputs is dropped randomly. Max pooling layer is used to reduce the dimensionality of the inputs and to help the representation becomes slightly invariant to small translation of the input data [18].

2.3. ConvNet Architectures

Here we will employ two frameworks, plain and resnet [19]. The resnet building block is built similarly to VGGNet [20] in which we have serially connected modules. Building blocks are defined to represent convolutional neural network and other necessary modules. The shortcut link connection is applied every two of these building blocks, starting at the input of building block 1 and ending at right before the ReLu module of building block 2.
Figure 2. Shortcut link in resnet.

Shortcut link contains addition of $x$ and $(x)$, thus their dimensions must be equal. If they are not, we need to do matrix linear projection to equalize the two.

Figure 3. Plain and resnet framework.

3. Result and Discussion

3.1. Accuracy and Precision
The first calculations show that plain and vgg-16 model seems to show a better predicting power than resnet, as shown in Figure 4. The plain and resnet models were used with depth of layer equals to $6n+2$. 
3.2. Overfitting test

Here we want to see how both models suffer from overfitting issue in which the predicting accuracy over train data is better than test data.

![Figure 4. Accuracy and Precision on Train data.](image)

![Figure 5. Overfitting plot on plain and resnet.](image)

The calculation indicated that both models suffer from overfitting as it gets worse as depth of layer increases as shown in figure 5. It is also clear that overfitting issue is worsened in resnet model. This conclusion is in agreement with the He [19].

3.3. Augmented Data

Machine learning requires a relatively large dataset to train the model to have high predicting power. One of the issues in predicting breast cancer is the limited availability of mammogram dataset. Although we have used independent lesion sampling to improve our size of input data, it is still possible that the model doesn’t completely capture all the relevant features as predicting basis. One of the possible remedies to this problem is to produce additional data from the current training set. This new data is generated by modifying the images through horizontal and vertical flipping and through images shifting. This method is called augmented data and is applied only on training dataset.

In Figure 6 we show that the predicting accuracy of resnet are the same for \( n=1 \) and 4 which converge to around 90%. The calculation also shows that predicting accuracies are roughly the same for test and training data. The robust convergence of resnet is in agreement with previous finding [19].
3.4. Area Under Curve (ROC)

ROC curve [21] is a plot of true positive rate vs false positive rate; thus each point of data represents a pair of sensitivity/specificity. A perfect predictor would have an ROC that passes through the most left upper point and area under curve is equal to one. Area under the curve represents how well the model can distinguish lesion with malignancy and non-malignancy. Our calculations indicate that the resnet with deeper layer performs better than plain and shallow resnet as shown in Figure 7.

![Figure 7. ROC Curve for resnet and plain with depths of 8 and 26. Random Predictor is shown as straight line.](image)

3.5. False Predictions

In this section we want to understand why and how our model fails to detect malignancy or normal lesions. The two possible miss classifications are the false positive (normal identified as malignant) and false negative (malignant identified as normal). We sort the mammograms based on how large the predictions deviate from the true annotations and how many lesions are misclassified in a single mammogram. Here we show the WSI with highest number of wrongly identified lesions, ones that are belong to the false positive and false negative. The false positive predictions tend to occur in whiteish mammogram image that is usually associated with dense breast tissue. On the other hand, the false negative predictions tend to occur in greyish mammogram image that is associated with less dense breast tissue. It is interesting to note that the high dense vs low dense breast tissue has been identified in previous study as one of prominent feature differences between Asian and Caucasian women [22]-[23].
4. Conclusions
In the present work we have been able to automatically detect a lesion in mammogram and classify it into either malignant or non-malignant. This automated method is based on a class of machine learning known as deep neural network, specifically plain and resnet. The two frameworks were applied with varying depths and were demonstrated to reach accuracies more than 90%. The convergence of resnet prediction is ensured with applying augmented data which includes vertical and horizontal flipping, and image shifting. We have also shown that the resnet has the improved ROC curve over the plain’s. The areas under the roc curve are 0.943 and 0.923 for resnet and plain with depths of 26. The present work, which is lesion detection, is part one of our project and should be continued with part two, in which we incorporate this resnet model into faster r-cnn and mask r-cnn method [24]-[25] to automatically detect, locate, and segmentize malignant patches in a mammogram WSI.

References
[1] Siegel R L, Miller K D and Jemal A 2019 Cancer statistics, 2019 CA. Cancer J. Clin. 69 7–34
[2] Bleyer A, Baines C and Miller A B 2016 Impact of screening mammography on breast cancer mortality Int. J. Cancer 138 2003–12
[3] Mayor S 2017 Mammography screening reduces breast cancer deaths in Ireland, study finds BMJ 359
[4] Kopans D B 2003 Opinion. The most recent breast cancer screening controversy about whether mammographic screening benefits women at any age: Nonsense and nonscience Am. J. Roentgenol. 180 21–6
[5] Feig S A 2014 Screening Mammography Benefit Controversies. Sorting the Evidence Radiol. Clin. North Am. 52 455–80
[6] Lee C H 2002 Screening mammography: proven benefit, continued controversy Radiol. Clin. North Am. 40 395–407
[7] Green B B and Taplin S H 2003 Breast cancer screening controversies J. Am. Board Fam. Pract. 16 233–41
[8] Anon Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion : American Journal of Roentgenology : Vol. 181, No. 4 (AJR)
[9] Lehman C D, Wellman R D, Buist D S M, Kerlikowske K, Tosteson A N A and Miglioretti D L
2015 Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern. Med.* **175** 1828–37

[10] Wei L, Yang Y, Nishikawa R M and Jiang Y 2005 A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications *IEEE Trans. Med. Imaging* **24** 371–80

[11] Ramos-Pollán R, Guevara-López M A, Suárez-Ortega C, Díaz-Herrero G, Franco-Valiente J M, Rubio-Del-Solar M, González-De-Posada N, Vaz M A P, Loureiro J and Ramos I 2012 Discovering mammography-based machine learning classifiers for breast cancer diagnosis *J. Med. Syst.* **36** 2259–69

[12] Kool T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez C I, Mann R, den Heeten A and Karssemeijer N 2017 Large scale deep learning for computer aided detection of mammographic lesions *Med. Image Anal.* **35** 303–12

[13] Dhungel N, Carneiro G and Bradley A P 2015 Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests *2015 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2015* (Institute of Electrical and Electronics Engineers Inc.)

[14] Wang J, Yang X, Cai H, Tan W, Jin C and Li L 2016 Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning *Sci. Rep.* **6** 1–9

[15] Becker A S, Marcon M, Ghafoor S, Wurnig M C, Frauenfelder T and Boss A 2017 Deep Learning in Mammography *Invest. Radiol.* **52** 434–40

[16] Moreira I C, Amaral I, Domingues I, Cardoso A, Cardoso M J and Cardoso J S 2012 INbreast: Toward a Full-field Digital Mammographic Database. *Acad. Radiol.* **19** 236–48

[17] D’orsi C, Bassett L, Berg W, Feig S, … V J-A C of and 2003 undefined Breast imaging reporting and data system: ACR BI-RADS-mammography

[18] Aloysius N and Geetha M 2018 A review on deep convolutional neural networks *Proceedings of the 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017* vol 2018-January (Institute of Electrical and Electronics Engineers Inc.) pp 588–9

[19] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vol 2016-Decem pp 770–8

[20] Chung C, Patel S, Lee R, Fu L, Reilly S, Ho T, Lionetti J, George M D and Taylor P 2018 VGGNET *Am. J. Heal. Pharm.* **75** 398–406

[21] Bradley A P 1997 The use of the area under the ROC curve in the evaluation of machine learning algorithms *Pattern Recognit.* **30** 1145–59

[22] Bae J M and Kim E H 2016 Breast density and risk of breast cancer in Asian women: A meta-analysis of observational studies *J. Prev. Med. Public Heal.* **49** 367–75

[23] Mariapun S, Li J, Yip C H, Taib N A M and Teo S H 2015 Ethnic differences in mammographic densities: An Asian cross-sectional study *PLoS One* **10**

[24] Ren S, He K, Girshick R and Sun J 2017 Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1137–49

[25] He K, Gkioxari G, Dollár P and Girshick R 2020 Mask R-CNN *IEEE Trans. Pattern Anal. Mach. Intell.* **42** 386–97