



Investigations on Deep Learning Techniques for Analysing Mammograms

Satish Babu Bandaru^{1*}, Natarajasivan Deivarajan², Rama Mohan Babu Gatram³

¹ Department of Computer Science and Engineering, Annamalai University, Annamalinagar, Tamil Nadu 608002, India

² Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Annamalai University, Annamalinagar, Tamil Nadu 608002, India

³ Department of Computer Science and Engineering (AI & ML), RVR & JC College of Engineering, Guntur, A.P. 522019, India

Corresponding Author Email: researchbsbabu@gmail.com

<https://doi.org/10.18280/ria.360313>

ABSTRACT

Received: 8 January 2022

Accepted: 25 May 2022

Keywords:

mammograms, deep learning (DL) techniques, residual network convolutional neural networks (CNNs), inception

Mammograms have been acknowledged as one of the most reliable screening tools as well as a key diagnostic mechanism for early breast cancer detection. Though mammography is a valuable screening tool for detecting malignant growth in breasts, its competence as a diagnostic tool is heavily reliant on the radiologists' understanding. Automated systems are now widely used for detection of breast cancer. Image processing techniques were widely used in automated systems for classifying mammograms. Of late with the advent of deep learning (DL) where images can be processed directly for classification, the DL is widely researched for medical image classification. Basically, DL techniques are representation-learning methods which aid in understanding data like sounds, images as well as texts. DL algorithms have the ability to learn multiple levels of representation as well as abstraction. Residual network (ResNet) is given due consideration as a kind of highly advanced Convolutional Neural Networks (CNNs). This work has offered a potential application of Visual Geometry Group (VGG), Residual network (ResNet) and Inception based CNN model for differentiating the mammograms into the abnormal class and the normal class. Experimental results demonstrated that the deep learners are effective for classifying mammograms and Inception deep learner achieved the best accuracy of 91.49%.

1. INTRODUCTION

The identification of breast cancer at initial stage, improves the likelihood of successful treatment, and thus boost the illness forecast. Breast cancer detection can be performed using a multitude of screening techniques [1]. In contrast to the earlier devices, the radiation levels of the recently used mammograms are much lower. Mammography's chief task will involve dependable early identification as well as determination of breast cancer. While mammography is a valuable screening tool for detecting malignant growth in breasts, its competence as a diagnostic tool is heavily reliant on the radiologists' understanding. For early breast cancer detection, attempts are being made to develop the mammography's ability by boosting its preciseness as well as minimizing inconsistencies in its interpretation [2].

Presently, computer-aided detection and diagnosis (CAD) systems [3] offer vital support to the radiologists' process of decision-making. These systems are capable of drastically mitigating the time and effort required for assessing a lesion in the clinical practice as well as in reducing the number of false positives which in turn, will result in biopsies that are distressing and unnecessary. Mammography-related CAD systems are the goal of which is to identify and classify worrisome lesions on a mammogram (CADe) and to diagnose any such discovered lesions (CADx) into either malignant or benign. Normally, conventional CAD methods will require the features to be manually extracted from the images. Such

features are inclusive of original features like texture, and shape as well as features that are extracted using algorithms like Gabor filter, Local Binary Pattern, and Histogram of Gradient. Nevertheless, these conventional methods do suffer from certain constraints due to selection as well as combination of the features being heavily dependent on the designers' experience. Thus, in this work, the efficacy of Deep Learning (DL) methods is explored.

Nowadays, deep learning modeling has been found to be quite promising in numerous artificial intelligence (AI) applications, inclusive of biomedical imaging analysis [4]. A significant role is served by mammographic databases in the DL methods' training, testing as well as evaluation. Unlike the data required for training conventional neural networks, a DL network's training will require tremendous data amounts. For the growth of DL in medical imaging, a comprehensive annotated databases are vital. Asymmetry, architectural distortion (AD), calcification (MC), and abnormal areas of mass can be commonly observed in mammography.

Predominantly for a complex problem's resolution, a few extra layers are stacked in the Deep Neural Networks so as to obtain improved accuracy as well as performance. The underlying premise is that the addition of these extra layers will result in the layers' progressive learning of more complicated features. As an example, in the case of image detection, the first layer may identify edge detection, the second layer may identify texture identification, the third layer may learn object detection, and so on. On the other hand, the

conventional Convolutional Neural Network (CNN) model is encumbered with a maximum threshold for depth.

Being a machine learning subset, the DL will require a huge number of labeled data for model training. Typically, the term “deep” will signify the neural network’s number of hidden layers, for example, The ResNet is 8 times as deep as the VGG-Net, having 152 layers. Because to the rise in computational power, low-cost hardware, and open-source algorithms as well as the emergence of Big Data [5, 6], it has resulted in the CNNs garnering tremendous popularity as well as interest.

In this work, the deep learners like VGG16, VGG19, Resnet18, Resnet34 and Inception are evaluated for the classification of mammograms as Benign or Malignant. The paper is organized as follows: the related work in the literature has been explained in Section 2. Section 3 explains all the techniques used in this investigation. Section 4 discusses the findings of the experiments, and Section 5 closes this study.

2. LITERATURE SURVEY

Gardezi et al. [7] had presented a technique for classifying the abnormal and normal tissues in mammograms with utilization of a deep learning approach. Mammogram ROIs from the IRMA dataset was deployed with VGG-16 CNN deep learning architecture having convolutional filter of (3×3). The first FC layer was used for the deep feature matrix’s computation. This technique was able to yield 100% classification accuracies having AUC 1.0.

Carneiro et al. [8] had given the detailed description of an automated methodology to analyze the unregistered MLO and cranio-caudal CC mammography views for estimation of the breast cancer risk for the patient development. The key novelty of proposed method involved the utilization of DL methods for resolving the issue of joint classification using the segmentation maps of breast lesions (that is, micro-calcifications and masses) and the unregistered mammogram views. The biomedical imaging field’s chief standard approach involved classification of the individual lesions. Moreover, the proposed methodology had accomplished accurate outcomes despite the utilization of automatic detection methods for bulk and micro-calcification provide segmentation maps. The INbreast and DDSM dataset were used for testing the semi-automated method. The results of this methodology indicated that the VUS was over 0.9 for a three-class issue and over 0.7 for a fully automated method.

Becker et al. [9] had assessed the DL with ANNs for an independent and dual-center mammography dataset for breast cancer detection. There was selection of patients with cancer as well as a matched control cohort (n=35×2) from the publicly available dataset. Moreover, this external dataset was used for testing the trained neural network’s performance. The test dataset was assessed by three different radiologists (with each individual having an experience of 3, 5, and 10 years, respectively). The second step involved training the ANN. This second test dataset was also assessed by the aforementioned radiologists. There was comparison of the areas under the ROC between readers and ANN. Statistical significance was offered to a Bonferroni-corrected P value of less than 0.016. For general image analysis, the existing highly advanced artificial neural networks were capable of cancer detection in mammography with an accuracy which was same as that of the radiologists, even in a group similar to a

screening for breast cancer but with a low prevalence.

Sarwinda et al. [10] had analyzed a DL approach, ResNet architecture, for detecting colorectal cancer. Researchers were prompted to deploy this proposed method of deep learning classification for medical image analysis due to its outstanding performance. Images of colon glands were used to train ResNet-18 and ResNet-50 in the research. These models were trained or categorization of the colorectal cancer as either malignant or benign. Assessment of the prototypes was done on three distinct kinds of testing data (that is, 20%, 25%, and 40% of whole datasets). It was confirmed from the empirical outcomes that, when compared against the application of ResNet-18, The ResNet-50 programme was the most accurate, sensitive, and valuable in terms of accuracy, sensitivity, and value of specificity for three distinct kinds of testing data. It was demonstrated from this study that the deep learning method had the ability to accomplish reliable as well as repeatable outcomes for the analysis of biomedical images [11].

Tsochatzidis et al. [3] had examined the CNNs’ efficacy in the CAD of breast cancer. Highly advanced CNNs were trained as well as assessed Malignant and benign mass lesions were represented in two separate mammographic datasets. Two distinct training scenarios were used to address each examined network’s performance evaluation: while the first scenario had dealt with the network’s initialization with pre-trained weights, the second scenario had dealt with random network initialization. It was evident from the comprehensive simulated outcomes that, in comparison to training from scratch, fine-tuning a pretrained network was able to accomplish an excellent performance.

3. METHODOLOGY

Structure-wise, CNNs are no different from other types of neural networks. Non-linear, pooling, and loss functions may be found in the final Fully Connected (FC) layer of a basic CNN design, which includes a convolutional layer, a non-linear layer, and a pooling layer. Alternatively, the result might be a probability of classifications that best describe the picture (for example, malignant, benign, or normal). The Conv is the input layer and image with a fixed size is given as input. The image size is considered to have fixed width (W1), height (H1) and depth (D1), where the depth will indicate the number of channels (for example, D1=3 for a RGB image). The Conv will have F of size N×N×E1, in which N will be lesser than the image’s size, and E will be equivalent to the number of channels (for example, when the size is 5×5×3, the image’s width as well as height will be 5 pixels, and its depth will be 3 due to color channels). At the time of the convolution operation, each F filter will convolve with the image to yield K feature maps with X2×I2×E2 in volume size, in which $X2=G2=(X1-G+2q)/R+1$, and R will indicate the number of strides, $E2=G$, and q will indicate. The amount of zero padding. Each of the maps in this section is a feature map will get applied with a non-linear activation function (for example, ReLU). However, the non-linear activation function will have no impact on the volume size (W2×H2×D2). Upon the ReLU’s application, a down-sampling operation, referred to as Pool, will be employed along the resultant feature map’s spatial dimensions (namely, width and height). After the pooling, it is possible to have many layers of FC which are able to evaluate the class scores [6].

This section details on Visual Geometry Group (VGG), Residual Network (ResNet) and inception.

3.1 Visual geometry group (VGG)

Visual Geometry Group (by Oxford University) had invented VGGNet [11]. It was essential to fully comprehend VGGNet since it formed the basis of the construction of majority of the contemporary image classification models. The process of classifying mammograms using VGG is shown in Figure 1.

Table 1. VGG configuration

VVG16	VGG19
Input Image (224×224)	Input Image (224×224)
Conv3-64	Conv3-64
Conv3-64	Conv3-64
Maxpool	Maxpool
Conv3-128	Conv3-128
Conv3-128	Conv3-128
Maxpool	Maxpool
Conv3-256	Conv3-256
Conv3-256	Conv3-256
Conv3-256	Conv3-256
maxpool	maxpool
Conv3-512	Conv3-512
Conv3-512	Conv3-512
Conv3-512	Conv3-512
Maxpool	Maxpool
Conv3-512	Conv3-512
Conv3-512	Conv3-512
Conv3-512	Conv3-512
maxpool	maxpool
Fully Connected	Fully Connected
Fully Connected	Fully Connected
Fully Connected	Fully Connected
Softmax	Softmax

The VGG net is defined. The network’s input will be an image sized $224 \times 224 \times 3$ pixels. The first two layers 3×3 filters and 64 channels as well as same padding. Then a stride max pool layer is applied on top of it (2, 2) will be followed

by 2 Conv having 256 filters of (3, 3) filter size. Once again, just like in the earlier layer, there will be a stride pooling maximum layer (2, 2) followed by 2 Conv having 256 filters of (3, 3) filter size. Finally, two distinct sets of 3 convolution layer as well as a max pool layer. The most popular VGG models are VGG16 (with 16 layers) and VGG19 (with 19 layers). The VGG-19 has one more layer in each of the three convolutional blocks (as shown in Table 1). 224×224 , 227×227 , 256×256 , and 299×299 are typical picture sizes for a CNN trained on ImageNet; in this work, 224×224 images are used. The architecture for VGG16 and VGG19 is shown in Table 1.

Upon stacking convolution and max-pooling layers, a feature map is attained. This output can be flattened to turn it into a feature vector. Later on, these are moved on to the FC layers. After the last FC layer, the output will get transferred to Softmax layer for the classification vector’s normalization. All the hidden layers will employ ReLU as their activation function. ReLU will result in quicker learning as well as will mitigate the vanishing gradient problem’s likelihood. Due to these reasons, it will have more computational efficiency.

Challenges of the VGG:

A long time is needed for its training (2-3 weeks were required for training the original VGG model on Nvidia Titan GPU).

VGG-16 trained picture with a 528 pixel dimension. Because it consumes so much disc space and bandwidth, it will be rendered inefficient by net weights.

3.2 Residual network (ResNet)

In 2015, introduction of a novel architecture termed the Residual Network (ResNet) was made by researchers at Microsoft Research. ResNet network will employ a VGG-19-inspired 34-layer plain network architecture along with the addition of shortcut connection. Later on, these shortcut connections will transform the architecture into the residual network.

In comparison to other architectural models, the ResNet model is quite beneficial since its performance will not decrease despite the architecture’s increase in depth. Furthermore, it has lighter computational calculations as well as better network training capability.

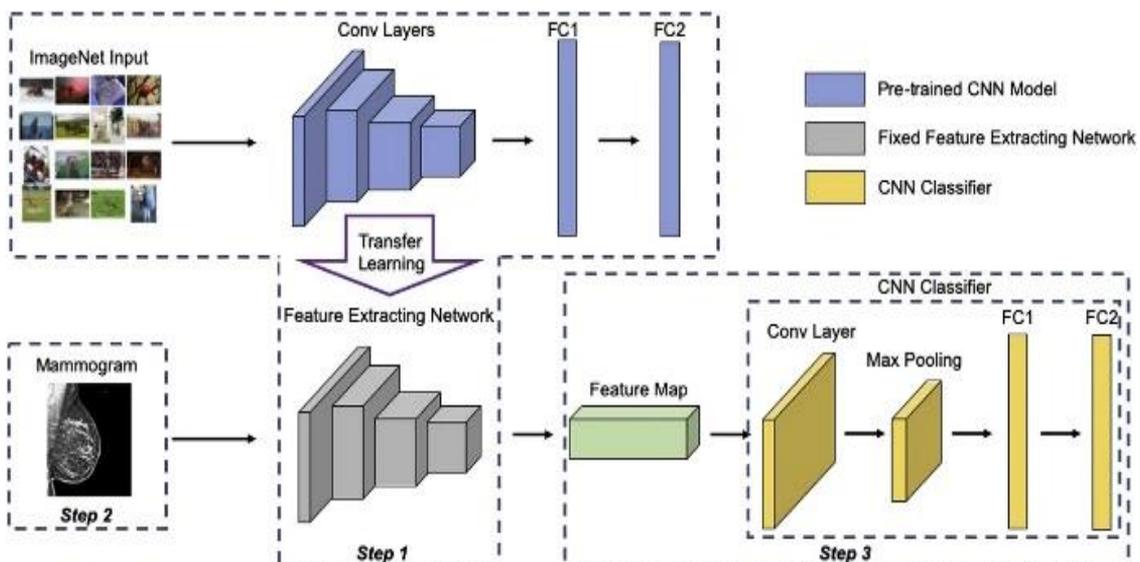


Figure 1. Framework for analysing Mammogram with VGG-16

Accomplishment of residual block on the ResNet is possible if the dimensions of the input and output data are same. It's also worth noting that the blocks of ResNet will have either two or three unique layers, depending on the kind of network (in case of ResNet-50 and ResNet-101 networks). Two first layers of the ResNet architecture mimic Inception's convolution 7×7 and max-pooling 3×3 with stride number 227 convolutions.

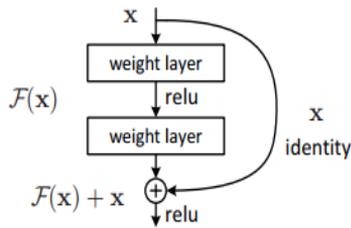


Figure 2. Deep residual learning for image recognition

With these Residual blocks' introduction, there is mollification of problem of training very deep networks. Moreover, these blocks are the ResNet model's constituents. Upon observation of the above Figure 2, it is seen that there seems to be a direct connection which will skip certain layers of the model. Referred to as the 'skip connection', this connection is the residual blocks' core. This skip connection will result in a different output. When there is the skip connection's absence, input 'X' will undergo multiplication with the layer's weights, and later, will be added with a bias term.

Upon utilization of the activation function, $f()$, the output can be expressed as $G(x)$:

$$I(y)=g(xy+c) \text{ or } I(y)=g(y)$$

The output will increase as a result of a new skip connection method $G(x)$ can be altered as below:

$$G(y)=g(y)+y$$

However, there may be a variation on the input's either a convolutional layer or a pooling layer. These two methods will allow us to deal with the issue at hand:

By adding padding with the skip connection, you may expand Zero's size.

One may add 11 convolutional layers to the input to make the dimensions equal. In some cases, the result will look like this:

$$G(y)=g(y)+x1.y$$

Here, there is an inclusion of an extra parameter, $x1$, while the first approach did not make use of any additional parameter. These skip connection techniques in the ResNet will facilitate alternative gradient flow shortcut pathways and so fix the issue of the deep CNNs' vanishing gradient. Furthermore, in the event that any layer results in hurting the architecture's performance, the skip connection will aid in skipping that particular layer by means of regularization. The architecture has a 34-layer plain network which is inspired by the VGG-19 wherein there is the addition of either the shortcut connection or the skip connections. In the Table 2, the dimension of the convolutional kernels at every point in the structure of ResNet18 and ResNet34 is given 3×3.

Table 2. Structure of ResNet

ResNet18 layer	ResNet34 layer
7×7, 64, stride 2	
3×3 maxpool, stride 2	
$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 3$
$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 4$
$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 6$
$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 3$
Average pool	
Fully connected	
Softmax	

3.3 Inception (GoogLeNet)

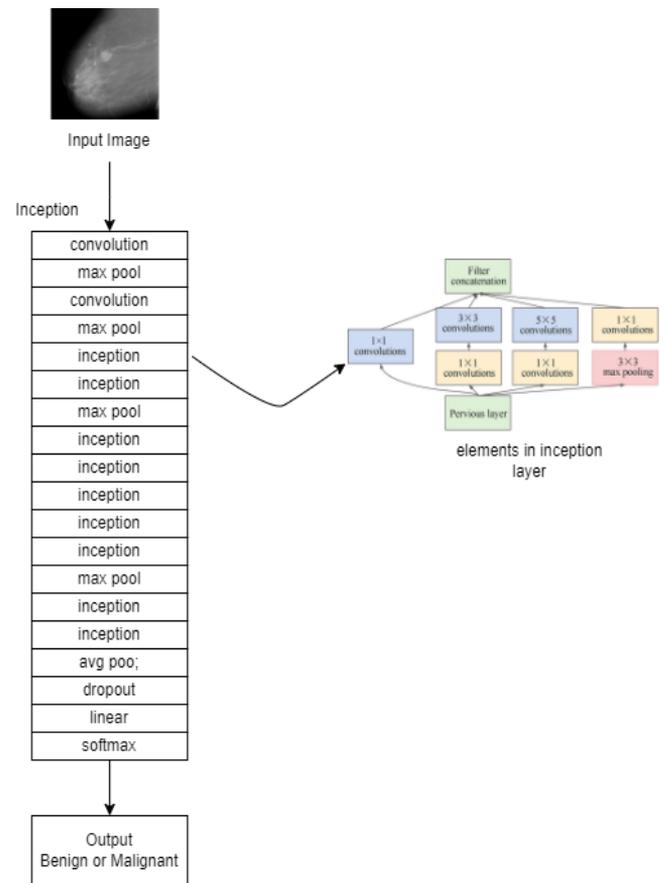


Figure 3. Framework of inception network for classifying mammograms

Inception/GoogLeNet is the Inception module's first implementation. This module's underlying concept is based on the authors' findings which pertain to the dense components' approximation of a local sparse structure. A multi-layered network was to be constructed by finding and repeating the best local structure. The Inception module has four unique branches with the same input as its constituents. The first branch performs a linear change on the input channels by filtering the input with a 1*1 convolution. When it comes to convolution, the second branch will use a kernel size of 3×3 for the first kernel and 1x1 for the second kernel, the third branch will execute the 1×1 kernelled convolution followed by convolutional layer with a kernel of size 5×5. On the other hand, the fourth branch will carry out convolution with 1 1 kernels after max-pooling. At last, every node's output will get

concatenated and then, fed as the next block's input. Stacking of nine Inception modules will result in the Inception's construction. Selected locations will be situated with a layer of maximum pooling between the modules of inception so as to minimize the feature maps' dimensionality. The incorporation of auxiliary classifiers is one of the notable features of Inception. It is assumed that a CNN's middle layers must yield discriminative features. To this end, they have included simple classifiers (two fully connected layers as well as one Softmax layer) which have the ability to perform operations on a network intermediate point's produced feature. In the back-propagation step, these classifiers' judgments are used to evaluate new gradients, which put up to the teaching of the relevant convolutional layers. There will be elimination of the auxiliary classifiers during the inference time. Figure 3 shows the architecture of Inception.

4. RESULTS AND DISCUSSION

To evaluate the various deep learning techniques, CBIS-DDSM: Breast Cancer Image Dataset used. Number of samples - 550 Benign and 625 Malignant. Figure 4 shows the sample images 1 and 2 respectively. Tables 3 to 5 and Figures 5 to 7 shows the Classification Accuracy, recall and precision respectively for VGG16, VGG19, Resnet18, Resnet34 and Inception.

Table 3. Classification accuracy for inception

Classification Accuracy	
VGG16	88.51
VGG19	89.19
Resnet18	89.53
Resnet34	90.21
Inception	91.49

Table 4. Recall for inception

	Recall for Benign	Recall for Malignant
VGG16	0.8964	0.8752
VGG19	0.9018	0.8832
Resnet18	0.9055	0.8864
Resnet34	0.9109	0.8944
Inception	0.9236	0.9072

Table 5. Precision for inception

	Precision for Benign	Precision for Malignant
VGG16	0.8634	0.9056
VGG19	0.8717	0.9109
Resnet18	0.8752	0.9142
Resnet34	0.8836	0.9194
Inception	0.8975	0.931

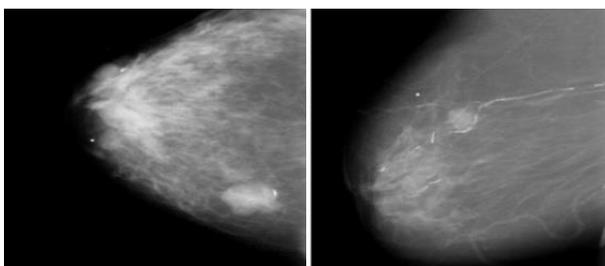


Figure 4. Sample images

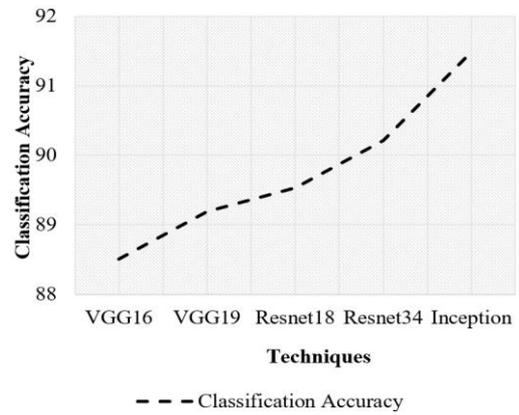


Figure 5. Classification accuracy for inception

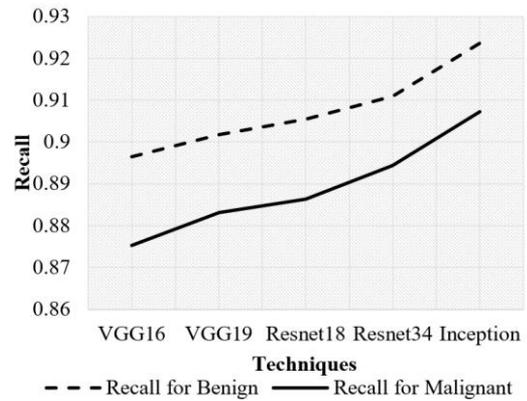


Figure 6. Recall for Inception

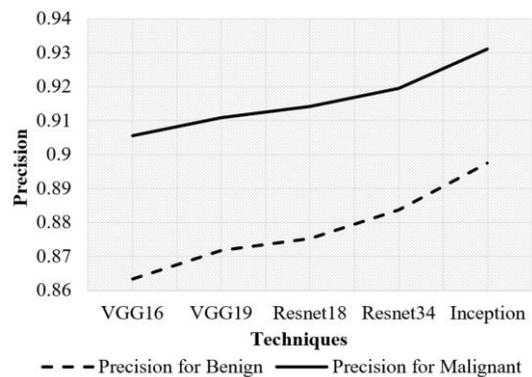


Figure 7. Precision for inception

Figure 5 shows that the Classification Accuracy of Inception performs better by 3.3%, by 2.55%, by 2.17%, and by 1.41% respectively than VGG16, VGG19, Resnet18, and Resnet34.

Figure 6 shows that the recall of Inception performs better by 2.99%, by 2.39%, by 1.98%, and by 1.4% respectively than VGG16, VGG19, Resnet18, and Resnet34 for Benign. Also the recall of Inception performs better by 3.6%, by 2.7%, by 2.32%, and by 1.42% respectively than VGG16, VGG19, Resnet18, and Resnet34 for Malignant.

Figure 7 shows that the precision of Inception performs better by 3.87%, by 2.92%, by 2.52%, and by 1.56% respectively than VGG16, VGG19, Resnet18, and Resnet34 for Benign. Also the precision of Inception performs better by 2.77%, by 2.2%, by 1.82%, and by 1.25% respectively than VGG16, VGG19, Resnet18, and Resnet34 for Malignant. Figure 8 shows sample output for a sample image.

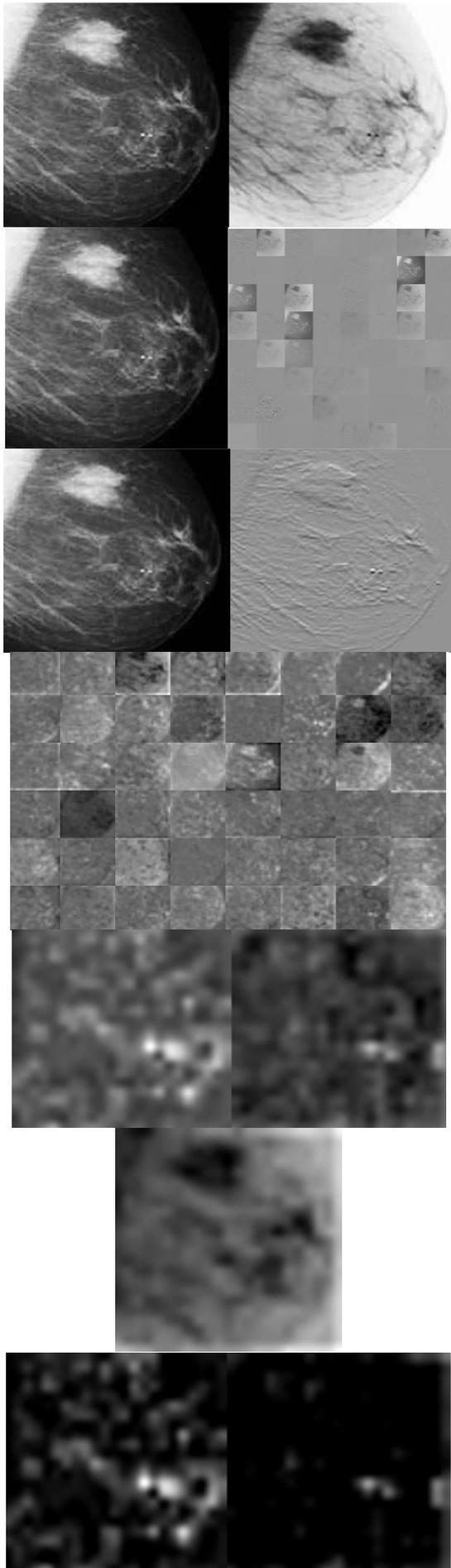


Figure 8. Output images of sample breast image

5. CONCLUSIONS

Among women, Cancer deaths from breast disease are on par with those from lung cancer as the second most prevalent cause. It has made important contributions towards lowering the death rate by means of early cancer detection. Mammography is an extensively employed in the screening method of cancer caused in breast. In this work, the deep learning techniques are used for classifying mammograms. The VGG16, VGG19, Resnet18, Resnet34 and Inception are evaluated. Result shows that the Classification Accuracy of Inception performs better by 3.3%, by 2.55%, by 2.17%, and by 1.41% respectively than VGG16, VGG19, Resnet18, and Resnet34. The recall of Inception performs better by 2.99%, by 2.39%, by 1.98%, and by 1.4% respectively than VGG16, VGG19, Resnet18, and Resnet34 for Benign. When compared against very deep plain networks, the ResNet has been found to achieve better outcomes with residual mapping as well as shortcut connections. Moreover, the training is much easier in the ResNet. Further investigations for optimizing the deep learners to enhance the mammogram classification needs to be carried out.

REFERENCES

- [1] World Health Organization. <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>, accessed on 1 January 2020.
- [2] Meenalochini, G., Ramkumar, S. (2021). Survey of machine learning algorithms for breast cancer detection using mammogram images. *Materials Today: Proceedings*, 37: 2738-2743. <https://doi.org/10.1016/j.matpr.2020.08.543>
- [3] Tsochatzidis, L., Costaridou, L., Pratikakis, I. (2019). Deep learning for breast cancer diagnosis from mammograms-A comparative study. *Journal of Imaging*, 5(3): 37. <https://doi.org/10.3390/jimaging5030037>
- [4] Arefan, D., Mohamed, A.A., Berg, W.A., Zuley, M.L., Sumkin, J.H., Wu, S. (2020). Deep learning modeling using normal mammograms for predicting breast cancer risk. *Medical Physics*, 47(1): 110-118. <https://doi.org/10.1002/mp.13886>
- [5] Abdelhafiz, D., Yang, C., Ammar, R., Nabavi, S. (2019). Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinformatics*, 20(11): 1-20. <https://doi.org/10.1186/s12859-019-2823-4>
- [6] Srinivas, S., Sarvadevabhatla, R.K., Mopuri, K.R., Prabhu, N., Kruthiventi, S.S., Babu, R.V. (2017). An introduction to deep convolutional neural nets for computer vision. In *Deep Learning for Medical Image Analysis*, pp. 25-52. <https://doi.org/10.1016/B978-0-12-810408-8.00003-1>
- [7] Gardezi, S.J.S., Awais, M., Faye, I., Meriaudeau, F. (2017). Mammogram classification using deep learning features. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 485-488. <https://doi.org/10.1109/ICSIPA.2017.8120660>
- [8] Carneiro, G., Nascimento, J., Bradley, A.P. (2017). Automated analysis of unregistered multi-view mammograms with deep learning. *IEEE Transactions on Medical Imaging*, 36(11): 2355-2365. <https://doi.org/10.1109/TMI.2017.2751523>

- [9] Becker, A.S., Marcon, M., Ghafoor, S., Wurnig, M.C., Frauenfelder, T., Boss, A. (2017). Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Investigative Radiology*, 52(7): 434-440.
- [10] Sarwinda, D., Paradisa, R.H., Bustamam, A., Anggia, P. (2021). Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Computer Science*, 179: 423-431. <https://doi.org/10.1016/j.procs.2021.01.025>
- [11] Hammad, I., El-Sankary, K. (2018). Impact of approximate multipliers on VGG deep learning network. *IEEE Access*, 6: 60438-60444. <https://doi.org/10.1109/ACCESS.2018.2875376>