# A transfer learning approach to implementation of pretrained CNN models for Breast cancer diagnosis

**[1]Rajan Prasad Tripathi, [2]Sunil Kumar Khatri, [3]Darelle Van Greunen**

[1]*Dept. of IT and Engineering, Amity University in Tashkent, Tashkent, Uzbekistab, rajantripathi22@gmail.com*
[2]*Dept. of IT and Engineering, Amity University in Tashkent, Tashkent, Uzbekistab, skkhatri@amity.edu*
[3]*Center for Community Technologies, Nelson Mandela University, Port Elizabeth, South Africa,*
*Darelle.vanGreunen@mandela.ac.za*

## Abstract

In this study, different  pre trained model based on CNN for breast cancer picture segmentation and classification has been analysed . Different models including InceptionV3, DenseNet121, ResNet50, VGG16 and MobileNetV2 models are used to categorize Mammographic Image Analysis Society (MIAS), This approach will help radiologist's aide in early recognition and increment the productivity of our framework. Broad exploratory outcome showed the predominant presentation accomplished on account of calibrating a pretrained network. In our study we found out VGG16 and Resnet101 perform better in predicting the cancer classification, VGG16 performs slightly better than Resnet101 in predicting the malignant Tissue. Furthermore, in the study it was found out that VGG 16 and Resnet delivers the greatest accuracy.

**Keywords**: CNN, Breast Cancer, Pretrained network,ANN,MIAS,Mammography.

## I. INTRODUCTION

Malignancy of the breasts is the most frequent kind of malignancy in women, in accordance with a recent study [1], and accounts for close to 33% of all recently distinguished growths in the United States [2]. Breast malignancy mortality is also important and usually accounts for seventeen percent of all cancer deaths [3]. Accurate diagnosis and assessment of the early periods of breast cancers are important to minimize mortality. Mammography has until now recently been a very useful strategy for screening large populations. However, effective detection associated with breast skin lesions absolutely founded on mammography information is perplexing and relies by and large upon the skill of radiologists, creating a high proportion of false-positive results and extra tests. Masu [4].

Computer-aided diagnosis and diagnostic (CAD) technology is as of now being utilized to offer significant help to the radiologist's decision-making process. This sort of solution can significantly reduce the energy required to check lesions in specialized medical practice while lowering the number of false benefits wasteful and inconvenient biopsies. CAD systems for mammography are able to do two different tasks: identifying skin lesions appealing in mammography (CADe) and what's more, sorting out identified lesions (CADx). Order as harmless or malignant.

Deep learning has been seen as a notable discovery in recent years, as it shows cutting edge performance in many machine learning questions such as object location and arrangement. Unlike standard machine learning methods, which require hand-made feature removal steps to depend on domain experience, deep learning techniques take into consideration the intended outcome and the appropriate features from the input data.

Adaptively train extraction techniques. This reduces the hassle of deciding and evaluating useful features while minimizing the reproducibility of the approach. Due to the fact the debut of deep learning, various studies using serious architecture have recently been published. [5][31]. The most well-known structure of CNN. Arevalo et 's. [6] Various CNNs were tested in comparison with two hand-made instructions for size diagnostic purposes. Their very own tests were run on the BCDRFM dataset. They saw performance improve through a combo of learning and hand-crafted rendering. On the other hand, mcdougal did not look at the performance of the pre-trained network and used a less difficult CNN design.

Carneiro et al. [7] developed a fine-tuned preformat Convolution Neural Network using unrecorded mammogram images and distinct masses and microcalcifications. They examined the risk of cancer of the breast. They established that the pre-trained models beat the arbitrarily evolved models. Huynh et 's. [8] implemented pre-trained AlexNet [9] minus any additional improvement to fulfill the group diagnostic challenge. They will test the efficiency of classification using attributes from different intermediate amount communities. Jiao et al. [10] propose an approach where a pre-trained CNN has been refined by partial application of the DDSM repository. Then this properties of the mass are from the several levels of the model. With this method, "advanced" and "intermediate" characteristics were created with independent scales, correspondingly.. Levy and Jain [11] uncovered mammograms using GoogleNet.. Particularly, they researched the aftereffect of information setting, showing that cutting bigger jumping boxes of ordinary size round the illness twisted substantially more viable than cutting with equivalent cradles. Rollator walker et al. [12] developed and trained their network to categorise breast masses from the beginning. Their network has 28 fully connected and complex levels and it is powered by the suggested ROI found by a oneshot detector. They directed their trials in the MIAS data source. Rampon et ing. [13] deployed a arranged of lightly altered versions of AlexNet pre-trained and funely-tuned on CBISDDSM. All through the surmising system, they pick 2

best performing models and blend their expectations.

Most sophisticated programs encourage the use of pre-trained networks rather than training from the beginning. Nevertheless, modern networks are created and analyzed on datasets that are much diverse, variable in character, and larger by a quantity of purchases of magnitude than currently available mammography datasets. As a result, the usefulness and intricacy of the organizations can extend a long ways past the limits of more modest datasets, prompting huge undesirable aftereffects while preparing from the very beginning. While a result, various publications emerged that the authors strongly suggested to train from damage. Considering the abovementioned, in this investigation, we check out the presentation of numerous organizations. A large number of us look at the presentation of every single organization in two situations: the first is to get started on training using the pre-trained weights even though the second is for the networks to get started on using the weight loads random. Other report is organized as follows: Section 2 examines the genuine learning structures applied in the current review to order mammogram pictures as harmless or dangerous. Segment 3 subtleties of the experiment related to the pertinent outcomes. Finally, in segment 4, the conclusion is discussed.

## 2. Methodology

### 2.1. Convolutional Neural Networks

### 2.1.1. VGG

Within [16], the researchers investigated the impact of local area profundity while to get convolutional filtration suprisingly low. They confirmed that a sizable enhancement could be executed by increasing the depth to 15-18 layers. The suggestions to the convolutional layer is an picture of set size $224 \times 224$. The is processed by way of a collection of convolutional levels with ReLU service where the filter systems have a very modest receptive field (3 X 3). ) was applied. The convolutional step is also established to 1. Space aggregation is executed using five maximum sharing layers, executed after several levels of complexity. Since with AlexNet, a fully attached three-layer

stack is built over the network's composite component. The particular good thing about VGG is that by putting numerous convolutional layers with little centres, the viable responsive field of the framework is expanded while decreasing how many rules contrasted with involving less mind-boggling levels with bigger induration for the same procuring school. The editors considered different blends of different absolute depths (9, 11, fourth there's 16 and 19 layers) (9, 11, fourth there's 16 and 19 layers) (9, 11, fourth there's 16 and 19 layers) (9, 11, fourth there's 16 and 19 grades).. Some arrangements incorporates an additional layer of LRN. As it has been studied in the review, the best outcomes are gotten for profundities from fifteen to nineteen. The architecture of VGG16 is shown in Figure 2.
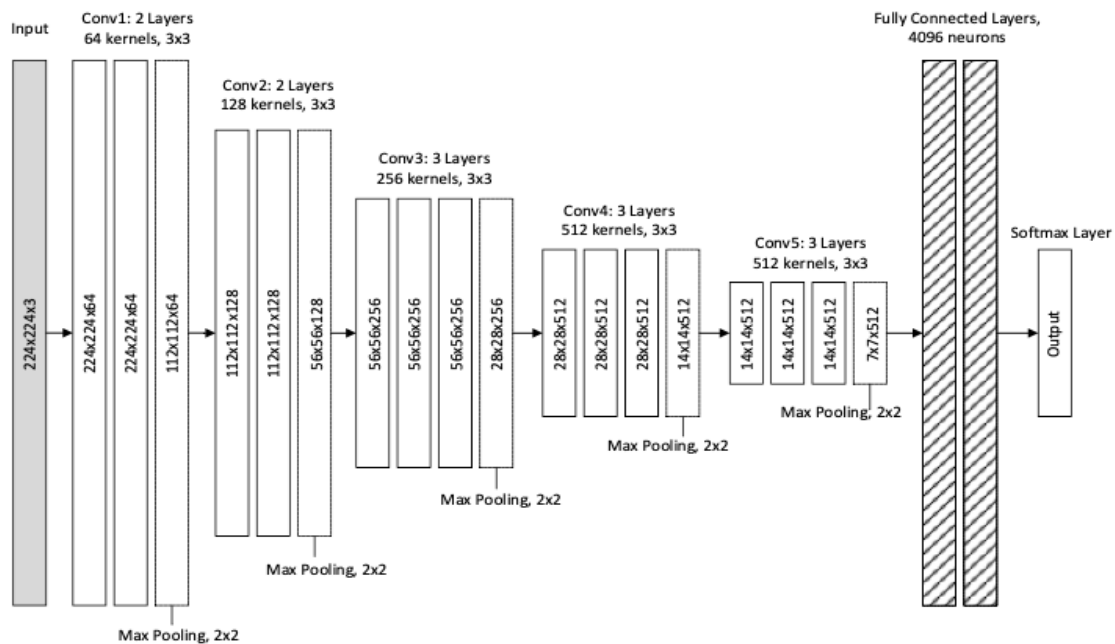


Fig1.

*VGG-16 Structure[30]*

Fig1Alt text. The block diagram shows input flatten layer of 224*224*3 is given to convolution layer and is fed to maxpooling layer in total there are 5 pair of Convolution layer and maxpooling layer which is fed to 4096 fully connected dense layer and finally it is fed to output layer.

### 2.1.2 GOOGLE/NET INCEPTION

GoogLeNet [17] is the main rendition to utilize the Invention module. The essential standard behind such a module depends on the creators' revelation showing the way that inadequate nearby constructions can be addressed by thick parts. Their goal was going to determine the optimum local structure and reconstruct it, creating a multi-layer system[17]. GoogLeNet is created by stacking seven Inception modules. at some stage in described regions, a most shared overlaying is positioned between starter modules to reduce the size of the function map.

Losing caused by the assessments of these classifiers is utilised throughout the backpropagation stage to give extra gradient that help to the education of the appropriate convolutional levels. At inference time the extra classifiers are removed.
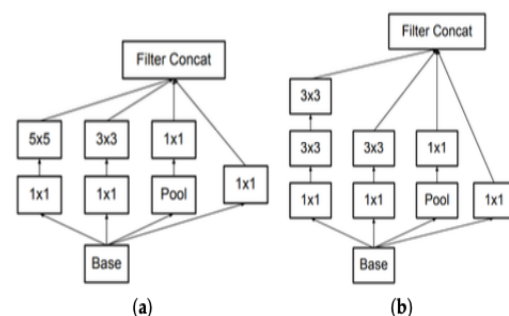


Fig2. *(a) Inception module of GoogLeNet; (b) Inception-v2 module [30].*

Fig2 Alt text. Part a shows the connected layers of Inception module of GoogleNet and Part b shows the connected layers Inception V2

In later articles [18], a more current variant of the Inception module was really delivered, alongside somewhat unique organization structures. The creators recommended Batch Normalization (BN) and put it in the Start-up Network. BN is a standardization strategy intrinsic in the model design that performs standardization for every little preparation cluster. The creators contend that BN offers higher learning rates and easier introduction techniques without adverse results.In addition, a linear transformation was performed, whose parameters were learned during the learning process. The network used in [18], named Inceptionv2, is a modest modification of GoogLeNet. Besides the inclusion of BN, the most significant modification was that the Starter module's 5 X 5 convolutional layers were replaced by two consecutive $3 \times 3$ layers (Fig.2)

### 2.1.3 Residual Networks

It consist of reconstructed convolutional layers to learn input related residual functions[19]. The particular authors suggest that such networks are easy to improve and can be much deeper. Creating a "residual block", as portrayed in [19] is direct: for each and every convolutional level, a "shortcut connection" is established that runs in suite with those levels and performs the mapping. identity. Apart from using shortened cable connections, network design is essentially influenced by VGG network principle. All composite courses have little parts of size 3 X 3 and adhere to two essential plan guidelines for a similar result including graph size, layers with a similar measure of channels; (ii) when how huge the convolutional object is divided, the assortment of channels is multiplied to save the time intricacy for every single layer. The creators concentrated on designs of fluctuating profundities in the scope of 33 and 151 layers.
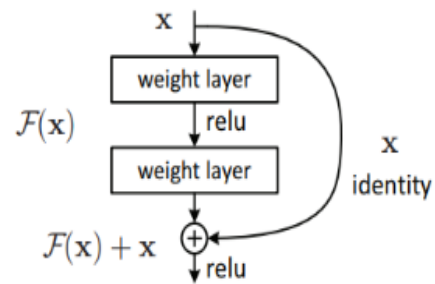


Fig3. *Building block of ResNet [19]*

Fig3 Alt text. It id building block of ResNet which shows how input is fed to weighted layer and activation function is relu and the input is also connected to output

### 2.1.4 InceptionResNet V2

Inception ResNet-v2 is a network that develops the Inception group of models furthermore contains leftover associations, They are somewhat Residual Connections are a kind of skip-association that learn lingering capacities concerning the layer inputs, as opposed to simply learning unreferenced capacities.

The idea is that it is less complex to streamline the remaining planning than to upgrade the underlying, unreferenced planning.
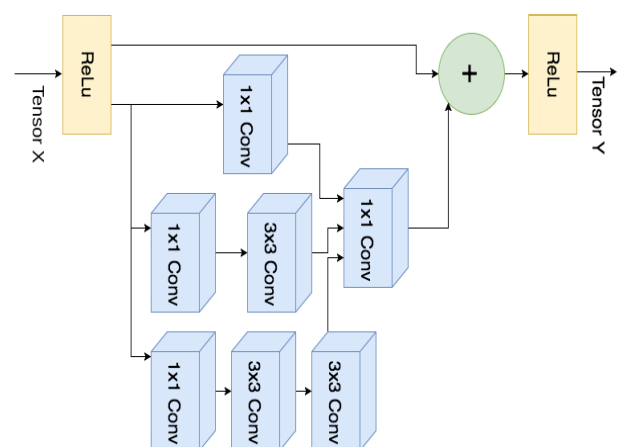


Fig4. *InceptionResNet V2 architecture*

Fig4 Alt text. The figure shows the block diagram of InceptionResNet V2 architecture it shows how the input tensor is also connected to

several Convolution layers before finally feeding to output

### 2.1.5 Inception V3

Introduced by Szegedy et al is a redesigned architecture for Computer Vision that delivers various enhancements including apply Label Smoothing. The inception V3 of course performs better than other inception network . It has been made as modification over the previous version of inception . It has 42 layers and the error rate when compared with other previous model is very less. It was made in 2014. It performs the factorization into smaller convolutions .It also utilises auxiliary classifiers and has efficient grid size reduction technique. It has cumulative convolution which uses an auxiliary classifier to propagate label information further into the network (in combination with the use of batch normalization (batch normalization attempts to reduce the error rate). internal covariate delay and is therefore intended to speed up the training of deep neural networks It does this through a normalization that adjusts the mean and variance of the layer inputs. also have a favourable effect on the gradient flow through the network, by minimizing the dependence of the gradient on the scale of the factors or on their factors. This allows the use of high learning rates significantly more without the risk of bias. Consequently, batch normalization regularizes the model and decreases the requirement for Dropout. We apply a batch normalization layer as follows for a minibatch B:

i. $\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$ ……………..(1)

ii. $\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2$ ……(2)

iii. $\hat{x}_i = \frac{x_i - \mu_B}{\sigma_B^2 + \epsilon}$ ..........(3)

iv. $y_i = \gamma\hat{x}_i + \beta = BN_{\gamma,\beta}(x_i)$….(4)

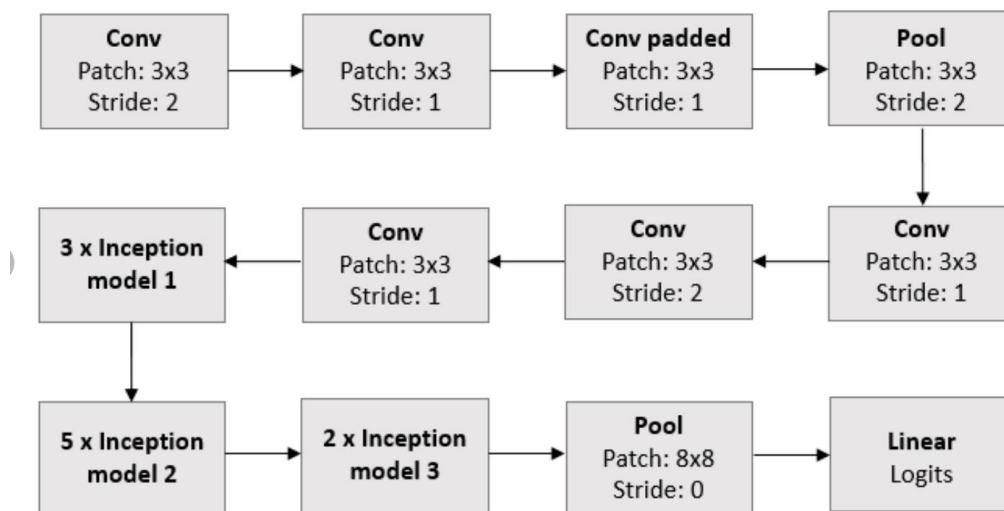Where $\gamma$ and $\beta$ are learnable parameters.



Fig5. *InceptionV3 architecture*

Fig 5 Alt. The figure shows the block diagram Inception V3 architecture and how there's a cascading of different layers involving convolution layer , padding layer and inception model 1,2 and 3.

### 2.1.6 DenseNet

The problem with the traditional convolution neural network is that all the layers receives the input from the previous layer and hence except for the first layer all the layers are connected from the previous layer and now creates the problem of losing gradient descent and as the layer gets more deep there is high probability that some of the information might get lost .Hence for deeper layers we have to use DenseNet , Now it provides the solution to the vanishing gradient descent problem that all the layers are connected to each other and hence the loss of the information is not there.

We analyse our proposed architecture on four extremely competitive standard object recognition tasks (CIFAR10, CIFAR100, SVHN and ImageNet) (CIFAR10, CIFAR100,

SVHN and ImageNet). DenseNets get a significant boost over modern on many of

them, although using less memory and computation for good performance.

| Layers | Output Size | DenseNet-121 | DenseNet-169 | DenseNet-201 | DenseNet-264 |
|---|---|---|---|---|---|
| Convolution | 112 × 112 | 7 × 7 conv, stride 2 | | | |
| Pooling | 56 × 56 | 3 × 3 max pool, stride 2 | | | |
| Dense Block (1) | 56 × 56 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 |
| Transition Layer (1) | 56 × 56 | 1 × 1 conv | | | |
| | 28 × 28 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (2) | 28 × 28 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 |
| Transition Layer (2) | 28 × 28 | 1 × 1 conv | | | |
| | 14 × 14 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (3) | 14 × 14 | [1 × 1 conv, 3 × 3 conv] × 24 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 48 | [1 × 1 conv, 3 × 3 conv] × 64 |
| Transition Layer (3) | 14 × 14 | 1 × 1 conv | | | |
| | 7 × 7 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (4) | 7 × 7 | [1 × 1 conv, 3 × 3 conv] × 16 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 48 |
| Classification Layer | 1 × 1 | 7 × 7 global average pool | | | |
| | | 1000D fully-connected, softmax | | | |

Fig6. *DenseNet  model architecture for ImageNet*

Fig 6 Alt text. DenseNet model architecture series  convolution layer , pooling layer , dense block , transition layer and classification layer.

Fig 7 Alt text. The block diagram of MobileNet V2 Architecture with  1 block stride and 2 block stride is shown

### 2.1.7 MobileNetV2

MobileNetV2 is a convolutional neural network plan that attempts to perform well on cell phones. It is based on an upset lingering structure where the remaining associations are between the stifle layers. The moderate development layer utilizes reasonable profundity-wise convolutions to channel factors as a wellspring of non-linearity. As an aggregate, the plan of MobileNetV2 highlights the first totally convolution layer containing 32 channels, trailed by 19 remaining bottleneck layers.
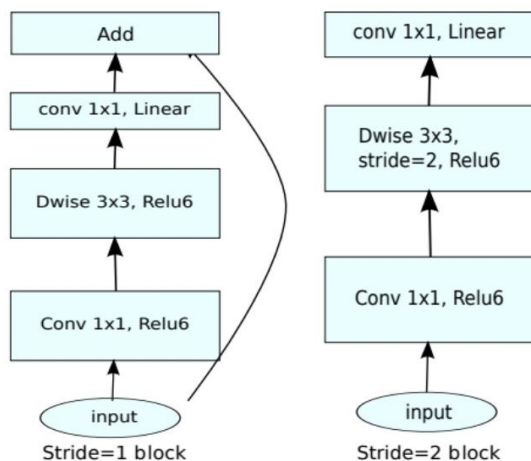
### 2.1.7 Resnet

Deep neural networks are significantly more diligently prepared. We give a residual learning structure to work with the preparation of organizations a lot further than recently utilized. We are purposefully reformatting classes as gathering lingering capacities regarding class inputs, rather than learning unassigned capacities. We present broad trial proof that these remaining organizations are more straightforward to tune and can accomplish exactness from a lot more noteworthy profundities. On the ImageNet dataset, we investigate lingering networks with an increase to 152 layers, multiple times further than the VGG network yet less mind-boggling. A bunch of these lingering lattices accomplishes a blunder of 3.47% in the ImageNet test suite. This performance has earned it the first place in the 2015 ILSVRC binary rankings. We also do CIFAR10 studies with 100 to1000 classes.



Fig7. *MobileNet V2 architecture*

The profundity of portrayals is of fundamental importance for some, visual ID assignments. Exclusively inferable from our remarkably profound portrayals. we gain a 27 percent significant enhancement on the COCO object identification dataset. Each ResNet block is

either two layers deep (used in small networks like ResNet 18, 34) or 3 layers deep (ResNet 50, 101, 152).

| layer name | output size | 18-layer | 34-layer | 50-layer | 101-layer | 152-layer |
|---|---|---|---|---|---|---|
| conv1 | 112×112 | 7×7, 64, stride 2 | | | | |
| conv2_x | 56×56 | 3×3 max pool, stride 2 | | | | |
| | | $\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$ |
| conv3_x | 28×28 | $\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$ | $\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$ | $\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$ | $\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$ |
| conv4_x | 14×14 | $\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$ | $\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$ | $\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$ | $\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$ |
| conv5_x | 7×7 | $\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$ |
| | 1×1 | average pool, 1000-d fc, softmax | | | | |
| FLOPs | | $1.8\times10^9$ | $3.6\times10^9$ | $3.8\times10^9$ | $7.6\times10^9$ | $11.3\times10^9$ |

Fig8. *ResNet Architecture*

Fig8. Alt text. The figure shows in tabular form the convolution stride followed up by max pool layer for ResNet Architecture.
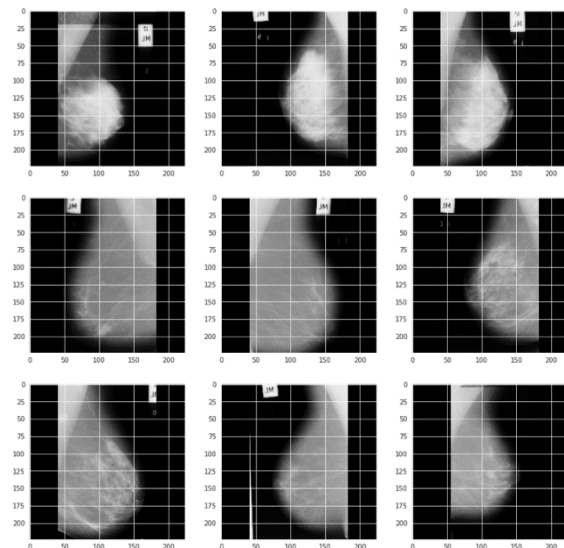
Fig8. *Visualisation of Random 9 images of MIAS dataset*

Fig 8 Alt text. The 9 random images from MIAS data set

## 3.    Dataset

MIAS Dataset

The Mammographic Image Analysis Society (MIAS) is an association of UK research organizations engaged in the comprehension of mammograms and has produced a database of mammographic images. The data is images and labels / annotations for mammography scans.





< **Info.txt (6.42 kB)**

```
REFNUM BG CLASS SEVERITY X   Y   RADIUS
mdb001 G  CIRC  B        535 425 197
mdb002 G  CIRC  B        522 280 69
mdb003 D  NORM
mdb004 D  NORM
mdb005 F  CIRC  B        477 133 30
mdb005 F  CIRC  B        500 168 26
mdb006 F  NORM
mdb007 G  NORM
mdb008 G  NORM
mdb009 F  NORM
mdb010 F  CIRC  B        525 425 33
mdb011 F  NORM
mdb012 F  CIRC  B        471 458 40
mdb013 G  MISC  B        667 365 31
mdb014 G  NORM
mdb015 G  CIRC  B        595 864 68
mdb016 G  NORM
```

Fig9.  *Various features of MIAS dataset*

Fig9 Alt text. The various feature of MIAS data set is shown like REFNUM, BG, Class, Severity X Y Radius

## 4.    Result

(a) Train Data Size = 0.7 * Total_Data

(b) Validation Data Size = 0.21 * Total_Data

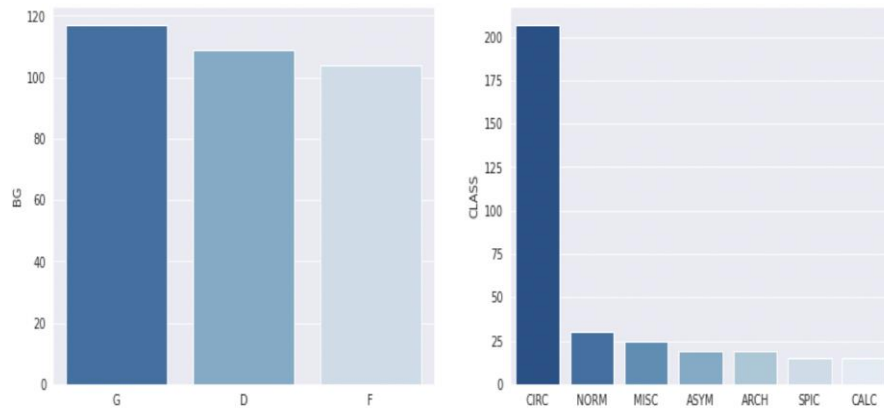(c) Test Data Size = 0.09 * Total_Data

Fig10. *Visualizing the MIAS dataset Class wise bifurcation*

Fig 10 Alt text. The MIAS dataset class wise distribution with respect to G,D and F as well as CRC, NORM, MISC, ASYM, ARCH, SPIC and CALC

## 4.1 Model summary

### 4.1.1   Inception V2

Table1. *Model Summary for InceptionV2*

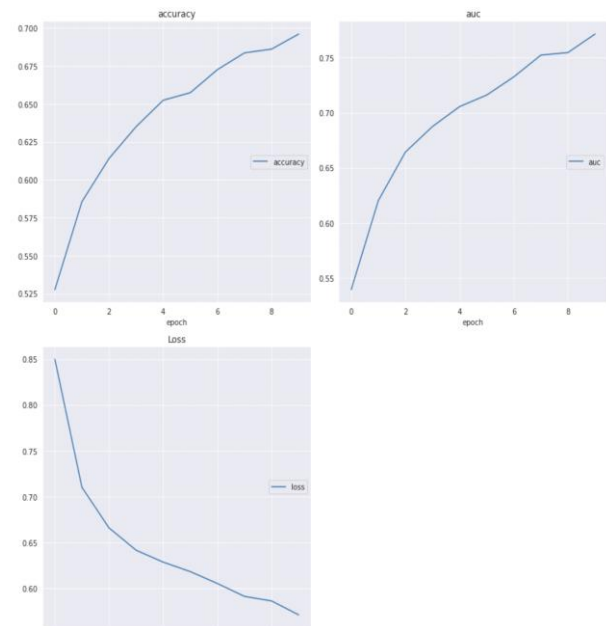| Layer | Output Shape | Parameter Count |
|---|---|---|
| inception_resnet_v2 | None,5,5,1536 | 54336736 |
| Flatten | None, 38400 | 0 |
| Batch normalization 203 | None, 38400 | 153600 |
| dense | None,256 | 9830656 |
| Batch normalization 204 | None,256 | 1024 |
| Activation 203 | None,256 | 0 |
| dropout | None,256 | 0 |
| Dense 1 | None,2 | 514 |
| Total_parameters | 6,43,22,530 | |
| Trainable_parameters | 99,08,482 | |
| Non-trainable_parameters | 5,44,14,048 | |



Fig 11. *Accuracy, AUC and Loss variation curve over the epochs for Inception V2*

Fig 11 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases

Table2. *Model Summary for InceptionV3*

| Inception V3 | | |
|---|---|---|
| Layer | Output Shape | Parameter Count |
| Inception v3 | None,5,5,2048 | 21802784 |
| Flatten 1 | None, 51200 | 0 |
| Batch normalization 299 | None, 51200 | 204800 |
| Dense 2 | None,256 | 13107456 |
| Batch normalization 300 | None,256 | 1024 |

| Activation 298 | None,256 | 0 |
|---|---|---|
| Dropout 1 | None,256 | 0 |
| Dense 3 | None,2 | 514 |
| Total parameters | 3,51,16,578 | |
| Trainable_parameters | 1,32,10,882 | |
| Non-_Trainable parameters | 2,19,05,696 | |

| Trainable parameters | 2,10,37,570 | |
|---|---|---|
| Non-trainable parameters | 1,28,06,464 | |



Fig 12. *Accuracy, AUC and Loss variation curve over the epochs for DenseNet169*

Fig 12 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases

Table4. *Model Summary for Dense121*

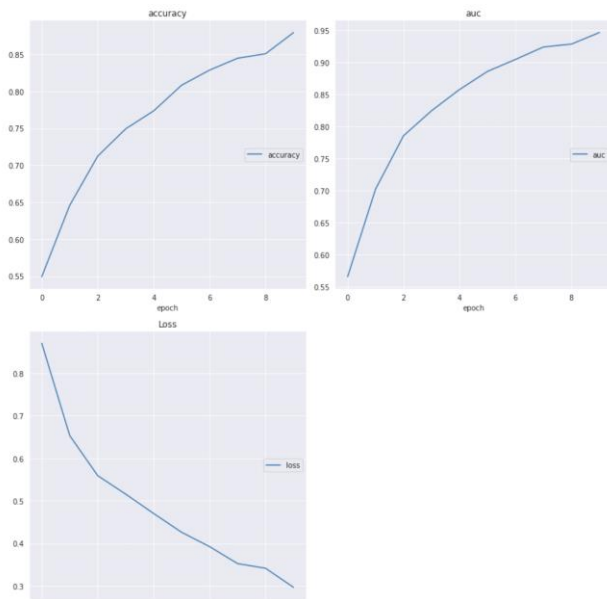| Dense 121 | | |
|---|---|---|
| Layer | Output Shape | Parameter Count |
| Densenet121 | None,7,7,1024 | 7037504 |
| Flatten 3 | None , 50176 | 0 |
| batch_normalization 303 | None , 50176 | 200704 |
| Dense 6 | None,256 | 12845312 |
| batch_normalization 304 | None,256 | 1024 |
| Activation 300 | None,256 | 0 |
| Dropout 3 | None,256 | 0 |
| Dense 7 | None,2 | 514 |
| Total parameters | 2,00,85,058 | |
| Trainable | 1,29,46,690 | |



Fig 12. *Accuracy, AUC and Loss variation curve over the epochs for Inception V3*

Fig 12 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases

Table3. *Model Summary for DenseNet169*

| DenseNet169 | | |
|---|---|---|
| Layer | Output Shape | Parameter Count |
| Densenet169 | None,7,7,1664 | 12642880 |
| Flatten 2 | None, 81536 | 0 |
| Batch normalization 301 | None, 81536 | 326144 |
| Dense 4 | None,256 | 20873472 |
| batch_normalization 302 | None,256 | 1024 |
| Activation 299 | None,256 | 0 |
| Dropout 2 | None,256 | 0 |
| Dense 5 | None,2 | 514 |
| Total parameters | 3,38,44,034 | |

| parameters | | |
|---|---|---|
| Non-trainable parameters | 71,38,368 | |

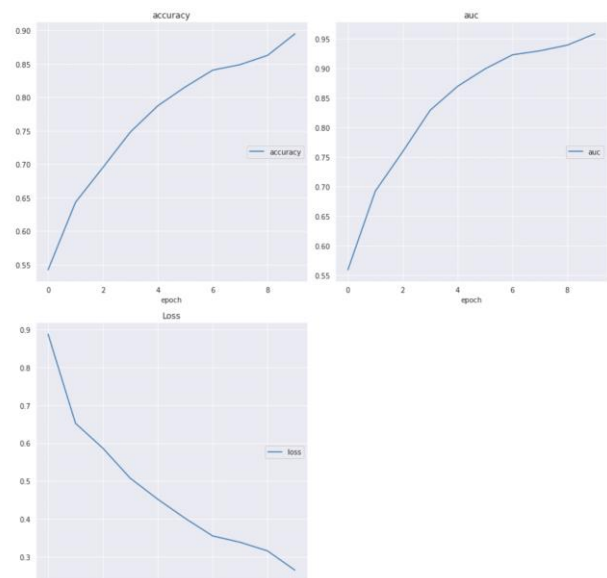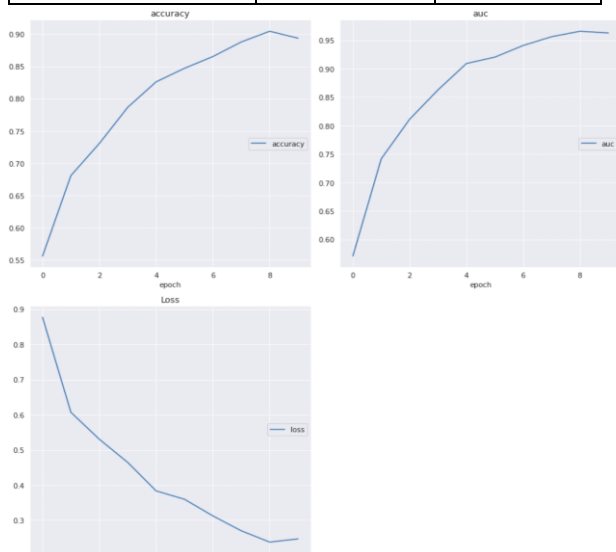| Non-trainable parameters | 23,83,936 | |
|---|---|---|



Fig 13. *Accuracy, AUC and Loss variation curve over the epochs for DenseNet 121*

Fig 13 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases

Table5. *Model Summary for MobileNet V2*

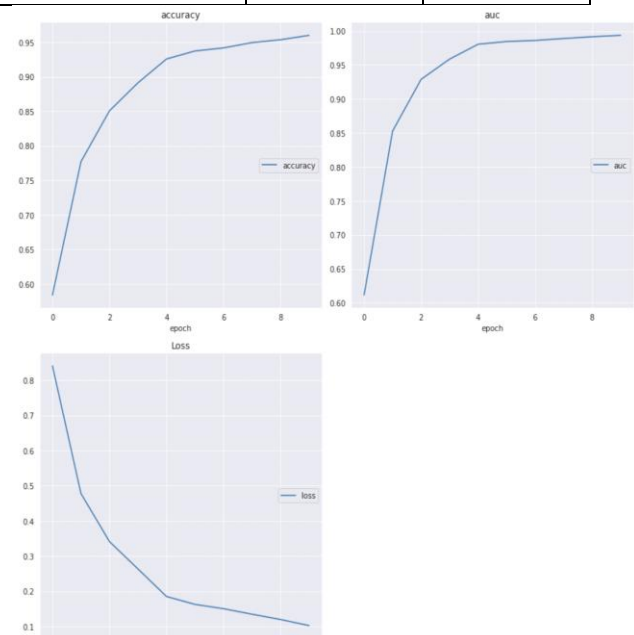| MobileNetV2 | | |
|---|---|---|
| Layer | Output Shape | Parameter Count |
| mobilenetV2 1.00_224 | None,7,7,1280 | 2257984 |
| Flatten 4 | None, 62720 | 0 |
| Batch normalization 305 | None, 62720 | 250880 |
| Dense 8 | None,256 | 16056576 |
| Batch normalization 306 | None,256 | 1024 |
| Activation 301 | None,256 | 0 |
| Dropout 4 | None,256 | 0 |
| Dense 9 | None,2 | 514 |
| Total_parameters | 1,85,66,978 | |
| Trainable_parameters | 1,61,83,042 | |



Fig 14. *Accuracy, AUC and Loss variation curve over the epochs for MobileNet V2*

Fig 14 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases

Table6. *Model Summary for VGG-16*

| VGG-16 | | |
|---|---|---|
| Layer | Output Shape | Parameter Count |
| Vgg_16 | None,7,7,1280 | 2257984 |
| Flatten 5 | None , 62720 | 0 |
| Batch normalization_ 07 | None , 62720 | 250880 |
| Dense 10 | None,256 | 16056576 |
| Batch normalization 308 | None,256 | 1024 |
| Activation 302 | None,256 | 0 |
| Dropout 5 | None,256 | 0 |
| Dense 11 | None,2 | 514 |
| Tota_ parameters | 2,12,39,362 | |
| Trainable_parameters | 64,73,986 | |

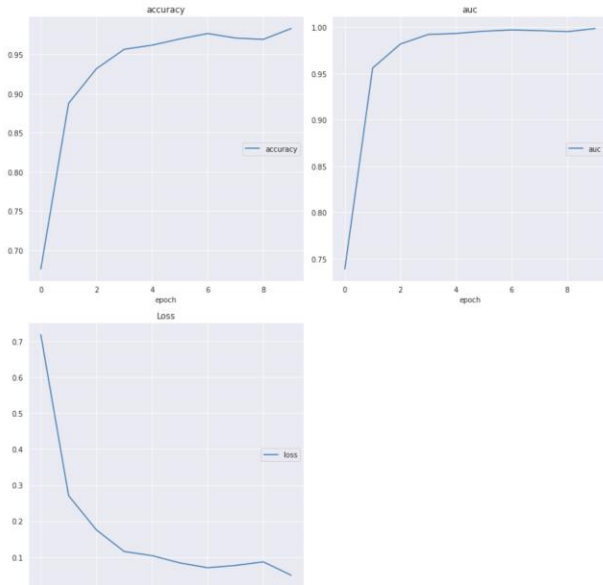| Non-trainable parameters | 1,47,65,376 | |
|---|---|---|



Fig 15. *Accuracy, AUC and Loss variation curve over the epochs for VGG-16*

Fig 15 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases

Table7. *Model Summary for ResNet101*

| ResNet 101 | | |
|---|---|---|
| Layer | Output Shape | Parameter Count |
| Resnet_101 | None,7,7,2048 | 42658176 |
| Flatten 6 | None , 100352 | 0 |
| Batch normalization 309 | None , 100352 | 401408 |
| Dense 12 | None,256 | 25690368 |
| Batch normalization 310 | None,256 | 1024 |
| Activation 303 | None,256 | 0 |
| Dropout 6 | None,256 | 0 |
| Dense 13 | None,2 | 514 |
| Total_parameters | 6,87,51,490 | |
| Trainable_parameters | 2,58,92,098 | |

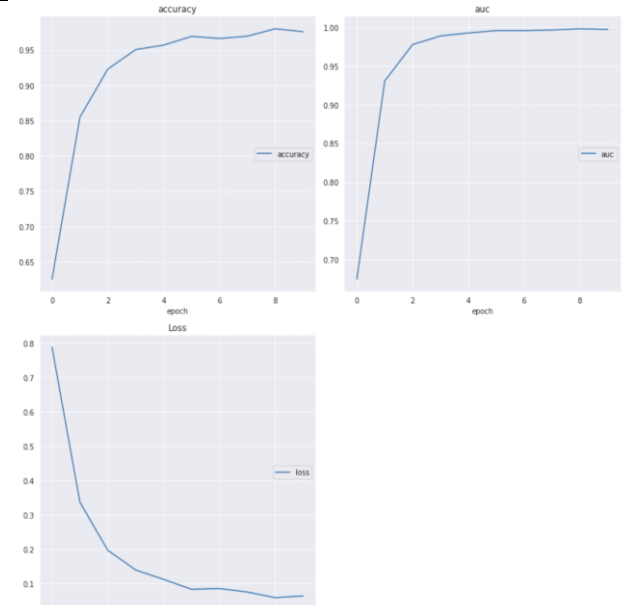| Non-trainable parameters | 4,28,59,392 | |
|---|---|---|



Fig 16. *Accuracy, AUC and Loss variation curve over the epochs for ResNet 101*

Fig 16 Alt text. The graph shows how accuracy and AUC increases with increasing number of epochs and while the loss decreases
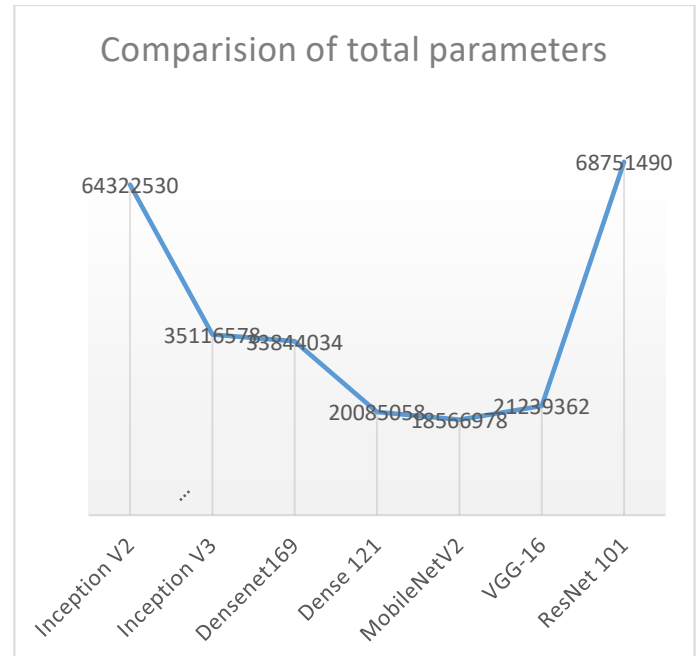


Fig 17. *Total parameters comparison for different pre trained models*

Fig 17 Alt text. The comparison graph shows the difference between total parameters for pre trained network

Table 8. *Performance Report*

| Performance report | | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy Score | Precision score | Recall score | F1 Score | ROC AUC Score | Cohen Kappa Score |
| InceptionResNetV2 | 0.69 | 0.6866 | 0.6888 | 0.6868 | 0.6778 | 0.3592 |
| Inception_V3 | 0.79 | 0.7874 | 0.7876 | 0.7875 | 0.7832 | 0.5667 |
| DenseNet169 | 0.79 | 0.792 | 0.7918 | 0.7919 | 0.7881 | 0.5759 |
| DesneNet121 | 0.81 | 0.8065 | 0.8069 | 0.8066 | 0.802 | 0.6054 |
| MobileNetV2 | 0.85 | 0.8474 | 0.8476 | 0.847 | 0.8414 | 0.6054 |
| VGG-16 | 0.95 | 0.9549 | 0.9549 | 0.9549 | 0.9538 | 0.6874 |
| ResNet101 | 0.95 | 0.9485 | 0.9485 | 0.9484 | 0.9463 | 0.9081 |



Fig 18. *Performance report chart for different Pre-Trained models*

Fig 18 Alt text. The graph shows the accuracy , precision , recall, F1, ROC AUC and Copen kappa Score for all the studies pre trained model

Table 9. *Classification Report*

| Classification Report | | | | | |
|---|---|---|---|---|---|
| | | precision | recall | f1-score | support |
| InceptionResNetV2 | B | 0.71 | 0.76 | 0.73 | 265 |
| | M | 0.65 | 0.6 | 0.62 | 201 |
| | | | | | |

| Inception_V3 | B | 0.81 | 0.82 | 0.81 | 265 |
|---|---|---|---|---|---|
|  | M | 0.76 | 0.75 | 0.75 | 201 |
|  |  |  |  |  |  |
| DenseNet169 | B | 0.82 | 0.82 | 0.82 | 265 |
|  | M | 0.76 | 0.76 | 0.76 | 201 |
|  |  |  |  |  |  |
| DesneNet121 | B | 0.83 | 0.84 | 0.83 | 265 |
|  | M | 0.78 | 0.77 | 0.77 | 201 |
|  |  |  |  |  |  |
| MobileNetV2 | B | 0.85 | 0.89 | 0.87 | 265 |
|  | M | 0.84 | 0.8 | 0.82 | 201 |
|  |  |  |  |  |  |
| VGG-16 | B | 0.96 | 0.96 | 0.96 | 265 |
|  | M | 0.95 | 0.95 | 0.95 | 201 |
|  |  |  |  |  |  |
| ResNet101 | B | 0.95 | 0.96 | 0.96 | 265 |
|  | M | 0.95 | 0.93 | 0.94 | 201 |

## 5. Conclusion

The use of pre trained model of CNN has been in use recently and they provide a very reliable and robust method of training our model to do prediction. Especially in the case of health care filed, the transfer learning approach becomes game changer as even with a smaller number of data we can use the knowledge gained by the models with one task and it can be applied to any other task of our choice in our case detection of breast cancer malignancy task. We have used various pre trained model for computer vision problems. In our study we found out VGG16 and Resnet101 perform better in predicting the cancer classification, VGG16 performs slightly better than Resnet101 in predicting the malignant Tissue.

## Future work

In the primary survey carried out in various hospital in Uzbekistan, it has been observed that ultrasound reports are easily available of the breast cancer patient, hence it will be interesting to implement image segmentation model in breast cancer radiology report which could be ultrasound report to detect and predict Breast Cancer. The ultrasound machines are widely available and with the availability of large number of datasets. the performance of the model can be increased significantly.

## Reference

[1] Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2015. CA Cancer J. Clin. 2015, 65, 5–29. [CrossRef] [PubMed]

[2] American Cancer Society. Breast Cancer Facts & Figures; American Cancer Society, Inc.: Atlanta, GA, USA, 2015.

[3] Eurostat. Health Statistics: Atlas on Mortality in the European Union; Office for Official Publications of the European Union: Luxembourg, 2009.

[4] Hubbard, R.A.; Kerlikowske, K.; Flowers, C.I.; Yankaskas, B.C.; Zhu, W.; Miglioretti, D.L. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: A cohort study. Ann. Intern. Med. 2011, 155, 481–492. [CrossRef] [PubMed]

[5] Hamidinekoo, A.; Denton, E.; Rampun, A.; Honnor, K.; Zwiggelaar, R. Deep learning in mammography and breast histology, an overview and future trends. Med. Image Anal. 2018, 47, 45–67. [CrossRef] [PubMed]

[6] Arevalo, J.; González, F.A.; Ramos-Pollán, R.; Oliveira, J.L.; Lopez, M.A.G. Representation learning for mammography mass lesion classification with convolutional neural networks. Comput. Methods Programs Biomed. 2016, 127, 248–257. [CrossRef] [PubMed]

[7] Carneiro, G.; Nascimento, J.; Bradley, A.P. Unregistered multiview mammogram analysis with pre-trained deep learning models. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 652–660.

[8] Huynh, B.Q.; Li, H.; Giger, M.L. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imaging 2016, 3, 034501. [CrossRef] [PubMed]

[9] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.

[10] Jiao, Z.; Gao, X.; Wang, Y.; Li, J. A deep feature based framework for breast masses classification. Neurocomputing 2016, 197, 221–231. [CrossRef]

[11] J. Imaging 2019, 5, 37 11 of 11

[12] Lévy, D.; Jain, A. Breast mass classification from mammograms using deep convolutional neural networks. arXiv 2016, arXiv:1612.00542.

[13] Ting, F.F.; Tan, Y.J.; Sim, K.S. Convolutional neural network improvement for breast cancer classification. Expert Syst. Appl. 2019, 120, 103–115. [CrossRef]

[14] Rampun, A.; Scotney, B.W.; Morrow, P.J.; Wang, H. Breast Mass Classification in Mammograms using Ensemble Convolutional Neural Networks. In Proceedings of the 20th International Conference on e-Health Networking, Applications and Services (Healthcom), Ostrava, Czech Republic, 17–20 September 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.

[15] Krizhevsky, A. One weird trick for parallelizing convolutional neural networks. arXiv 2014, arXiv:1404.5997.

[16] Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks

[17] by preventing co-adaptation of feature detectors. arXiv 2012, arXiv:1207.0580.

[18] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv

[19] 2014, arXiv:1409.1556.

[20] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.

[21] Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern

[22] Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

[23] Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate

[24] shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015;

[25] pp. 448–456.

[26] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE

[27] Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

[28] Heath, M.; Bowyer, K.; Kopans, D.; Kegelmeyer, P., Jr.; Moore, R.; Chang, K.; Munishkumaran, S. Current status of the digital database for screening mammography. In Digital Mammography; Springer:

[29] Berlin/Heidelberg, Germany, 1998; pp. 457–460.

[30] Tsochatzidis, L.; Zagoris, K.; Arikidis, N.; Karahaliou, A.; Costaridou, L.; Pratikakis, I. Computer-aided

[31] diagnosis of mammographic masses based on a supervised content-based image retrieval approach.

[32] Pattern Recognit. 2017, 71, 106–117. [CrossRef]

[33] Arikidis, N.; Vassiou, K.; Kazantzi, A.; Skiadopoulos, S.; Karahaliou, A.; Costaridoua, L. A two-stage method

[34] for microcalcification cluster segmentation in mammography by deformable models. Med. Phys. 2015,

[35] 42, 5848–5861.

[36] Lee, R.S.; Gimenez, F.; Hoogi, A.; Miyake, K.K.; Gorovoy, M.; Rubin, D. A curated mammography data set

[37] for use in computer-aided detection and diagnosis research. Sci. Data 2017, 4, 170–177.

[38] Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Chia

[39] Laguna Resort, Sardinia, Italy, 13–15 May 2010; pp. 249–256.

[40] Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database.

[41] In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009),

[42] Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 248–255.

[43] Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.

[44] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.;

[45] Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 2015,

[46] 115, 211–252. [CrossRef]

[47] Rouhi, R.; Jafari, M.; Kasaei, S.; Keshavarzian, P. Benign and malignant breast tumors classification based on

[48] region growing and CNN segmentation. Expert Syst. Appl. 2015, 42, 990–1002. [CrossRef]

[49] Xie, W.; Li, Y.; Ma, Y. Breast mass classification in digital mammography based on extreme learning machine.

[50] Neurocomputing 2016, 173, 930–941. [CrossRef]

[51] Tsochatzidis, L.; Costaridou, L.; Pratikakis, I. Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. J. Imaging 2019, 5, 37. https://doi.org/10.3390/jimaging5030037

[52] Mohan kumar, Sunil kumar khatri, Masoud Mohammadian;" Breast cancer identification and prognosis with machine learning techniques - An elucidative review

[53] " Journal of Interdisciplinary Mathematics 2020,pp503-521

[54] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna," Rethinking the Inception Architecture for Computer Vision" Computer Vision and Pattern Recognition, arXiv,2015

[55] G. Huang, Z. Liu, van, and Weinberger, Kilian Q, "Densely Connected Convolutional Networks," arXiv.org, 2016. arxiv.org/abs/1608.06993.