Mass Detection in Breast Using Transfer Learning for Computer Aided Diagnosis

Md. Kamrul Hasan, Lavsen Dahal, Fakrul Islam Tushar

Erasmus Scholar on Medical Imaging and Applications
Università degli Studi di Cassino e del Lazio Meridionale
Cassino, Italy

Email: {md-kamrul_hasan, lavsen.dahal, fakrulislam.tushar}@etu.u-bourgogne.fr

Abstract **- Mammography is the most widely used gold standard method for the screening of the breast cancer and Mass detection is the prominent pre-processing step. State-of-art performances of the DCNN architectures in the field of classification made them an obvious choice for the image classifications. However, due to limited availability of medical data, application of DCNN architectures in medical images is challenging. In this project our contribution was mass detection in mammographic images using two different approaches. First approach was applying transfer learning concept which lessen the demand of data to use a pre-trained publicly available VGG16 model and second approach was training a Alexnet from the scratch to classify mass and non-mass mammographic images. Both the models were trained and tested with different hyperparameters and different data size. From the experiment results, it can be concluded that transfer learning approach for VGG16, training from fully connected layer 2 (fc2) has promising expected result for mass detection. In our research, maximum accuracy for mass detection was 93.60 % for VGG16 with 13500 train images.**

Index Terms – Mammography; Mass and Non-mass; AlexNet; VGG16; Deep convolutional neural network (DCNN).

I. Introduction

Cancer is a foremost public health problem worldwide and is considered as the second leading cause of death. Approximately, 70% of deaths from cancer occur in low- and middle-income countries [1]. Among all types of cancer like Bladder cancer, Lung cancer, Brain cancer, Breast cancer, Non-Hodgkin lymphoma etc., Breast cancer is the severe type of cancer and is taken as a second cause of death especially for the women [2]. A tumor is an uncontrolled growth of breast cells which are generally two types e.g. non-cancerous or 'benign', and cancerous or malignant. The term "breast cancer" refers to a malignant tumor as shown in Fig. 1. that has developed from cells in the breast [3]. There are plenty of techniques for the breast mass detection e.g. X-ray (Mammography, Digital breast tomosynthesis, Xeromammography, Galactography), MRI, CT, PET, Ultrasound, and Scintimammography etc. Among all those methods for the breast cancer detection, Mammography is the most commonly used method and key screening tool of breast imaging that uses low-dose amplitude-X-rays to examine the patient breast. Cancerous masses and calcium deposits appear brighter on the mammogram and now a day, mammography is the gold standard method to detect early stage breast cancer before the lesions become clinically palpable [5].

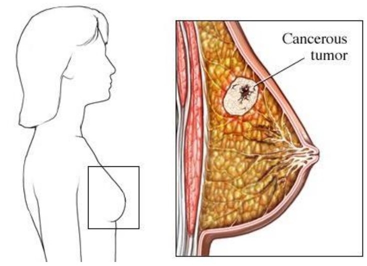
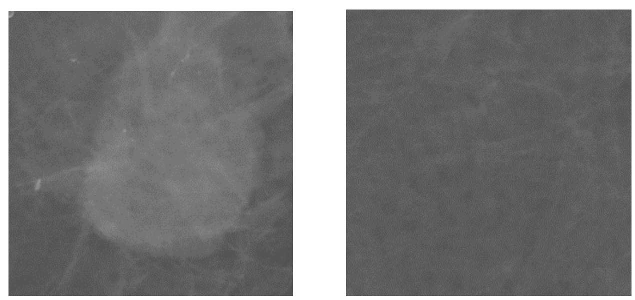


Fig. 1. Typical presentation of breast with corresponding mass [4].

Mammographic radiographic units use x-rays to produce images of the breast that provide information about breast morphology, normal anatomy, and gross pathology. Mammography is used primarily to detect and diagnose breast cancer and to evaluate palpable masses and nonpalpable breast lesions [6]. After getting mammography images, foremost challenges are to classify the mass and non-mass region of breast for the CAD systems. Typical mammogram of mass and non-mass region of breast is shown in Fig. 2.



(a) (b)

Fig. 2. Mammography image a) mass and b) non-mass

There are plenty of methods in CAD systems which have been done to detect the mass and non-mass of the breast. In ref [7], K-means clustering, Template-matching, Simpson's Diversity Index and finally SVM was used that had 83.94 % accuracy. In ref [8], Principal components analysis, Gabor wavelet, independent component analysis, and finally linear discriminant analysis was used with the accuracy 90.07 %. In ref [9], Transfer learning e.g. AlexNet with 1656 mammography images (454x454) was used that provided 85.35 % accuracy. In ref [10-13], mass candidate or region of interest (ROI) was detected, and the feature vector of the ROI was extracted based on special knowledge and then classified the ROI into mass or non-mass classes.

In recent years, deep learning has attracted great attention in the field of image classifications. Deep convolutional neural network (DCNN) has become fruitful learning due to having outstanding performance in recognition of natural images. The applications of deep learning on medical images are not abundant yet due to lack of training images for learning parameters. In this research, our main contribution is to use transfer learning approach for VGG16 Deep Convolution Neural Network models to train last layer and Alexnet for all the layers with mammography images.

The remaining sections of this paper are organized as follows: section 2 discusses the overall pipeline for this research to classify the regions of interest (454x454) extracted from mammographic images as mass and non-mass. Section 3 is for results and discussion for the proposed pipeline. Finally, the paper is concluded in section 4.

II. Pipeline of The Research

Overall pipeline used for this research is mentioned in a block diagram as shown in Fig. 3.

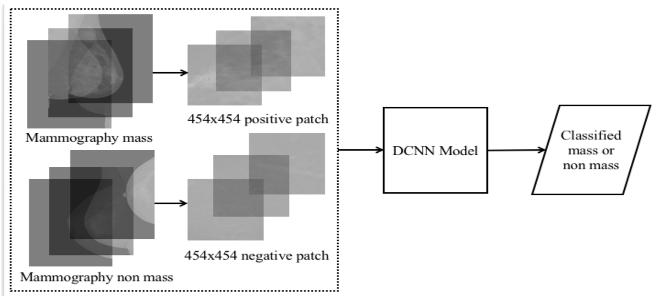


Fig. 3. Proposed pipeline of the research

2.1. Patch Extraction

The dataset used for this research is obtained from INbreast Challenge. The dataset has 107 mammography images which have masses. There is also a ground truth for these images where the masses have been manually segmented by the experts. Firstly, from the original images, the desired patch size (454×454) was extracted. The pseudocode used for the patch creation is given in Table 1.

Table 1: Pseudocode for patch extraction

```
Load Target Image (Im).
Find rows, columns and select height, width, stride.
for i=1:1:rows
      Final_row=initial_row+height;
      if  Final_row < rows
            for j=1:1:columns
                  Final_column=initial_column+width;
                  if  Final_column < columns
                        NewImage=Croped(TargetImage);
                        Saved;
                  end;
                  initial_column = initial_column + stride;
            end;
            initial_column =1;
      end;
      initial_row = initial_row + stride;
end;
```

2.2. Experimental Set up

The experiment was carried out in two workstations with configurations as follows:

	CPU	RAM	GPU	CPU
Work station 1	1 (8 Hyper Threaded Core)	256 GB	2 TitanXP (12 GB onboard)	Work station 1
Work station 2	1 (4 Hyper Threaded Core)	32 GB	1 TitanXP (12 GB onboard)	Work station 2

The software configurations that were used for this experiment are as follows:

CUDA Version: 8.0

DL4J Framework Version: 1.0.0-alpha

JAVA SDK Version: 1.8

Maven: 4.0.0

2.3. Architecture of the DCNN

The deeper the convolutional neural network more good the results but the limitation is deeper the CNN more data needed for training the network. DCNN needs a lot of data to learn the features which comes in form of training data, and it becomes very crucial in case of high dimensional samples like images. So, in practice it becomes very different to train the DCNN from scratch with random initialization due to the lack of dataset of sufficient size. The idea of transfer learning contributes a lot to that limitation.

Transfer learning is the approach of using a pre-trained model which has already been trained on a huge dataset and it has learned the weights and parameters. The concept is that this pretrained model will act as feature extractor and the layers of the network will be edited as the requirement of the specific classification problem. Although, still is not so clear to the reaches worldwide how DCNN visualize and works but recent researches provide few concepts why transfer learning works. Authors in [14] shows the visualization of the network. The lower layers of the network will detect low level features like edges, curves and most last years are responsible for specific features. So, applying transfer learning concept except training the whole network through random initialization of weights, the weights of the pretrained network was used and more important layers for training. Major transfer learning scenarios are-

Feature Extractor: The pre-trained network is use as a fixed feature extractor by removing the output layer that gives the probabilities.

Fine Tune: The model can be trained partially trained. Where the output layer can be replaced by the specific classifier depending on the task. Another way was to freeze the weights of the initial layers of the model and retrain the higher layer on the dataset.

In this research, Alexnet [14] and VGG16 [15] were adopted to perform the non-mass and mass detection. AlexNet is one of the most influential works in the field of CNNs, which changed the thinking of the computer vision community and CNNs became a common choice for classification and detection tasks. In the ImageNet LSVRC-2010 contest AlexNet achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. Authors in [14] discussed the architecture of the AlexNet as shown in Fig. 4.

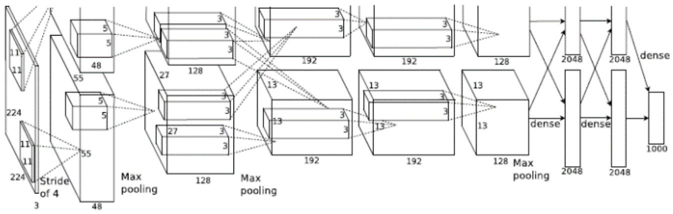


Fig. 4. AlexNet architecture

Alexnet used simple approach and their optimization methods data augmentation and dropout still so effective compared to modern architectures such as GoogleNet, ResNet. Brief discussion about the architectures of AlexNet is given below-

- AlexNet architecture has 60 million parameters and 650,000 neurons, consist 5 conv layers, maxpolling layers, dropout layers and 3 fully connected layers with a final 1000-way softmax activation [14].
- For faster training with gradient descent, Rectified Linear Unit (ReLu) is used which was different from that time the standard method tanh.
- For artificial data enlarging data augmentation techniques used. Three different approaches have been used: image translation, patch extraction and horizontal reflection.
- To reduce test error and avoid overfitting dropout layer was implemented.
- The model was trained using batch stochastic gradient descent, with specific values for momentum and weight decay [14].

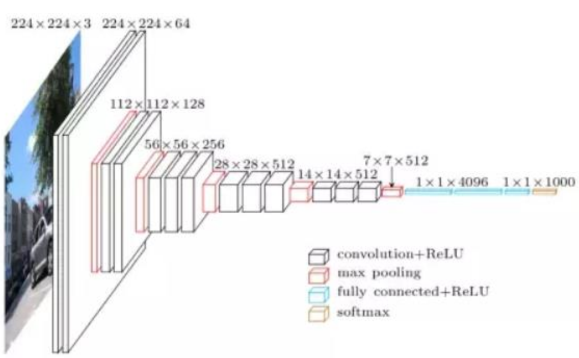


Fig. 5. Visualization of VGG architecture

VGG network as shown in Fig. 5 is characterized by its simplicity, using only 3×3 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier

III. Results and Discussions

The overall results are divided into three parts e.g. train AlexNet from the scratch, train of VGG16 from fc2 others are frozen and train from bottleneck of VGG16 others are frozen. To get train and test images, firstly whole 107 images are split into the two subsets and they are mutually exclusive. Then train and test patches are separately selected by following the pseudocode in Table 1. The whole process is given in Fig. 6. There are several versions of the image sets to check the effect of the number of image and image augmentations to train the DCNN. The used image sets for this research are given in Table 2. There are only five cases of the image set. But, there are lots of experiment have been done by different combinations of those five cases which will describe later. After extracting patches of desired size, some patches that were inside and around borders of mask of the breast as shown in Fig. 7 were not included in our dataset.

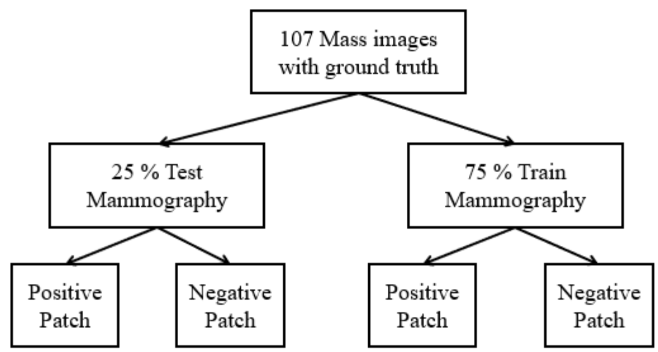
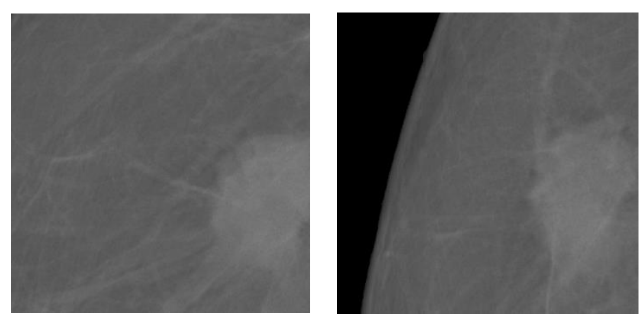


Fig. 6. Train (Positive and Negative) and Test (Positive and Negative) Patch extraction



(a) (b)

Fig. 7. Patch having mass a) inside mask b) inside and around borders of mask

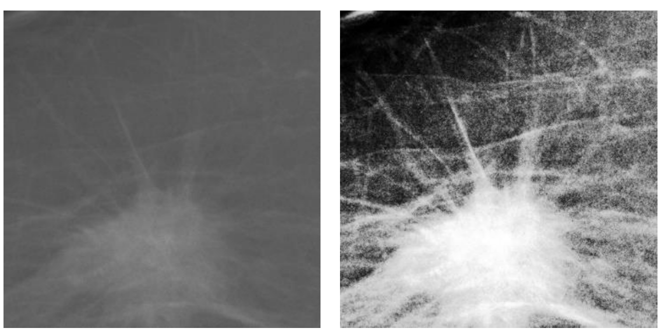
Table 2: Possible cases of data for experiment

Case #.	Train (75%)	Test (25%)
Case-1	P:750 and N:750 Inside mask No augmentation No pre-processing	P:250 and N:250 Inside mask No augmentation No pre-processing
Case-2	P: 2250 and N:2250 Inside mask	P: 750 and N:750 Inside mask

	Two augmentations No pre-processing	Two augmentations No pre-processing
Case-3	P: 7500 and N:7500 Inside mask Nine augmentations No pre-processing	P: 2500 and N:2500 Inside mask Nine augmentations No pre-processing
Case-4	P: 6750 and N:6750 Inside mask Eight augmentations No pre-processing	P: 2250 and N:2250 Inside mask Eight augmentations No pre-processing
Case-5	P: 1800 and N:1800 Inside and around borders of mask Eight augmentations No pre-processing	P: 600 and N:600 Inside and around borders of mask Eight augmentations No pre-processing

In case-2, two augmentation namely horizontal flipping and rotation have been done on both train and test images to create more data.

Similarly, in case-3, there are nine augmentations e.g. Adaptive Histogram Equalization, flipping about both axis, Gamma correction for the lighting, Perspective Translation for re calibration of the image frame, Rotations, Scaling, Normal Histogram Equalization, Flip Vertically, and Flip Horizontally have been done. But normal histogram equalization gives some images that look like synthetic image as shown in Fig. 8.



(a) (b)

Fig. 8. a) original Patch having mass b) normal histogram equalization as an augmentation

For that reason, in case-4, t normal histogram equalization has been excluded to examine the effect of this augmentation on both training and testing. In case 5, some patches inside and around borders of mask as shown in Fig. 7 have been included to verify their effects on the classifications of mass and non-mass.

In our first experiment, we have used VGG16 models. In this case, only last layer (fc2) has been trained with different cases formulated from Table 2. In all the cases, some performance metrics e.g. True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Accuracy, Precision, Recall and F score have been measured. Some of the research outcomes using the several cases of the train and test data as mentioned in Table. 2. have been shown in Table 3. Similarly, results for the VGG16 train from bottleneck and AlexNet train from scratch are given in Table 4 and Table 5.

Table 3: Different Train and Test outcomes from VGG16 (from fc2 others in frozen)

Exp. No.	**Types of Patch/ Hyperparameters of DCNN**	**Performance metrics**	
1.0	Train and Test from Case 1	TP	155
		FP	1
		FN	95
		TN	249
		Accuracy	0.8080
		Precision	0.8587
		Recall	0.8080
		F1 score	0.7635
2.0	Train and Test from Case 2	TP	729
		FP	89
		FN	21
		TN	661
		Accuracy	0.9267
		Precision	0.9302
		Recall	0.9267
		F1 score	0.9298
3.0	Train and Test from Case 3	TP	2403
		FP	223
		FN	97
		TN	2277
		Accuracy	0.9360
		Precision	0.9371
		Recall	0.9360
		F1 score	0.9376
4.0	Train from Case 3 and Test from Case 1	TP	211
		FP	39
		FN	39
		TN	211
		Accuracy	0.8440
		Precision	0.8440
		Recall	0.8440
		F1 score	0.8440
5.0	Train from and Test from Case 4	TP	2195
		FP	392
		FN	55
		TN	1858
		Accuracy	0.9007
		Precision	0.9099
		Recall	0.9007
		F1 score	0.9076

Table 4: Different Train and Test outcomes from AlexNet (from scratch)

Exp. No.	**Types of Patch/ Hyperparameters of DCNN**	**Performance metrics**	
1.0	Train and test both from case 1 No augmentations have been done on Train. Positive: 750, Negative:750 Epoch: 50, batch 20 and channel 1	TP	133
		FP	7
		FN	17
		TN	143
		Accuracy	0.9200
		Precision	0.9219
		Recall	0.9200
		F1 score	0.9172
2.0	Train and test both from case 1 2 augmentations (Warping and Flipping) have been done on Train. Positive: 1800, Negative:1800 Epoch: 50, batch 20 and channel 1	TP	149
		FP	43
		FN	1
		TN	107
		Accuracy	0.8533
		Precision	0.8834
		Recall	0.8533
		F1 score	0.8713

3.0	Train and test both from case 1 2 augmentations have been done on Train. Positive: 1800, Negative:1800 Epoch: 50, batch 20 and channel 1	TP	145
		FP	17
		FN	5
		TN	133
		Accuracy	0.9267
		Precision	0.9294
		Recall	0.9267
		F1 score	0.9295
4.0	Train and test both from case 1 6 augmentations have been done on Train. Positive: 4200, Negative:4200 Epoch: 50, batch 20 and channel 1	TP	147
		FP	17
		FN	3
		TN	133
		Accuracy	0.9333
		Precision	0.9371
		Recall	0.9333
		F1 score	0.9363
5.0	Train and test both from case 1 10 augmentations have been done on Train. Positive: 4200, Negative:4200 Epoch: 50, batch 20 and channel 1	TP	136
		FP	2
		FN	14
		TN	148
		Accuracy	0.9467
		Precision	0.9495
		Recall	0.9467
		F1 score	0.9444

Table 5: Different Train and Test outcomes from VGG16 (Bottleneck and others layer keep freezing)

Exp. No.	**Types of Patch/ Hyperparameters of DCNN**	**Performance metrics**	
01	Train and Test from Case 2	TP	0
		FP	0
		FN	750
		TN	750
		Accuracy	0.5000
		Precision	0.5000
		Recall	0.5000
		F1 score	0.0000
02	Train and Test from Case 3	TP	0
		FP	0
		FN	2500
		TN	2500
		Accuracy	0.5000
		Precision	0.5000
		Recall	0.5000
		F1 score	0.0000
03	Train from case 3 and Test from Case 1	TP	82
		FP	23
		FN	168
		TN	227
		Accuracy	0.6180
		Precision	0.6778
		Recall	0.6180
		F1 score	0.4620

IV. Conclusion and Future Work

In this study, we presented a DCNN model for mass detection in mammography images. A transfer learning approach was used for VGG-16 model (train from fc2 and train from bottleneck layers), as well as Alexnet model was also trained for all layers. The experimental results showed that transfer learning e.g. train from fc2 approach worked well, which shows promising potential for it to be used for medical images where there is limited labelled data. But on the other hand, train VGG16 from the bottleneck layer does not have promising results like train from fc2 of VGG16. The AlexNet model that trained from scratch also performed very well for more number of images in train. To the best our knowledge, to get promising results from bottleneck layer, more data must be used since its learning near about 100M parameters. Moreover, our VGG16 edit last layer and AlexNet from the scratch providing promising results, we need more and more experiment to validate the results with our experimental set up.

Acknowledgment

Author would like to thanks to Professor Mario Molinara for his crucial and valuable suggestion and guideline over the research period and Università degli Studi di Cassino e del Lazio Meridionale for providing two workstations to perform research.

References

[1] World Health Organization: Cancer. Access on: 01 June 2018. Available at: http://www.who.int/news-room/fact-sheets/detail/cancer

[2] Cancer treatment centers of America. Access on: 01 June 2018. Available at: https://www.cancercenter.com/cancer/

[3] Breast Cancer. Access on: 01 June 2018. Available at: http://www.breastcancer.org/symptoms/understand_bc/what_is_bc

[4] Breast Cancer. Access on: 01 June 2018. Available at: https://www.kpwomenshealth.org/breast_health_breast_cancer.asp

[5] S. V. Sree et al. "Breast Imaging: A Survey," *World Journal of Clinical Oncology*, vol. 2, no. 4, p.p. 171–178, 2011.

[6] Radiographic units, mammography. Access on: 01 June 2018. Available at: http://www.who.int/medical_devices/innovation/hospt_equip_25.pdf

[7] A. Nunes, A. Silva and A. Paiva, "Detection of masses in mammographic images using geometry, Simpson's Diversity Index and SVM," in International Journal of Signal and Imaging Systems Engineering (IJSISE), vol. 3, No. 1, 2010.

[8] D. Costa, L. Campos and A. Barros, "Classification of breast tissue in mammograms using efficient coding," In Bio Medical Engineering on Line, vol. 10, no. 55, 2011.

[9] S. Suzuki, X. Zhang, N. Homma, K. Ichiji, N. Sugita,Y. Kawasumi, T. Ishibashi, and M. Yoshizawa, "Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis," In 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 1382–1386, 2016.

[10] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsit, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space," Phys. Med. Biol., vol. 40, no. 5, pp. 857- 876, 1995.

[11] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolutional neural network classifier with spatial domain and texture images," IEEE Trans. Med. Imag., vol. 15, no. 5, pp. 598-610, 1996.

[12] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Gradient and texture analysis for the classification of mammographic masses," IEEE Trans. Med. Imag., vol. 19, no. 10, pp. 1032-1043, 2000.

[13] J. Wei, B. Sahiner, L. Hadjiiski, H. Chan, N. Pet- rick, M. Helvie, M. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," Med. Phys., vol. 32, no. 9, pp. 2827-2837, 2005.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, p.p. 1097-1105, 2012.

[15] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556.