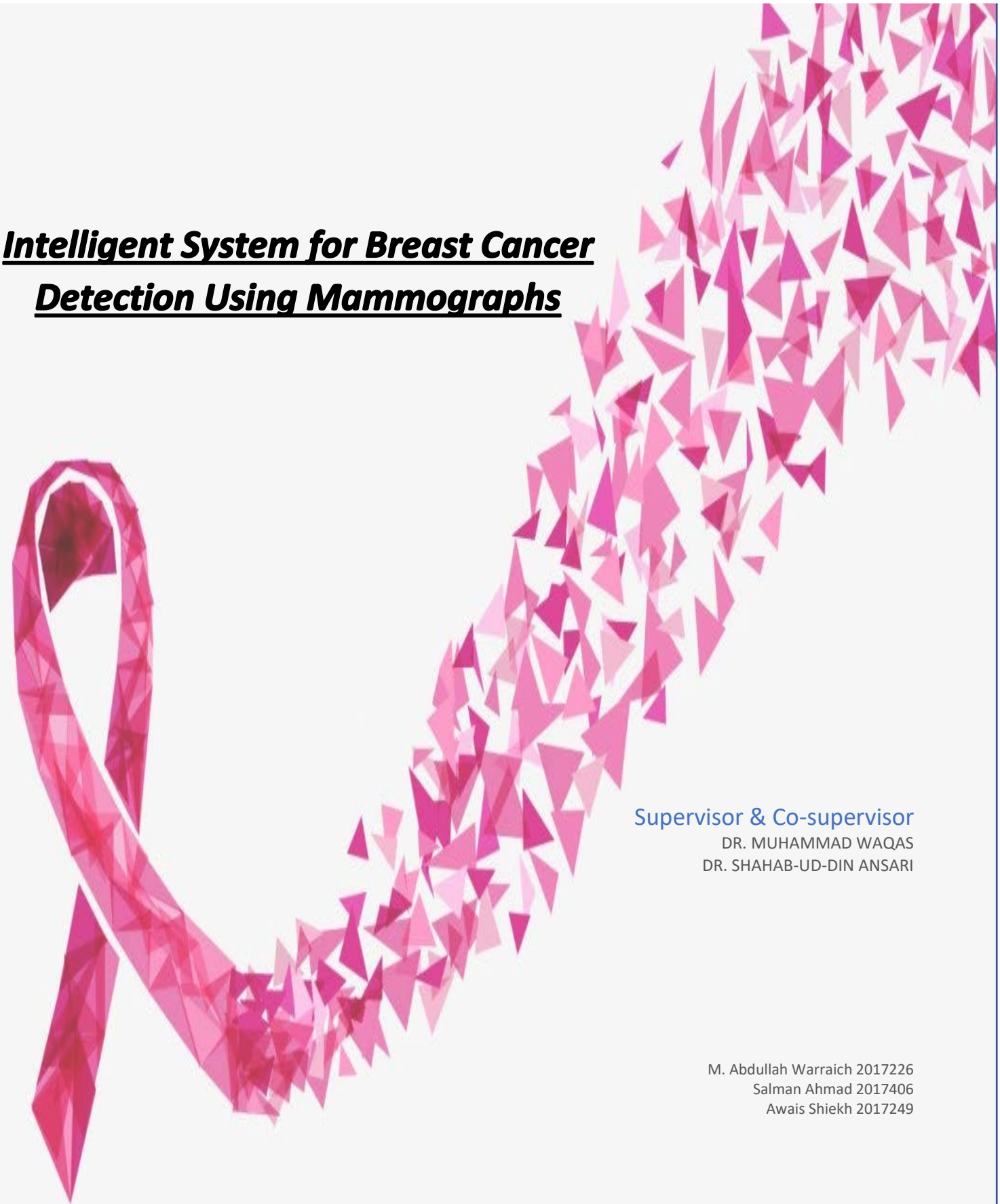
Intelligent System for Breast Cancer Detection Using Mammographs



Supervisor & Co-supervisor

DR. MUHAMMAD WAQAS

DR. SHAHAB-UD-DIN ANSARI

M. Abdullah Warraich 2017226

Salman Ahmad 2017406

Awais Shiekh 2017249

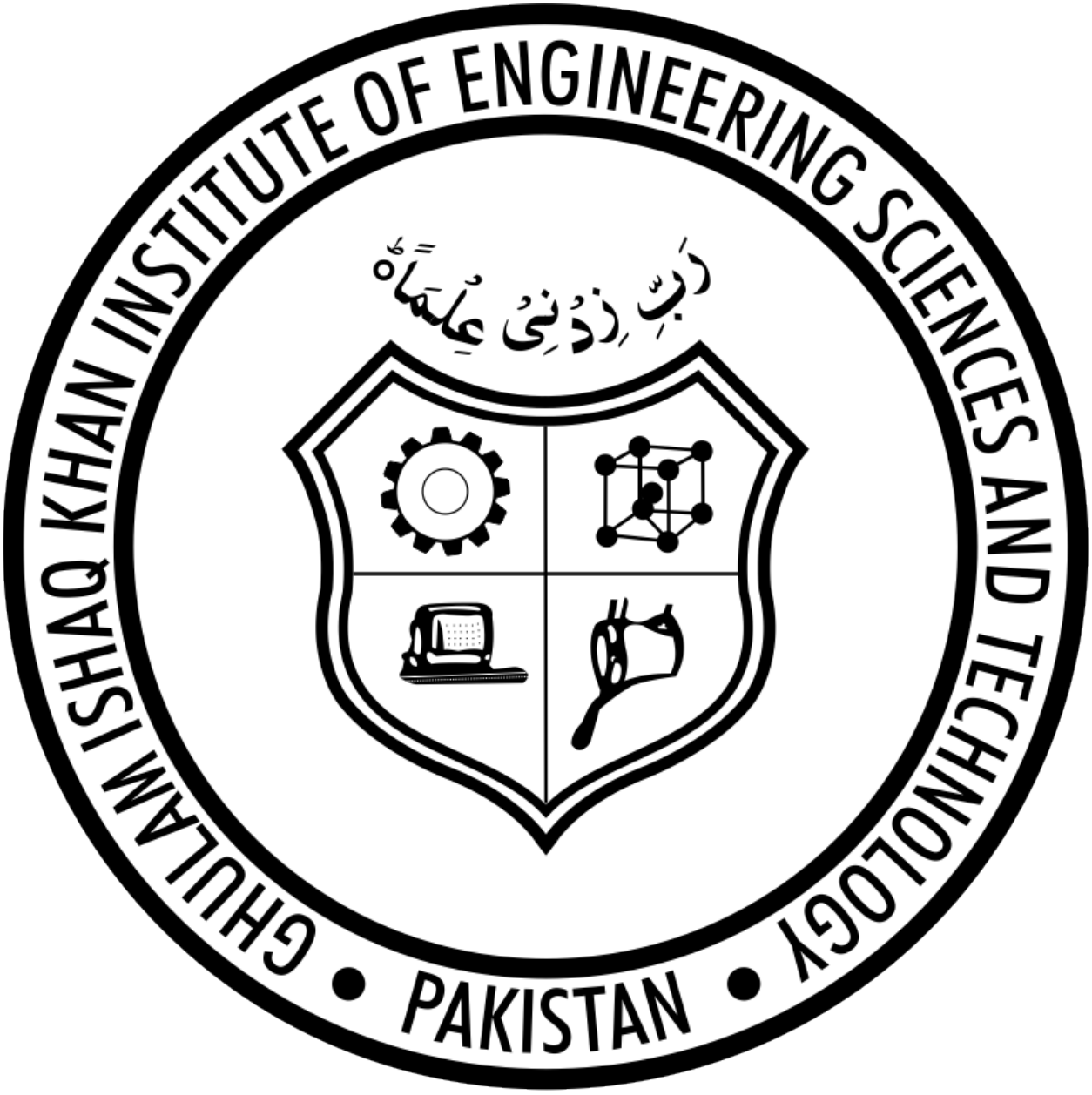

GHULAM ISHAQ KHAN INSTITUTE OF ENGINEERING SCIENCES AND TECHNOLOGY
PAKISTAN
رَبِّ زِدْنِی عِلْمًا

Table of Contents

Chapter no.1 Introduction 3

1.1 Introduction 3

1.2 Objectives: 3

1.3 Project Motivation: 3

Chapter No. 2 Literature Survey 4

2.1 Prior Work 4

2.2 Current Work 5

Chapter No. 3 Functional Requirements 6

3.1 Android Application Design 6

Chapter # 04 Proposed Solution (Methodology, Implementation) 8

Methods 8

4.1 Conversion of a patch classifier from recognizing patches to whole image classifier 8

4.2 Network design 10

4.3 Computational environment 12

Chapter 5: Results and discussion 13

Results 13

5.1 Classifier training on DDSM 13

5.1.1 Setup and processing of the dataset 13

5.1.2 Developing image classifiers 14

5.1.3 Model averaging 16

5.2 Transfer learning on INbreast 177

5.2.1 Dataset preprocessing and setting up 177

5.2.2 Efficiency of transfer learning 17

Discussion: 19

Chapter 6: Conclusion and Future Work 22

References: 23

Chapter no.1 Introduction

1.1 Introduction

Breast cancer is ranked to be among the leading cause of deaths for women every year. The major causes of breast cancer are obesity, radiation exposure, more alcohol consumption, and chances of having it in women increase with the age. Some of the previously used methods to detect the breast tumor detection are Breast Ultrasound (BUS), Magnetic resonance imaging (MRI) and Microwave Imaging Technique (MIT) etc. Unable to detect tumors in early stages from the above-mentioned methods. Due to similarity in cancerous tissue and normal tissue early detection becomes too difficult.

Our project aims to detect cancerous breast tumors from mammographs using a Convolutional Neural Network (CNN) in early stage. Firstly, capture image from the camera. Then doing image enhancement followed by image segmentation. Then using the CNN and extracting features from the image and finally making decision whether tumor detected or not.

1.2 Objectives:

Through our project devise a system that helps doctors and surgeons increase their vision and knowledge about early detection of breast cancer. Our project helps in bringing awareness among the women in the sense that substantial work as well as effort is being done in this area of research to help as well to treat and save their lives. Also, showing the power of using machine learning tools and methods in breast tumor detection.

1.3 Project Motivation:

Since, ample amount of work is being done in detecting the breast tumor detection in the past like breast tumor detection from Breast Ultrasound (BUS) method, Magnetic resonance imaging (MRI) and Microwave Imaging Technique (MIT) etc. So, there remains an urge to do some more work and innovations in this field. So, we decided to devise a technique and detect breast tumor from mammographs using a CNN.

Chapter No. 2 Literature Survey

Breast Tumor Detection Using Ultrasound: Previously, breast tumor detection was done using ultrasound(Wu et al., 2020). Usually, tumor detection through ultrasound is invaluable for imaging the internal organs due to the noise present in the images. So, for this very reason tumor detection could not be done in the early stages. The technique proposed in our report is to detect the breast tumor in the early stages using the CNN.

Breast Tumor Detection using Microwave Imaging: In this technique, certain number of antenna's are used to for the purpose of transmitting and receiving. Also, using these antenna's a full 3-D visual is obtained from every angle of the breast is obtained. This technique helps in telling the difference between the healthy tissue and abnormal tissue. The microwave imaging(Malla et al., 2020) methodology is close to the technique proposed in our report. But the methodology used in the report is breast tumor detection through mammograms using a CNN.

Breast Tumor Detection in Automated Breast Ultrasound: The mentioned technique is being widely used for detection of breast related abnormalities. To fasten this technique candidate aggregation algorithm is used. As automated Breast Ultrasound(Chiang et al., 2019) is a slow approach. In this technique, volume of interest is selected first and then each VOI is assigned a probability score and finally aggregation technique is applied. This paper infers to get good results from the applied technique. The proposed methodology is a bit old, and technique proposed by our report much good in terms of early detection of breast tumor.

Breast cancer detection using CPW-Antenna: In this technique, CPW antenna(Slimi et al., 2019) from certain distance throw electric signal around the breast to detect the cancerous and usually represents the cancerous cells by the sphere. The technique proposed in the conference paper claims good results using this technique. The technique proposed by our report is far simpler and easier. As, it would help doctors and surgeons easily get used to with our proposed methodology.

2.1 Prior Work

With the advance in technology humans had been more dependent on the mobile phones, performing different tasks such as bill payments, money transfers and many more. Today mobile phones can be used for detection of tumors.

One such example can be D3 system created by Massachusetts General Hospital by Ralph Weissleder. For improving the health of the people in Africa Dr. Weissleder travelled to Africa and deduced that technology could be used to improve health care conditions for people in South Africa since the number of practicing pathologists is less overall and lab do not return test results on time. "People walk for 20 miles to get an answer to why they are sick," Dr. Weissleder says. "They can't come back the next day.". On observation he found out that mobile is used by people in remote areas of Africa.

The D3 system was a quick and responsive system where you just had to slide the blood sample and the camera of the smartphone captures the image which detects whether you have cervical cancer or not.

Upon further research we found out that there is also a device called smartphone dongle for cancer biomarker measurement. The core concept behind the system was, patients had to suffer needless anxiety for many days when they sent their reports to detect cancers. This device takes a blood sample of the patient and mix this sample with the reactive fluid on the strip which then inserted into the reader, the reader can show results in the smartphone.

Many such evidence of using smartphone in diseases detection were found and then a decision was made to work on the breast tumor detection using the mammograms images from the smartphone.

2.2 Current Work

What we planned in our work was using the machine learning advance algorithms to detect the breast tumors through mammograms as this can be nothing but an image detection task. The CNN was implemented to in the process of detecting harmful and harmless tumors, trained on a CBIS-DDSM dataset. The base paper for our learning was Deep Learning to Improve Breast Cancer Detection on Screening Mammography.

Once we got our model, the work on android application was started using the android studio and language used was java.

Chapter No. 3 Functional Requirements

Software development life Cycle (SDLC) process is followed to build a system or application with improved performance, quality, tracking, controlling and maintenance. It has different models.

- Waterfall Model
- Incremental
- Hybrid
- RAD Model

RAD model is being selected for this model. Because it is based on prototyping, and there is not any specific planed path involved, in it we somehow use CASE toll (Computer Assisted Software Engineering). It is used due to its quick method.

- It uses less development time and get edge to increase the time of re-usability of available component.
- Due to a smaller number of projects member, it increases their productivity.
- Due to less time and high-speed change in requirement it accommodates it easily.
- Due to less time for development, it helps in quick development of system.

3.1 Android Application Design

The model was firstly converted to tflite file and then uploaded to firebase custom MLkit. The internet permissions were then granted to the application in the android manifest file to access the firebase server. The use of firebase was a clever approach as upon updating the model the application automatically downloads the updated model.

Next the model should be bundled with application for that tflite file was added to application's assets folder. The assets folder in the res package also includes a doc file for the labels.

A FirebaseCustomRemoteModel object is created to load the model then downloading the model was done by giving the conditions such as whether internet connection is available or not.

After the model is loaded and downloaded, configuring, and interpreting of the model is done. Our model was remotely located on the firebase server so FirebaseModelInterpreter object is used, and the inputs and outputs specification are given to the model through interpreter.get_input_details method. FirebaseInputOutputOptions method was used to give all the necessary specifications of the inputs and outputs our model accepts and give. As our model accepts the mammographs images so an image from camera is saved as bitmap and passed to the model.

MLKit

- Offline Models
- Remote serving
- Remote Config

Android

- User experience
- Performance

The outputs were then mapped to the labeled file and results were then displayed in a new activity. For the android application, the firebase authentication was also used to login or signup.

Chapter # 04 Proposed Solution (Methodology, Implementation)

Methods

4.1 Conversion of a patch classifier from recognizing patches to whole image classifier

In order to divide and place a set of complex pictures into specific groups, a typical methodology includes the utilization of a classifier in slider style to perceive neighboring patches on a picture to produce a framework of probabilistic yields. This is trailed by yet another cycle of summing up the patch classifier's yields to provide the last categorization or division result. Such techniques have been utilized to distinguish spreading breast cancer, also known as last stage or advanced-stage cancer, utilizing entire slide pictures of sentinel lymph node biopsies and section neural layers in infinitesimal pictures.

Regardless, this method requires two phases that each ought to be smoothed out autonomously. Here, we propose a system to join the two phases into a singular development for setting up all in the gave figure (Fig. 1). Consider it along these lines, we have an underlying patch $X \in IR_{p \times q}$ $X \in IR_{p \times q}$ to get contribution from and a patch classifier which establishes a capacity f with the objective such that f(X) ∈ IRc f(X) ∈ IRc, where the particular capacity's yield placates f(X)i ∈ [0, 1] and Σci=1 f(X)i=1 Σi=1 cf(X) i=1 and here the letter c addresses the multitude of classes present in the patches. In the previously mentioned strategy, c is equal to 5 and the defined classes are: benign calcification, malignant calcification, benign mass, malignant mass, and background. These classes are for every single patch that has been fragmented from a mammogram.

Consider the input patch is extricated via a picture $M \in IRr \times s$ $M \in IR\ r \times s$ where p is really less than r and q really less than s. Assuming the coded function f is used to address a CNN, f can be used with M with no changes to the defined bounds so $f(M) \in IR_{u \times v \times c}$ $f(M) \in IR_{u \times v \times c}$, where u is less than 1 and v is less than 1, rely upon the picture size and the step used for the patch classifier. This has been made feasible due to the weight distribution and the region property of a CNN. On the off chance that the function f addresses an alternate class of neural network, for example, the multilayer perceptron (MLP), at that point this turn out to be impracticable since an MLP needs the input to be consistent. Subsequently, if the input is to be changed from X to M, we have a u × v network of selective yields of c number of classes (called a "heatmap") rather than a solitary yield of the c classes. Thus, giving us a heatmap of the size of u × v × c. This would hence give us the chance to add more layers to the existing heatmap to change the yields and associate with the last classification yield of the picture. to transform the whole patch

classifier into a screening element and broaden the respective responsive field we added a convolutional layer onto the patch classifier's yields.

For example, let us consider a patch classifier of size 300x300. If it has a stride of 32 then, using a 4x4 convolutional layer on top of it can be used to increase the receptive field dimensions, converting it into 300 + (4-1) x 32 = 396. Accordingly, the top layers successfully utilize the patch classifier to check the entire picture by 'scanning' it, searching for signs of cancerous lesions and extricating a more significant level element that can, at last, be utilized for entire picture characterization.
Utilizing function Y to address the layers above, the entire picture categorization procedure can be composed as h(M)=Y(f(M))∈IRih(M)=Y(f(M))∈IRi, where i is representing the number of classes of the entire picture. Ordinarily, i = 2 addresses the two classes we need to foresee: malignant and nonmalignant (benign or ordinary).

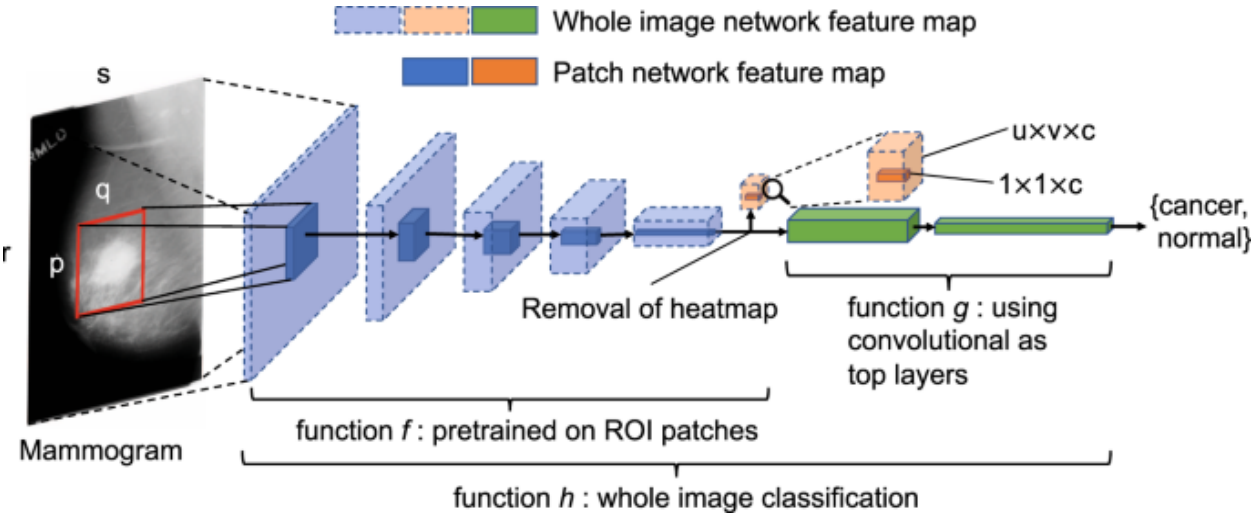


Figure 1: Changing a patch classifier over to an end-to-end complete picture classifier utilizing a fully convolutional disposition. The procedure f was refined resting on entire pictures after being trained and prepared originally on patches. the data stream was checked for improvements by removing the heatmap from the base layers of the patch classifier to the top convolutional layers in the entire picture classifier. The amplified part represented by the magnifier basically displays an amplified rendition of the heatmap. the figure was shaded in different colors to help understand better the different parts of the full picture classifier.

The function h supports entire pictures as input and returns defined characteristics for the entire picture level. Subsequently, this end to end teachable function gives two benefits for the process. In the first place, it tarries away from imperfect answers from each progression, by training all the networks together; Secondly, prepared system could be used on an alternative dataset devoid

of any dependency on ROI explanations. since the mammograms are hard to get and can be expensive if bought in bulk quantities. further getting labelled data set is difficult as well as expensive. The biggest publicly available source of breast cancer images with ROI explanations for digitized radiographic mammogram films is the DDSM, Digital Database for Screening Mammography. this huge database has thousands of pictures available with pixel-level annotations, which can be utilized accordingly to prepare the patch classifier f. When f is shaped and changed over into an entire picture classifier h, it very well may be calibrated on other available databases utilizing just picture-level labels. This methodology can be utilized to lessen the necessity for ROI explanations yet providing numerous usages in clinical imagery as well as breast cancer recognition on diagnosing mammograms.

4.2Network design

The cutting edge CNN are commonly built with putting convolutional layers over the input, trailed by at least one fully connected (FC) layer to get together with grouping yield. Max-pooling layers are frequently utilized between convolutional layers to enhance translational invariance and to shrink feature map size. In this report, two well-known CNN structures are analyzed:

- VGG network, and,
- ResNet.

Sequential network layers can be normally assorted into "blocks". This helps diminish the feature map size, regularly by a factor of 2. This size reduction can either be done initially or in the final stages of a block yet remains similar somewhere else in the block. A 2×2 max-pooling layer that lessens the feature map size by a factor of 2.

There is no justification for utilizing a 2x2 max-pooling layer alongside a 3x3 convolutional layer, except for the way that it is the most utilized sizes. These same sizes have been utilized in the entire report. To address a block of VGG we have a 'k' number of convolutional layers having a depth of 'n' for each layer. subsequently, we can express that the VGG block is of size k x n. To diminish the size of the feature map the ResNet block uses a step size of 2 in its first most convolutional layer, instead of having a max-pooling layer to do it. It at that point comprises a pile of various convolutional layers after the initial layer. A bottleneck configuration is carried out having a repetitive arrangement of three convolutional layers. These three convolutional layers are given by the spans of 1x1, 3x3 and at long last a 1x1 layer, all together. Quite possibly the most appealing qualities of this are, that it gives a detour on the layers in the middle, and straightforwardly associates the initial and the last layer in every unit. These aides in producing a quicker and more engaged learning technique by making the units train on the residual information. To accelerate the union and to have a regularization impact, the ResNet utilizes batch normalization (will be expressed as BN from now into the foreseeable future) in every one of its convolutional layers.

We can address the ResNet block by [L − M − N] × K, where L, M and N address the depths associated with the three convolutional layers present in a single unit and K addresses the quantity of present units. Here, the 16-layer VGG network (VGG16) and the 50-layer Resnet (Resnet50) are utilized as patch classifiers. The initial plan of the VGG16 comprised five VGG blocks trailing two FC layers. To be harmonious with the Resnet50, we supplanted the two FC layers with a global average pooling layer which figures the average activation of each feature map for the yield of the last VGG block. For instance, if the yield of the last VGG block dimensionally, 5 × 5 × 512 (height × width × channel), after the worldwide normal pooling layer the yield gets 512. This yield is afterwards linked to the grouping yield via a FC layer.

An immediate method to manage fostering a full picture classifier from a patch classifier incorporates smoothing the heatmap in addition to interfacing it to the image's classification yield using FC layers. To assemble the model's translational invariance to the patch classifier's yield, a max-pooling layer can be utilized after the heatmap. Further, a simple course can be made between the heatmap and the yield to simplify the process of training. The heatmap results clearly from the patch classifier's yield which uses the SoftMax:

$$f(z)j=ezjΣci=1ezi \text{ for } j=1,\ldots,c$$

$$f(z)j=ezjΣi=1cezi \text{ for } j=1,\ldots,c \quad (1)$$

Nonetheless, the SoftMax has the effect of reducing slopes in case of huge data inputs. This data can often disrupt the flow of the gradient when it is utilized in an in-between layer. Accordingly, the rectified linear units (ReLU) may also be utilized preferably:

$$f(z)j=\max(0, zj) \text{ for } j=1,\ldots,c$$

$$f(z)j=\max(0, zj) \text{ for } j=1,\ldots,c \quad (2)$$

It was additionally suggested that convolutional layers should be utilized as the topmost layers. this was done by keeping the security concerns in mind since this approach safeguards spatial data. Adding 2 blocks of FC layers, may it be VGG or ResNet, on top of the patch classifier layers, trailed with a single global avg pooling layer and then eventually in the end, the picture's categorization yield (Fig. 1). Hence, this method enables the creation of a fully convoluted network design for the entire picture classification process. As presented by Fig. 1, the depth of the feature map amongst the patch classifier layers and the surface layers is swiftly diminished using the heatmap. although the depth feature is diminished it has its effects, it can cause data loss in the entire picture classifier.

4.3 Computational environment

- Windows 10
- 16GB RAM
- 1TB
- NVIDIA GeForce RTX 2080 GPU card.

Chapter 5: Results and discussion.

Results

5.1 Classifier training on DDSM

5.1.1 Setting up and preprocessing of the dataset

The previously available dataset for digital mammography images was the DDSM dataset, which was available in the lossless jpeg format. This format has been outdated for quite a while now, so a new type of dataset was rolled out by the same providers, called CBIS-DDSM, Curated Breast Imaging Subset of DDSM. The team used this dataset which is formatted in the DICOM format. the CBIS-DDSM consisted of a total of 2478 images of a breast cancer diagnosis from the 1249 women. The dataset is available to download from the official website of CBIS-DDSM. For most of the obtained images it had 2 different views:

1. Craniocaudal (CC), and
2. Mediolateral oblique (MLO).

These different views were used as a different picture each for the sake of this research. Furthermore, the dataset was randomly split in a ratio of 85:15 to make training sets and test sets, respectively. From within the training set, another 10% was randomly selected to be used as the validation set. The division of these 3 sets was done in a strategic manner keeping the approximate number of cancerous images equal in all of them based on percentages. Following are the numbers of images that were split into their respective sets:

1. training = 1903,
2. testing = 199, and,
3. validation = 376.

The numbers of images of the four combinations of label and type are about the same. We create a patch image set by sampling 10 patches from each ROI having a minimum overlapping ratio of 0.9 and the same number of background patches from the same mammogram. According to the annotations of the ROIs, a patch is classified into five categories: background, calcification-benign, calcification-malignant, mass-benign, and mass-malignant.

The database consists of detailed explanation for the ROIs. They also have the medically approved terms to name these ROIs called.

- benign, or,
- malignant.

Not only these pictures have the terms associated with them but also their types, like:

- calcification, or,
- mass.

These 4 different types of images and labels are having the same quantity. A patch picture set was created by us by examining a total of 10 patches from every ROI. These selected patches had a few checks placed on them to be sampled fairly:

- the selected patch image had at least an overlapping ratio of 0.9
- the selected patches had a similar number of background patches from a similar mammogram.

The comments on the pictures with ROI split them into a total of 5 groups, namely:

- background,
- calcification-benign,
- calcification-malignant,
- mass-benign, and,
- mass-malignant.

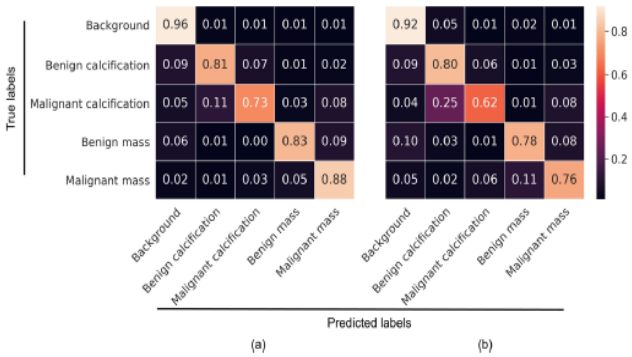


Figure 2. Confusion matrix showing 5-class patch classification for (a) Resnet50 and (b)VGG16 in the S10 test set. The matrices have been normalized to get individual row sum equal to 1.

5.1.2 Developing image classifiers

Two mainstream convolutional networks were utilized in the process of producing a network that can identify and group patches into the 5 classes. They are the ResNet50 and the VGG16 structures. The 5 classes are balanced based on the number of background patches that are checked and coordinated with the ROI patches on a similar mammogram. Accordingly, after running both the models for their test accuracies we concluded that the ResNet50 model can achieve higher accuracy than the VGG16 by 0.05 (Table 1).

Model	Accuracy	#Epochs
Resnet50	0.89	39
VGG16	0.84	25

Table 1: The above table represents best accuracies got using the Resnet50 and VGG16 patch classifiers. #Epochs shows us the epoch number on which the highest AUC was attained in the validation set.

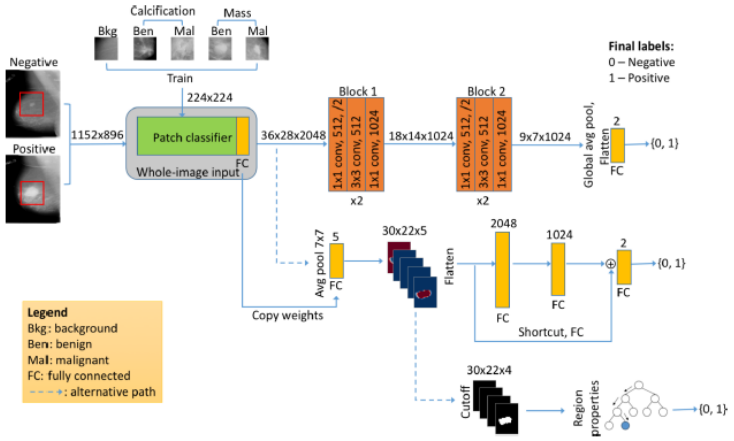
The patch classifier was transformed into the entire picture classifier through testing different arrangements for the top layers. The test sets were run through the per-image receiver operating characteristic curve (AUC) scores to evaluate various models. According to the table shown the initial test conversion was tested out on the ResNet50 model using a total of 2 similar residual blocks and in-between we used the bottleneck design of 512-512-2048 without the repetition of any residual unit. The obtained AUC score was around 0.85, but it was further increased to 0.86 by decreasing the size of the final block from 2048 to 1024. By doing so the initial restriction imposed by the size of the last layer was removed to allow the rehashing of the residual blocks. Having done this, we also tried reducing the size of the first 2 blocks, but they seemed to have decreased the accuracy by 0.2 and the new changes in dimensions were rejected based on results.

Patch set	Block 1	Block 2	AUC [95% CI]	A-AUC [95% CI]	#Epochs
S1	[512-512-2048] × 1	[512-512-2048] × 1	0.63 [0.58, 0.67]	NA	35
S1g	[512-512-2048] × 1	[512-512-2048] × 1	0.83 [0.79, 0.86]	NA	38
S10	[512-512-2048] × 1	[512-512-2048] × 1	0.85 [0.82, 0.88]	0.86 [0.83, 0.89]	20
S10	**[512-512-1024] × 2**	**[512-512-1024] × 2**	**0.86 [0.83, 0.89]**	**0.87 [0.83, 0.90]**	**34**
S10	[256-256-256] × 1	[128-128-128] × 1	0.84 [0.81, 0.87]	0.86 [0.82, 0.89]	25
S10	**256 × 1**	**128 × 1**	**0.87 [0.84, 0.90]**	**0.88 [0.84, 0.90]**	**36**
Insert heatmap between patch classifier and top layers					
S10	[512-512-1024] × 2	[512-512-1024] × 2	0.80 [0.76, 0.84]	NA	47
S10	[64-64-256] × 2	[128-128-512] × 2	0.81 [0.77, 0.85]	NA	41
Add heatmap and fully connected (FC) layers on top (S10 patch set)					
Pool size	**FC1**	**FC2**			
5 × 5	64	32	0.74 [0.69, 0.78]	NA	28
2 × 2	512	256	0.72 [0.67, 0.76]	NA	47
1 × 1	2048	1024	0.65 [0.60, 0.69]	NA	43

Table 2. Representing the AUC scores per-image for the whole image classifiers using the Resnet50 as patch classifiers in the mutually exclusive test set. #Epochs shows us the epoch number on which the highest AUC was attained in the validation set. The AUC scores shown in bold are the models having the optimal performance record.

Moving on we put the change to the test utilizing the VGG16 patch classifier (Table 2). Some of the most recent VGG blocks utilize 3x3 convolutions with BN. We discover that the VGG structure is more inclined to overfitting than the remaining construction Nonetheless, this can be moderated by diminishing the model's intricacy simply by add two 512x1 blocks increases the score to 0.83.

By reducing the dimensions of the VGG blocks further and find that the scores decrease by just 0.03. This corresponds to the results of the Resnet50-based models. Overall, Resnet50-based whole image classifiers outperform VGG16-based ones (Table 2). Furthermore, the Resnet50-based models appear to reach the highest validation score faster than the VGG16-based models to decide if the yield error is because of the patch network at the base or the recently added top layers, we add two residual blocks on top of the VGG16 patch classifier to make a new and improved model. It creates a score of 0.81, which is equivalent to VGG16-based organizations yet a couple of focuses lower than the best Resnet50-based models. After broad arrangement, nonetheless, this half breed model accomplishes a score of 0.85.



The deep learning structure for converting a patch classifier into a whole image classifier by adding convolutional layers on top. Shown is an example using two residual blocks with the same structure of [512-512-1024] x 2. Two alternative strategies are also presented: one is to add heatmaps and FC layers on top and the other is to add a random forest classifier on top of the heatmap.

5.1.3 Model averaging

The team selected various models to produce a total f 4 predictions for each model and calculate the average. This was done using inference-level augmentation by first performing a horizontal flip followed by a vertical flip. From the procedure we were able to get the top 3 models based on performance:

- Resnet50 with two [512-512-1024]x2 residual blocks,
- VGG16 with two 512x1 VGG blocks, and,
- VGG16 with two [512-512-1024]x2 residual blocks on top.

The augmented predictions for the three models improved the AUC scores from

- 0.86 to **0.88**,
- 0.83 to 0.86 and
- 0.85 to **0.88**, respectively (Table 2).

By averaging the AUC scores of the above 3 models we get an average AUC score of 0.91.

5.2 Transfer learning on INbreast

5.2.1 Dataset preprocessing and setting up

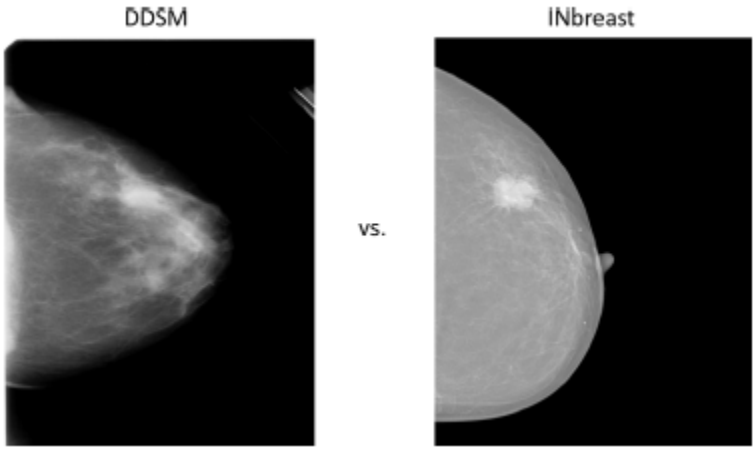
The INbreast dataset is another publicly available mammogram database that includes full-field digital images. These images differ from DDSM images in terms of colors. As a result, this proves to be an excellent source for portability testing of an entire image classifier from one database to another. There are a total of 115 patients with 410 mammographic images available in the INbreast database. The pictures have BI-RADS readings with each of them. The manual classification of BI-RADS containing images is as follows:

- values of 1 & 2 (negative samples),
- values 4,5, & 6 (positive samples), and
- value of 3 (irrelevant).

Using these class values, another 12 patients/23 mammograms can be excluded from the image dataset. We stratify the dataset and conduct a 70-30 split into train and validation sets depending on the patients. The sets contain:

- training = 280, and
- validation = 107

Now we apply steps like those applied on DDSM on the INbreast dataset too.



5.2.2 Efficiency of transfer learning

We fine-tune the entire image network on the train set without using ROI annotations, and we test model output using per-image validation AUC ratings. For transfer learning, the three best models from Section 3.1.3 are used. The refined Resnet50-based model earns a score of 0.84. Surprisingly, the fine-tuned VGG16-based model outperforms the Resnet50 with a score of 0.92. According to, the residual networks reduce the feature map sizes too aggressively, causing the ROI features in the first few layers to suffer. Based on this logic, the underperformance of

Resnet50-based models is most likely due to the bottom-most layers. This is supported by the hybrid model's high efficiency, which achieves a score of 0.95. As a result, we hypothesize that the VGG networks need more training time than the residual networks to achieve their maximum potential on the DDSM.

To demonstrate this, we run another model training run for the hybrid model with 200 more epochs. The model raises the test score from 0.81 to 0.85 (Table2), which is comparable to the best Resnet50-based models. We also want to know how much data is required to fine-tune an entire image classifier to achieve acceptable efficiency. This has significant practical consequences because acquiring marks, even at the entire picture stage, can be costly. We finetune the model by sampling a subset of 20, ..., 60 patients from the train set and evaluating its output on the same validation set (Table 3). The VGG16-based model and the hybrid model can obtain scores of 0.87 and0.89, respectively, with as few as 20 patients or 79 images.

As the train set size is increased, the scores seem to rapidly saturate. This fast change can be a huge benefit for end-to-end qualified whole image networks, reducing the strain of train set construction significantly. Finally, with augmented estimation, the VGG16-based model improves from 0.92 to0.94, while the hybrid model improves from 0.95 to 0.96. The combined score of the two augmented versions is 0.96. Incorporating the Resnet50-based model into model averaging does not affect the score.

#Pat	#Img	Resnet50	VGG16	Hybrid
20	79	0.78	0.87	0.89
30	117	0.78	0.90	0.90
40	159	0.82	0.90	0.93
50	199	0.80	0.93	0.93
60	239	0.84	0.95	0.91

Table 3: Efficiencies with various train set sizes, for transfer learning. Displayed as per validation scores from AUC scores per-image.

Discussion:

This investigation shows that exact arrangement of screening mammograms can be accomplished with a profound learning model prepared in a start to finish style that depends on clinical ROI explanations just in the pioneer stage. When the whole image classifier is raised, it very well may be calibrated utilizing more datasets that need ROI annotations, verily if the pixel power dispersions vary as is often the situation for datasets collected from blended mammography stages. These discoveries demonstrate that deep learning calculations can upgrade upon exemplary business CAD frameworks, like as iCAD SecondLook1.4 and R2 ImageChecker Cenova1.0, that are not deep learning based and were able to attain an average AUC of 0.726. Our all-convolutional networks prepared utilizing a start to finish approach have extensively aggressive execution and are more generalizable across various mammography stages contrasted and previous deep learning strategies that have achieved AUCs in the range of 0.65 –0.97 on the DDSM and INbreast databases, as well as in- house datasets12. Studies have reported that a new commercial CAD system, Transpara1.4.0, attained an AUC of 0.89 when used to support radiologists16 and 0.84 in standalone mode15. This business CAD utilized CNNs prepared utilizing the injury explanations from 9000 mammograms with malignancy to bring scores at the fix balance; the scores for all recognized areas were either joined into a score at the assessment balance. As far as anyone is concerned, the corporate CAD cannot unreservedly be adjusted on various mammography datasets without lesion annotations. Our methodology enjoys the benefit of testing just picture balance markers for tweaking once the entire picture classifier is developed to ease scaling to bigger datasets and moving to new mammography frameworks as they rapidly advance. Two ongoing investigations, grown profound learning grounded structures for breast cancer classification utilizing film and advanced mammograms, which were start to end teachable. The two investigations utilized multi-instance learning (MIL) and adjusted the entire picture classifier cost capacities to fulfill the MIL standard. In differentiation to our methodology, not one or the other. study applied ROI annotations to prepare the fix classifiers first and the AUCs were lower than detailed in this investigation.

#Patients	#Images	Resnet-Resnet	Resnet-VGG	VGG-VGG	VGG-Resnet
20	79	0.92	0.88	0.87	0.89
30	117	0.93	0.94	0.93	0.90
40	159	0.93	0.95	0.93	0.93
50	199	0.94	0.95	0.94	0.93
60	239	0.95	0.95	0.95	0.94
72 (All)	280	0.95	0.95	0.95	0.95

Table 4. Transfer learning efficiency with different training set sizes assessed by the per-image AUC on the INbreast test set.

We establish that the nature of the patch classifiers is basic to the exactness of the whole image classifiers. This was upheld by two lines of affirmation. To begin with, the whole image classifier grounded on the S10 fix set performed far superior to the one laid on the S1 fix set in light of the fact that the S10 fix set contained other data about the foundation than the S1 fix set. Second, it took any longer for the VGG16-rested whole image classifiers to accomplish a similar exhibition as the Resnet50-rested classifiers on the grounds that the VGG16 was less precise than Resnet50 in fix characterization. We additionally comprise that the precision of entire picture arrangement was refined by attempting more or bigger patches to incorporate adjoining districts around the ROI and other foundation locales. In any case, the computational trouble increments straightly with the number or size of patches attempted and the exhibition gain may quick decrease. Utilizing bigger patches can diminish the sign to-clamor rate, as demonstrated by the lower fix. arrangement exactness utilizing the S1gvs. S10 fix sets. Utilizing bigger fixes additionally requires progressed GPU memory, which may restrict network decisions. The saliency map investigation showed that our entire picture networks were reasonable to properly recognize the ROIs and utilize the data in that to understand malignant growth. It likewise showed that errors regularly went down in interesting cases, suchlike as benign lesions with malignant features, or malignant lesions that were interesting to recognize from foundation. Further investigation is requested to test how to test native fixes all the more proficiently, conceivably by intensifying the preparation information with precarious cases and focusing on the patches. that are bound to be misclassified.

This could help conquer the computational weight of preparing more precise classifiers. Albeit the VGG-based picture classifiers were more inclined to overfitting and requested longer preparing, the execution of VGG-based and Resnet-based picture classifiers was practically identical. The way that the outfit model performed better compared to any of the individual models additionally recommends that the VGG-based and Resnet-based classifiers can finish one another. The VGG16 (without the two FC layers), with 15 million weight boundaries, which is a lot lesser organization than the Resnet50, with 24 million weight boundaries. Having lesser boundaries diminishes memory requests and preparing time per age, which is significant when computational assets. are restricted. The Resnet is an all the newer grew profound learning approach, which is upgraded by alternate ways and. bunch standardization, the two strategies that may help the organization train rapidly

and sum up better. Similar methods can be utilized in the VGG-based organizations too in coming work, which may upgrade the VGG-based. Classifiers. This investigation had a few restrictions. Mammograms were dropped to fit the accessible GPU (8 GB). As farther GPU memory opens up, future examinations will actually want to prepare models utilizing bigger picture estimates, or hold the first picture goal without the requirement for scaling back.

Holding the full goal of new-designed advanced mammography pictures will give better subtleties of the ROIs and likely assistance execution. Albeit the CBIS-DDSM dataset included obsessive confirmation of all malignancy choices, the INbreast dataset didn't. Along these lines, we utilized the radiologists 'BI-RADS evaluations to allocate markers to the pictures in the INbreast dataset. which has the impediment of imitating radiologists ' dazzles as opposed to finding new qualities of harmful injuries. It would be of interest in coming work to incorporate span bosom tumors that were missed by radiologists, to help train calculations to decide more inconspicuous indications of danger that may not be outwardly obvious. At last, the CBIS-DDSM and INbreast datasets weren't broadly delegate tests and execution measurements. in these datasets aren't straightforwardly coordinating to public evaluations of radiologists ' perception and unequivocally. Coming direct examinations among calculations and radiologists will be facilitated by open sharing of the code and higher openness of agent benchmarking datasets.

Taking everything into account, our examination shows that profound learning models prepared in a start to finish design can be significantly precise and possibly promptly shippable across various mammography stages. Profound learning techniques can possibly additionally refine the exactness of breast cancer finding on screening mammography as the accessible preparing datasets and computational assets grow. Our methodology may help coming improvement of predominant CAD frameworks that could be utilized to help focus on the most dubious cases to be perused by a radiologist, or as a programmed diverse compilation in the wake of making a soonest autonomous understanding. Our start to finish approach can likewise be applied to other clinical imaging issues where ROI explanations are scant.

Chapter 6: Conclusion and Future Work

To conclude, the developed research paper can affirm that end to end trained models implementing deep learning algorithms can prove to be accurate to a high extent. These models can easily be transferred and used across multiple mammography manifestos. Given the working and proof shown in this paper we can safely say that the deep learning models are able to have a huge impact on increasing the output accuracy of breast cancer detection on screening mammograms, with the ease of obtaining dataset and technological advancements.

The reproduced method may be used as an initial step towards creation of more advanced and highly accurate computer aided design systems which can be used by radiologists to read even the most questionable of the diagnoses. The system can also be used as a double checking agent to aid the deductions without the need to wait for consultation with other radiologists. In future this intricate system can be built upon incrementally, to add other types of medical imaging issues and help detect multiple types of cancerous tumors, not just breast cancer. He hope to make this system as trained for further use such that the limitation of having ROI explanations is eliminated to some extent for training purposes.

References:

Chiang, T. C., Huang, Y. S., Chen, R. T., Huang, C. S., & Chang, R. F. (2019). Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation. *IEEE Transactions on Medical Imaging*, *38*(1), 240–249. https://doi.org/10.1109/TMI.2018.2860257

Malla, P. P., Sahu, S., & Routray, S. (2020). Investigation of breast tumor detection using microwave imaging technique. *2020 International Conference on Computer Communication and Informatics, ICCCI 2020*, 22–25. https://doi.org/10.1109/ICCCI48352.2020.9104205

Slimi, M., Jmai, B., Mendes, P., & Gharsallah, A. (2019). Breast cancer detection based on CPW antenna. *Mediterranean Microwave Symposium*, *2019-Octob*, 3–6. https://doi.org/10.1109/MMS48040.2019.9157301

Wu, H., Gole, R., Ghosh, S., & Basu, A. (2020). Alternative Techniques for Breast Tumour Detection using Ultrasound. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020-July*(Figure 2), 2047–2050. https://doi.org/10.1109/EMBC44109.2020.9175867

R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2014, pp. 580–587.

R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks," *arXiv:1605.06409[cs]*, May 2016.

A. R. Jamieson, K. Drukker, and M. L. Giger, "Breast image feature learning with adaptive deconvolutional networks," in *Proc.SPIE*, 2012, vol. 8315, pp. 831506-831506–13.

J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "CNNs for mammography mass lesion classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 797–800.

G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered Multiview Mammogram Analysis with Pre-trained Deep Learning Models," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 652–660.

N. Dhungel, G. Carneiro, and A. P. Bradley, “Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests,” in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2015, pp. 1–8.

M. G. Ertosun and D. L. Rubin, “Probabilistic visual search for masses within mammography images using deep learning,” in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2015, pp. 1310–1315.

A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, and E. Barkan, “A Region Based Convolutional Network for Tumor Detection and Classification in Breast Mammography,” in *Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings*, G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S.

Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and J. Cornebise, Eds. Cham: Springer International Publishing, 2016, pp. 197–205.

J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. Guevara Lopez, “Representation learning for mammography mass lesion classification with CNNs,” *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248–257, Apr. 2016.

Daniel Lévy and A. Jain, “Breast Mass Classification from Mammograms using Deep CNNs,” *arXiv preprint arXiv:1612.00542*, 2016.

N. Dhungel, G. Carneiro, and A. P. Bradley, “The Automated Learning of Deep Features for Breast Mass Classification from Mammograms,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, andW. Wells, Eds. Cham: Springer International Publishing, 2016, pp. 106–114.

A. S. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, “Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer,” *Invest Radiol*, Feb. 2017.

I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, “INbreast: Toward a Full-field Digital Mammographic Database,” *Academic Radiology*, vol. 19, no. 2, pp. 236–248, Feb. 2012.

L. Shen, “Breast cancer diagnosis using deep residual nets and transfer learning,” 2017. [Online]. Available: https://www.synapse.org/#!Synapse:syn9773182/wiki/426912. [Accessed: 23-Aug-2017].

Yaroslav Nikulin, “DM Challenge Yaroslav Nikulin (Therapixel).” [Online]. Available: https://www.synapse.org/#!Synapse:syn9773040/wiki/426908. [Accessed: 23-Aug-2017].

F. Chollet and others, *Keras*. GitHub, 2015.

Martín Abadi *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015.

Y. Jia *et al.*, "Caffe: Convolutional Architecture for Fast Feature Embedding," *arXiv:1408.5093 [cs]*, Jun. 2014.

T. Chen *et al.*, "MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems," *arXiv:1512.01274 [cs]*, Dec. 2015.

O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Dec. 2014.

S. Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated Breast Imaging Subset of DDSM," *The Cancer Imaging Archive*, 2016.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385 [cs]*, Dec. 2015.

K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Sep. 2014.