Analysis of Breast Cancer Classification for screening and diagnosis

A Project Report

submitted by

K V Gnana Vardhani (CED16I013)

in partial fulfilment of requirements
for the award of the dual degree of

Bachelor Of Technology And Master Of Technology

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, DESIGN AND MANUFACTURING KANCHEEPURAM

May 2021

DECLARATION OF ORIGINALITY

I, **K V Gnana Vardhani**, with Roll No: **CED16I013** hereby declare that the material presented in the Project Report titled **Analysis of Breast Cancer Classification for screening and diagnosis** represents original work carried out by me in the **Department of Computer Science and Engineering** at the Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram.

With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discussions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Place: Chennai

Date: 01.05.2021 **K V Gnana Vardhani**

Certificate

This is to certify that the report titled **Analysis of Breast Cancer Classification for screening and diagnosis**, submitted by **K V Gnana Vardhani (CED16I013)**, to the Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, for the award of dual degree of **BACHELOR OF TECHNOLOGY AND MASTER OF TECHNOLOGY** is a bona fide record of the work done by him/her under my supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Umarani Jayaraman
Project Guide
Assistant Professor
Department of CSE
IIITDM Kancheepuram, 600 127

Place: Chennai

Date: 01.05.2021

Acknowledgements

I sincerely express my gratitude to the project Guide Dr Umarani Jayaraman for her aspiring guidance and advice during the project.

I am thankful for the truth and illuminating views on the number of issues related to project. Her guidance was crucial in ensuring that I was able to learn as well as be productive during the project. I would like to express and extend my appreciation and gratitude to my mentors for their valuable guidance, professional knowledge, help, constructive criticism and frequent supervision through out the final year project helped to broaden my knowledge.

In conclusion, I would like to thank my parents, friends. They have been driving force throughout my internship journey and have kept me motivated during my toughest times as well.

Abstract

In this day to day life and era Breast malignancy growth is one of the primary issues looked by women. Recognizing malignancy or carcinoma growth is the primary stage and is continually difficult. Beginning phases of breast cancer improvement can be identified by radiologists with the help of computer based diagnosis. After screening, biopsy is used to confirm the cancer, histopathology classification using CNN with breakhis dataset. Preprocessing is used to improve the contrast and reduce the noise in histopathological images and otained the overall accuracy of classification model without using BatchNormalization is 83%. End to End CNN model perfonmance is improved 5% accuracy with swish activation function and improving the number of layers in CNN architecture. Solving the class imbalance issue using under sampling, end to end CNN model with swish activation function and batch normalization got the accuracy improve to 89% and AUC to 0.9624. Decision level fusion parallel ensemble model is incorporated to improve the performance further to 98 to 99% validation accuracy.

KEYWORDS: Breast Cancer; Classification; Histogram equalization; Convolutional Neural Network; Deep Neural Networks; Batch Normalization; Pre-processing

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i

ABSTRACT ii

LIST OF TABLES v

LIST OF FIGURES vi

ABBREVIATIONS vii

1 Introduction 1

1.1 Background 2

1.2 Motivation 2

1.3 Problem Formulation 3

1.4 Objectives of the work 4

2 Literature Survey 5

2.1 Attention Guided CNN 6

2.1.1 Introduction 6

2.1.2 Overview 6

2.1.3 Limitation 6

2.2 Guided Soft Attention Model 7

2.2.1 Introduction 7

2.2.2 Overview 7

2.2.3 Limitation 7

2.3 Patch Classifier with concatenated deep Features 7

2.3.1 Introduction 7

2.3.2 Overview 8

2.3.3 Limitation 8

2.4 Second Order Pooling 8
2.4.1 Introduction 8
2.4.2 Overview 8
2.4.3 Limitation 9

3 Methodology **10**
3.1 Pre-processing 10
3.2 Random Sampling 11
3.3 Feature Extraction and Classification 13
3.3.1 Vgg-16 13
3.3.2 ResNet50 15
3.3.3 Inception 17
3.3.4 CNN: 17
3.3.5 Improved End to End CNN: 18
3.3.6 Decision Fusion with Ensemble Model: 19

4 Work Done and Result Analysis **22**
4.1 Data Description 22
4.2 Data Preperation and Pre-processing 23
4.3 Classification 24
4.3.1 Proposed Model 24
4.3.2 Improved CNN model 24
4.3.3 Result analysis on deep CNN model 26
4.3.4 Fusion Model 28

5 CONCLUSION **30**
5.1 Conclusion 30

REFERENCES **31**

LIST OF TABLES

3.1 Simple CNN model 13

4.1 BreakHis Data Statistics 22

4.2 Data Stat: Under Sampling 24

4.3 Results of Histopathology 25

4.4 Results of Histopathology after under Sampling 26

4.5 Results of Histopathology after under Sampling 27

4.6 Decision Fusion 28

4.7 Parallel Ensembling 29

LIST OF FIGURES

3.1 Block Diagram 11

3.2 Under Sampling 12

3.3 Over Sampling 12

3.4 VGG-16 Architecture 14

3.5 Resnet50 Architecture 16

3.6 Inception Architecture 18

3.7 CNN Architecture 19

3.8 Improved CNN Architecture 20

4.1 Histopathology Pre-processing 23

4.2 End to End CNN model Graphs 25

4.3 ReLu vs Swish 26

4.4 End to End CNN model Graphs 27

Abbreviations

CLAHE	**C**ontrast **L**imit **A**daptive **H**istogram **E**qualization
SCDA	**S**caling and **C**LAHE **D**ata **A**ugmentation
CNN	**C**onvolution **N**eural **N**etwork
FNN	**F**uzzy **N**earest **N**eighbours
FRNN	**F**uzzy **R**ough **N**earest **N**eighbours
VQNN	**V**aguely **Q**uantified **N**earest **N**eighbours
KNN	**K** **N**earest **N**eighbours
SVM	**S**upport **V**ector **M**achine
BN	**B**atch **N**ormalization
MRI	**M**agnetic **R**esonance **I**maging

Chapter 1

Introduction

Breast carcinoma is the second most frequent issue for the cause of death in females. It is caused when some breast cells start to grow uncontrollably and abnormally. These tumour cells divide and grow faster and more rapidly than the normal healthy cell. They further continue to accumulate masses and result in formation of lumps. These cells might unfold (metastasize) from breast to different body parts or organs.

Early detection significantly improves the prognosis, it will increase the success of treatment, save lives, and reduce cost. Mammography is one of the foremost un-remarkably used screening techniques for detection of breast carcinoma, reported to be responsible for a twenty to forty out of hundred women reduction in carcinoma. Due to the cons of existing diagnostic procedure automating early and primary stage detection with better accuracy is crucial. Precise and accurate computerized diagnostic systems have become more important to patients, radiologists and also for clinical practice.

Traditional CAD systems have basically built on known options. Deep neural networks have succeeded in other fields in recent times, one could apply those methods on histopathology.

More than ninety percent of female who were diagnosed with breast malignancy at the earlier stages have probably survive. For at least five yrs around fifteen out of hundred females diagnosed with the most critical condition of the cancer issue, is possible with AI.

1.1 Background

For peer detection of breast carcinoma one ought to take frequent screening mammograms. There are several diagnostic techniques such as MRI screening, Mammography, Ultrasound screening before going to Biopsy. Breast ultra sound is not utilized as an evaluating apparatus for breast disease identification since it doesn't generally distinguish some early indications of malignancy like miniature calcifications, which are minuscule calcium deposits.

MRI isn't guided as an evaluating instrument for women who square measure at normal danger of creating malignancy - a significant drawback is that breast MRI screening brings about more bogus positives. X-ray is typically suggested exclusively along with different tests, like mammogram or ultrasound. X-ray is additionally more costly than mammography for a standard screening.

Malignant growth is one of the far and wide dangerous illnesses that would now be able to be recognized through AI and AI-empowered computerized machines. Throughout the most recent couple of many years, huge work has been finished in regards to breast malignancy growth analysis. Initially CADx systems used image processing techniques and traditional classification methods such as kNN, SVM, NaiveBayes, random forest, later on focus was shifted towards CNNs and transfer-learning for efficient resuts.

1.2 Motivation

When the tumor is affirmed through the manual screening or CAD, biopsy is performed. Eliminating an example of breast cells for testing (biopsy). All through the biopsy test, specialist utilizes a particular needle gadget radio-constrained by partner X-beam or another imaging method to remove a center of tissue from the space of doubt. Test outcomes are shipped off a research facility for investigation any place radiologists confirm whether the cells are destructive. These biopsy test is likewise checked and dissected

to confirm such a cells associated with bosom malignancy. The images taken through biopsy test under the assessment of magnifying lens are called histopathological images.

The long understanding and quality, intricacy, and varieties in biopsy tests can bring about a high bogus positive rate and all the more critically, misdiagnosis of manual screening. Manual arrangement of histopathological pictures of breast malignant growth tissues gets exhausted, dreary, and delicate to the abstract impact.So breast cancer diagnostic advancement is significant with the introduction of automation in the detection of breast cancer.

1.3 Problem Formulation

Given Histopathological images I_m, the final aim of our method is to classify whether the cancerous lesion is present in the image or not. It has two classes M-Malignant and B-Benign. I_m is the set of the input images and it is processed and enhanced. Here Function P(.) is a combination of the other two functions i.e. E(.) and T(.). E(.) is an noise removal function and T(.) is a resizing function.

I^t= C[P(X)]

where P(X)= E(T(X))

and C(X) is a CLAHE contrast enhancement.

The CNN feature extraction and classification process is denoted as [$prob_M$, $prob_{\mathrm{B}}$] = $Conv_{end}(I_i, \theta_{ei})$ (i $\in$ Biopsy images). The training images are given to the feature extraction and classification function defined by ConvNet end to end module $Conv_{end}$(.). $prob_M$ and $prob_{\mathrm{B}}$ are class probabilities from the layer with full connectivity with the features detected and extracted from Conv NN modules θ_i denotes the parameters of CNN for modality I to be learned. The output of the last max-pooling layer is the features extracted and given to fully connected layers to give class probabilities. The altered fine-tuned models are altered with additional batch normalization layers. Decision Fusion is incorporated using the trained models. Here w_i is the weight given to the

respective model

$$[prob_M, prob_{\mathrm{B}}] = \frac{\sum_{i=1}^{n} w_i \times Conv_{\mathrm{end}}(I_i, \theta_{\mathrm{e}i})}{\sum_{i=1}^{n} w_i}$$

1.4 Objectives of the work

- The major objective is to have a detailed literature analysis about the histopathology and analyse the current trends and developments in breast histology.
- Basic classification model of histopathology. A simple End to End classifier is proposed taking the histopathological images classify them into malignant and benign.
- Solve the class imbalance issue in histopathological classification model with random sampling and augmentation.
- Improve the contrast of the medical images. A pre-processing method CLAHE, with noise removal method, is incorporated.
- Improve the classification model. Due to limited training data, we carefully designed a classification module with Swish activation function and batch normalization to improve the training.
- Deep NNs and end to end CNNs are extensively used and we devised various model architectures that can automatically classify the lesion map. Deep NN models are taken and few alterations are made on top of that to improve and analyzed the model performance with and without batch normalization layers.
- Building and analysing the outcomes using decision fusion model with the models that are previously built.

Chapter 2

Literature Survey

Automation for breast histology began in the 1990s. In the course of the most recent couple of years, noteworthy work has been done with respect to a breast cancer diagnosis. At first CAD frameworks utilized image processing strategies and customary arrangement techniques, traditional machine learning classification models, later on, it was moved towards CNNs and move learning for effective outcomes. These CAD systems are designed aiming to increase efficiency and effectiveness. A complete CAD system consists of various steps i.e. pre-processing, classification of abnormal cases, detecting the abnormality, classification of abnormality type. For a CAD system, performing more accurate initial steps are crucial as they show an impact on the further process. Complete CAD system consists various steps i.e. pre-processing, classification of malignant cases, detecting the abnormality, classification of abnormality type. For a CAD system to perform more accurate initial steps are crucial as they show impact on the further process.

Following are the few methods explored earlier automation of procedure.

2.1 Attention Guided CNN

2.1.1 Introduction

An IEEE Transactions on Biomedicals-Health informatics, 2020 have proposed a three stage classification and detection architecture for histopathological images of breast.[35]

2.1.2 Overview

Train the fix based CNN to acquire the area of the mitotic competitors. After the back-bone network separated the highlights of the information picture, we utilize a RPN to produce a bunch of k mitotic. Generates region proposals using anchors of different sizes and uses NMS. It's subnet slides a small window uses scales and aspect ratios to the center of the candidates to generate reference boxes. Attention module re-encode the extricated highlights during the back-engendering of the model. It utilized the at-tention activity to feature the closer view locale in the common component map, which upgrades the heaviness of the frontal area district during back-spread to help the element encoding capacity. weights are allotted utilizing the sigmoid and redressed straight join initiation work. Multi-branch grouping subnet comprises of three branches, a bouncing box relapse branch, and two arrangement branches.(85 precision)

2.1.3 Limitation

The basic con of the attention methodology is that it adds more weight parameters to the model, which can increase the time especially take for training the input data for the model are long sequences.

2.2 Guided Soft Attention Model

2.2.1 Introduction

IEEE Transactions on Medical Imaging, 2019. They have utilixed a guided soft attention network to classify through CNN.[36]

2.2.2 Overview

A conv nn with extra locale level management which gives express direction on the attention maps for the grouping of bosom disease histopathological pictures was proposed in this investigation. Guided Soft Attention (GuSA) network which points at restricting these RoIs, and all the while use them to manage the NN. The NN comprises of two branches; (a) a RoI expectation branch to limit the symptomatically pertinent locales, and (b) a grouping branch to distinguish the sort of bosom histology image. The foundation of the network is a Conv NN based component extraction organization. The highlights at various layers of the highlight extraction network are collected, diminished and afterward taken care of to the RoI forecast branch and the grouping branch.(78% accuracy)

2.2.3 Limitation

Classification for histopathology is to confirm the presense of tumor, inter class similarities between the sub classes of malignant and benign are high, so separate cnn models would perform better.

2.3 Patch Classifier with concatenated deep Features

2.3.1 Introduction

IEEE Journal of Biomedical and Health Informatics, 2020.[37]

2.3.2 Overview

Proposed a profound feature extraction structure that forms a ROI-level element portrayal through weighted conglomeration of the portrayals of variable quantities of fixed-sized patches tested from cores thick districts in breast histopathology samples.

Techniques: First, the underlying patch-level component portrayals are separated from both fully connected layer initiations and pixel-level conv layer activations of a Deep NN, and the weights are gotten from the class expectations of a similar organization prepared on fix tests. At that point, the last patch level component portrayals are processed by connection of weighted examples of the removed element enactments. At last, the ROI-level portrayal is acquired by combination of the fix level portrayals by normal pooling.(72% accuracy)

2.3.3 Limitation

The weights obtained from class predictions are used to concatenate and extract the ROI through weighted instances. The accuracy of classification depends on the further ROI detection. The performance of classification can be highly improved.

2.4 Second Order Pooling

2.4.1 Introduction

International Joint Conference on Neural Networks (IJCNN) 2020.[39] The deep second-order pooling network is proposed.

2.4.2 Overview

Second-order measurements can catch the relationship between highlight measurements and altogether improve model order execution. Particularly, worldwide second-order covariance portrayals of profound highlights dependent on lattice power normalization

have outlined splendid execution on huge scope visual acknowledgment undertakings with covariance normalization in covariance pooling module. (89% accuracy)

2.4.3 Limitation

Incorporating second order pooling layers in pre-trained models is onerous. The input data is large in number and training an end to end model from scratch with second order pooling is time consuming.

Chapter 3

Methodology

An overview of the proposed network architecture, is illustrated in Fig. 1. For the first part, we take the I as inputs to pre-process using CLAHE i.e. and contrast enhancement with CLAHE and techniques are used to reduce noise, delete scratches and improve contrast. Once it is pre-processed, the images are passed through feature exatraction and classification module. They are End to End model and altered deep NN models. Altered fine-tuned models are pre-trained models such as VGG-16, Resnet, Inception, Mobilenet, etc. In each module, we analyze the classification performance with using batch normalization layers additionally. As patch classifier, CNN stream is combined with Batch Normalization Layers using max pooling. Few pre-processing CNN is used for classification.

3.1 Pre-processing

Before we give the data to classification module, images needs to be processed. Data is supposed to be messy in the event that it is missing trait, characteristic qualities, contain clamor or exceptions and copy or wrong information. Presence of any of these will corrupt nature of the outcomes. So to make consistency in the images, they are resized. To eliminate salt and pepper noise and scratches from biopsy images seen through microscope median blur is applied. The image patches from microscope are normalized using histogram equalization.

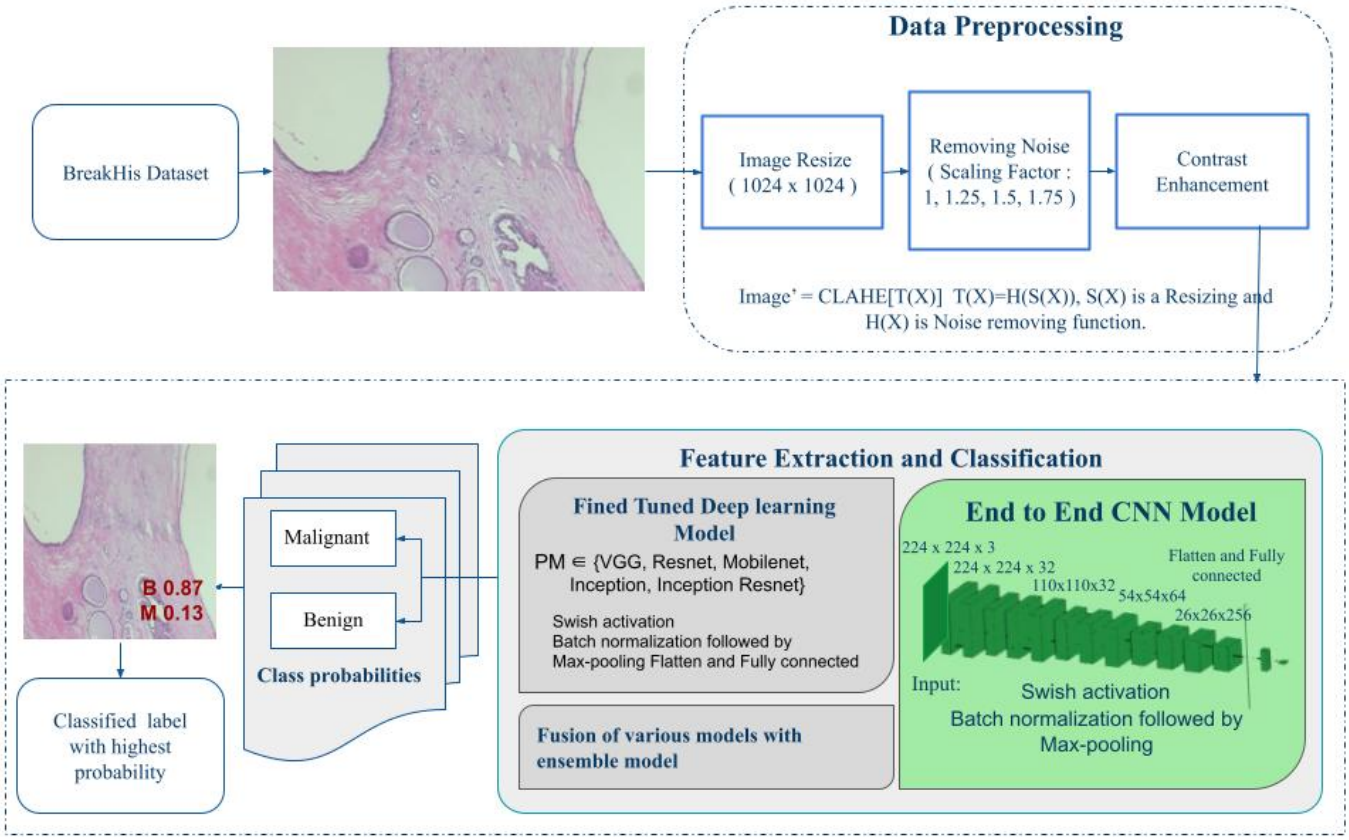


Figure 3.1: Block Diagram

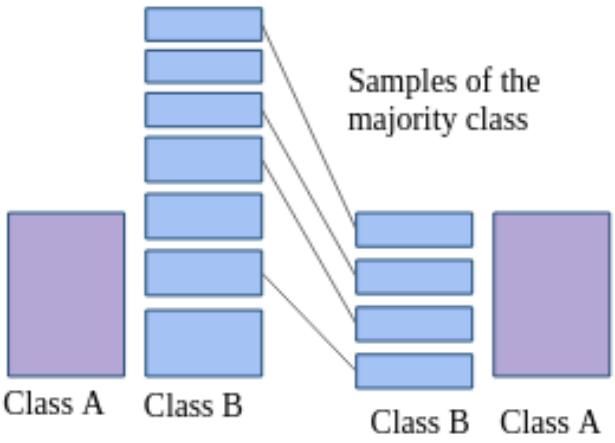
Contrast of the image can be improved using histogram equalization. Global equalization would be enough rather localization as the biopsy images are patches of cells seen through microscope.

3.2 Random Sampling

The Imbalanced classification issue is what we face when there is a genuine slant in the class appropriation of our preparation data.
An approach to manage this difficult test is Random Sampling. There are two crucial ways to deal with perform random resampling.

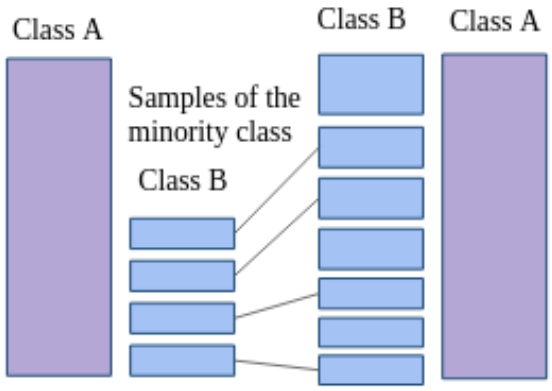
Thusly, Both oversampling and undersampling incorporate familiarizing an inclination with select a more prominent number of samples from one class than from another, to compensate for an imbalance that is either adequately present in the data, or inclined to make if a totally random model were taken.

Figure 3.2: Under Sampling



!htb

Figure 3.3: Over Sampling



The random oversampling may improve the likelihood of overfitting occurring, since it makes exact of the minority class models. Thusly, a meaningful classifier, for instance, may fabricate concludes that are clearly accurate, anyway straight one imitated example.

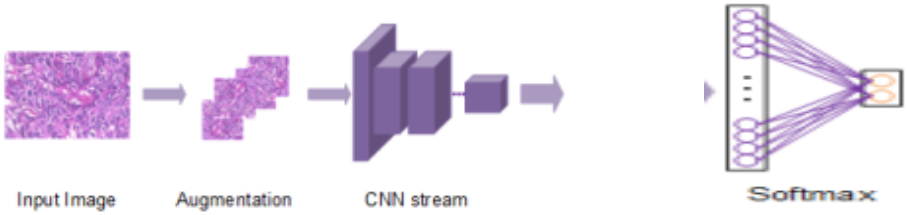
In random under-sampling (conceivably), huge measures of data are discarded. [. . .] This can be outstandingly perilous, as the lack of such data can make the choice boundary between the minority and larger part cases all the more excitedly to dominate,

achieving a setback in classification performance.

3.3 Feature Extraction and Classification

Improved CNN model and other deep NNs are used for feature extraction and classification of patches from biopsy into malignant or benign. The specific architectures are already mentioned above.

Table 3.1: Simple CNN model



The conv neural network is a part of Deep NN. These Deep NNs are most regularly used to investigate visual symbolism and are every now and again working in the background in image classification. For feature extraction convolutional neural network is used with classification fully connected layers.

3.3.1 Vgg-16

It is a pre-trained CNN model.VGG addresses another very important aspect of CNNs: depth. The quantity of layers was expanded from eight in AlexNet [32] to sixteen or nineteen, contingent upon the model, in VGGNet with a little convolutional channel size. VGG takes in a 224x224 pixel Red, Green, Blue image.

Convolutional Blocks: In this organization, you can distinguish a structure block as two Conv. layers followed by a pooling layer. This block is rehashed on different occasions. Something else, in VGGNet all the channel sizes utilized are three cross three

Figure 3.4: VGG-16 Architecture

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0

Total params: 14,714,688
Trainable params: 14,714,688
Non-trainable params: 0

.To join three non-straight amendment layers rather than a solitary one, which settles on the decision work more discriminative.

Reduction the quantity of boundaries: expecting that both the info and the yield of

a three-layer 3 × 3 convolution stack has C channels, the stack is parametrized by

$$3(3^2C^2) = 27C^2$$

Loads; simultaneously, a solitary 7 × 7 Conv. layer would require

$$7^2C^2 = 49C^2$$

boundaries, for example 81% more. Pooling decreases the dimensions of your data exponentially.
And so even if you have an image of size 256 x 256, you only get maybe 5 pools before you lose too much information (and therefore 5 convolutions). As a result, we would typically like to have multiple Conv layers before a pool.

3.3.2 ResNet50

The super deep architecture used a stack of residual blocks and each block had two 3x3 convolutional layers. The concept of residual mapping resolved the optimization problem in deeper models.
The skip connections or residual blocks in ResNet help the architecture to avoid gradient diminishing problems.
During the back propagation stage, the error is calculated and gradient values are determined. The gradients are sent back to hidden layers and the weights are updated accordingly. The process of gradient determination and sending it back to the next hidden layer is continued until the input layer is reached. The gradient becomes smaller and smaller as it reaches the bottom of the network. Therefore, the weights of the initial layers will either update very slowly or remains the same.

The very profound design utilized a pile of residual blocks and each had 2 three cross three conv layers. The idea of residual planning settled the advancement issue in more profound models.

The skip associations or residual blocks in resnet assist the design with keeping away from gradient reducing problem.

During the back propagation stage, the minute error is determined and gradient esteems are resolved. The gradients are sent back to covered up layers and the loads are updated in like manner. The cycle of gradient assurance and sending it back to the following hidden layer is proceeded until the input layer is reached. The gradient decreases and more modest as it arrives at the lower part of the NN. Along these lines, the loads of the underlying layers will either update gradually or stays as before. Initially it has

Figure 3.5: Resnet50 Architecture

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 7, 7, 2048)	23587712
global_average_pooling2d_3 (	(None, 2048)	0
dropout_10 (Dropout)	(None, 2048)	0
batch_normalization_31 (Batc	(None, 2048)	8192
dense_34 (Dense)	(None, 2)	4098

Total params: 23,600,002
Trainable params: 23,542,786
Non-trainable params: 57,216

conv(7*7) with batch normalization and max pooling layers. This is further followed by 4 sets of convolution and residual blocks respectively. First set has 2 identity blocks, second set has 3 identity blocks, third set has 5 identity blocks, and fourth set has 2 identity blocks. Finally resnet 50 have avg pooling with softmax activation function.
Each convolution block has three convolution layers with batch normalization and activation between base and add. There is a direct convolutional layer between base and add as a skip connection.
Each identity block has three convolution layers with batch normalization and activa-

tion between base and add. There is a direct weight between base and add as a skip connection.

3.3.3 Inception

This Inception architecture utilizes procedures, for example, one cross one conv in the design and global avg pooling. An inception module is the essential structure square of this network.
So, the Inception module does various conv, with various channel sizes, and just as pooling in single layer.

An issue with profound conv NN is that the quantity of feature maps frequently increments with the profundity and depth of the network. This issue can bring about a sensational and drastic expansion in the quantity of parameters and calculation required when bigger channel sizes are utilized, for example, five cross five and seven cross seven.
So a stand by solution for this issue, a one cross one conv layer could be utilized that offers a channel by channel pooling. This basic procedure is utilized for dimensionality decrease, diminishing the quantity of feature maps while holding their striking features. Global avg pooling is utilized toward the ending of the NN. This layer takes a feature guide of seven cross seven and midpoints it to one cross one.

3.3.4 CNN:

The CNN built has two convolutional blocks followed by maxpooling and dropouts. Each block ha two convolutional layers. After extracting features from CNN layers give it to the dense layers (1024,512,256,2) for classification. Built This is binary classification, converted to one-hot representation we 2 output neurons.

Figure 3.6: Inception Architecture

Layer (type)	Output Shape	Param #
inception_v3 (Functional)	(None, 5, 5, 2048)	21802784
global_average_pooling2d_5 (	(None, 2048)	0
dropout_12 (Dropout)	(None, 2048)	0
batch_normalization_330 (Bat	(None, 2048)	8192
dense_36 (Dense)	(None, 2)	4098

Total params: 21,815,074
Trainable params: 21,776,546
Non-trainable params: 38,528

3.3.5 Improved End to End CNN:

The CNN built has two convolutional blocks followed by maxpooling and Batch Normalization layers. Each block ha two convolutional layers. Increased the number of convolution and decreased number of dense layers to the previus model built for intern. After extracting features from CNN layers give it to the dense layers (1024,2) for classification. Built This is binary classification, converted to one-hot representation we 2 output neurons. Swish activation function is used instead of relu, which is differentiable at zero, and makes it easy to learn.

The ReLU function is an improvement over tanh because it is bounded above — this property is so important that every successful activation function after ReLU is unbounded above. Being bounded below may be advantageous because of strong regularization — This is important at the beginning of training when large negative activation inputs are common.

Swish provides smoothness which helps optimize and generalize the neural network. In the output landscapes below, it is obvious that ReLU's output landscape is sharp and jarring because of its non-smooth nature, whereas the Swish network landscape is much

Figure 3.7: CNN Architecture

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 224, 224, 32)	2432
conv2d_5 (Conv2D)	(None, 220, 220, 32)	25632
max_pooling2d_2 (MaxPooling2	(None, 110, 110, 32)	0
dropout_3 (Dropout)	(None, 110, 110, 32)	0
conv2d_6 (Conv2D)	(None, 110, 110, 64)	18496
conv2d_7 (Conv2D)	(None, 108, 108, 64)	36928
max_pooling2d_3 (MaxPooling2	(None, 54, 54, 64)	0
dropout_4 (Dropout)	(None, 54, 54, 64)	0
flatten_1 (Flatten)	(None, 186624)	0
dense_2 (Dense)	(None, 1024)	191104000
dense_3 (Dense)	(None, 512)	524800
dense_4 (Dense)	(None, 256)	131328
dense_5 (Dense)	(None, 2)	514

Total params: 191,844,130
Trainable params: 191,844,130
Non-trainable params: 0

smoother.

In the End to End CNN architecture we have incorporated batch normalization layers. Batch normalization improves the training curve and improves the performance of the model. It computes the normalized features among the single mini-batch with mean and variance and then forwards it to the next layer. We have used the Swish activation function instead of ReLu.Swish activation function used in this end to end CNN architecture. Swish is unbounded above and bounded below like ReLu activation function but it is non-monotonic and smooth

3.3.6 Decision Fusion with Ensemble Model:

It combines the decision of multiple classifiers and gives the resultant class probabilities. It gives resultant probabilities based on the weights given to the respective model outcome. It gives resultant probabilities based on the weights given to the respective

Figure 3.8: Improved CNN Architecture

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 224, 224, 32)      2432
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 220, 220, 32)      25632
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 110, 110, 32)      0
_________________________________________________________________
batch_normalization (BatchNo (None, 110, 110, 32)      128
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 110, 110, 64)      18496
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 108, 108, 64)      36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 54, 54, 64)        0
_________________________________________________________________
batch_normalization_1 (Batch (None, 54, 54, 64)        256
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 54, 54, 128)       73856
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 52, 52, 128)       147584
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 26, 26, 128)       0
_________________________________________________________________
batch_normalization_2 (Batch (None, 26, 26, 128)       512
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 26, 26, 256)       295168
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 24, 24, 256)       590080
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 12, 12, 256)       0
_________________________________________________________________
batch_normalization_3 (Batch (None, 12, 12, 256)       1024
_________________________________________________________________
flatten (Flatten)            (None, 36864)             0
_________________________________________________________________
dense (Dense)                (None, 4096)              150999040
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              4195328
_________________________________________________________________
batch_normalization_4 (Batch (None, 1024)              4096
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 2050
=================================================================
Total params: 156,392,610
Trainable params: 156,389,602
Non-trainable params: 3,008
```

model outcome. Ensemble techniques can be utilized as generally demonstrative methods for a more customary model structure. The bigger the difference in fit quality between one of the more grounded ensemble techniques and a customary factual model, the more data that the ordinary model is most likely absent. Ensemble techniques can be utilized to assess the connections between informative factors and the reaction in ordinary measurable models. Indicators or premise capacities neglected in an ordinary model may surface with an ensemble approach. With the assistance of the ensemble technique, the determination interaction could be better caught and the likelihood of enrollment in every treatment bunch assessed with less bias.

Chapter 4

Work Done and Result Analysis

4.1 Data Description

To successfully assess the presentation of the proposed model, utilized bosom disease histopathological picture dataset, for example BreaKHis dataset. BreaKHis dataset comprises of 7917 obsessive images from eighty two victims, in which each obsessive image has been clarified with a benign or malignant mark. In addition, two thousand four hundred and eighty eight examples have a place with Benign images and the stayed five thousand four hundred and twenty nine examples are malignant images. As indicated by the distinctive amplification factors, tests of every quiet can be ordered into four gatherings of forty, hundred, two hundred and four hundred, separately. It has further sub classes

under malignant and benign.

Table 4.1: BreakHis Data Statistics

class	Training	Validation	Total
Benign	1991	497	2488
Malignant	4343	1086	5429

- Malignant
  1. Ductal Carcinoma

2. Lobular Carcinoma
3. Mucinous Carcinoma
4. Papillary Carcinoma

- Benign
  1. Adenosis
  2. Fibroadenoma
  3. Tubular Adenoma
  4. Phyllodes Tumor

4.2 Data Preperation and Pre-processing

The comma-separated values (CSV) file for corresponding datasets contains many columns that are unnecessary for classification especially patient ID, Abnormality type, Age, View, Date, Sex, have been dropped from the CSV file. Further, the NaN values have also been removed. The images in the folders have different names from those given in the CSV file. Images have complete descriptions whereas a CSV file has just a unique code present in the image file name. So for convenience, and modified the CSV image names column to make it the same as image names in the folder.
One and all sample images with the size of seven hundred cross four sixty pixels in Red, Green, Blue channel pattern with an eight bit colour depth per each channel. Resized all the images to 224 x 224 pixel in RGB channel pattern. Images are processed using median blur with kernel size 5 x 5. Equalization of Histogram is done globally to so that the contrast of the image would improve .

```
image = cv2.resize(image, (224,224))
image = cv2.medianBlur(image, 5)
image = cv2.cvtColor(image, cv2.COLOR_RGB2HSV)
image[:, :, 2] = cv2.equalizeHist(image[:, :, 2])
image = cv2.cvtColor(image, cv2.COLOR_HSV2RGB)
cv2.imwrite(img_path,image)
```

Figure 4.1: Histopathology Pre-processing

Table 4.2: Data Stat: Under Sampling

class	Training	Validation	Total
Benign	1991	497	2488
Malignant	1991	497	2488

4.3 Classification

4.3.1 Proposed Model

Built an end to end CNN from scratch with 8 convolutional layers, four max pooling, and five batch normalization layers. The proposed CNN model has four blocks of 32, 64, 128, and 256. Each convolutional block has 2 conv layers and it can be clearly observed that a set of conv layer is followed by a maximum pooling layer. The extracted features from convolutional layers are flattened and given further to dense layers.
With the original data, generated the processed images and applied clahe to the images. Considering the processed data I have trained various pretrained models and a CNN model built. Pretrained models are given initial weights of Imagenet classification problem. To over come the class imbalance problem - give class weights for the model.
class weights = [class normal:1/902 class abnormal: 1/1442]
Number of Images: 2342
Image Size: (224,224,3)
Optimizer: Adam
Epochs : 15
Activation Funtion : Relu, Softmax
The same layers are used even for other pretrained models i.e. dense and output layers after feature extraction.

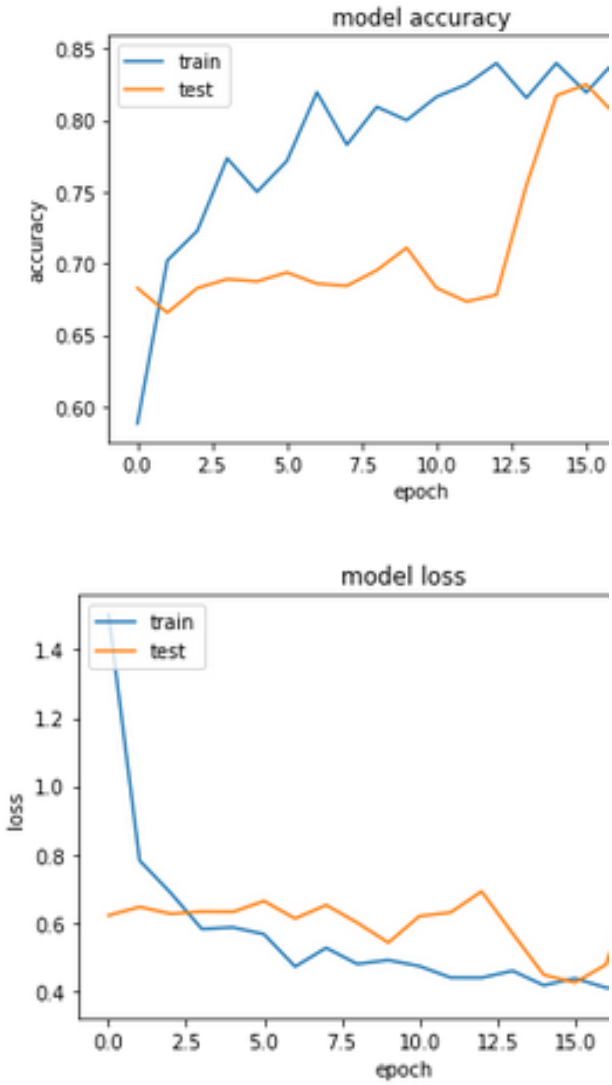
4.3.2 Improved CNN model

A maximum pooling layer and batch normalization layer are headed up with each and every set of Conv layer. The extracted features from conv layers are flattened given

Table 4.3: Results of Histopathology

Model	Val Accuracy	AUC	Specificity	Sensitivity
End to End CNN	83	0.91	0.84	0.81
Resnet50 BN	71	0.81	0.74	0.75

Figure 4.2: End to End CNN model Graphs



further to dense layers. This has two dense fully connected layers for classification. We have used the Swish activation function instead of ReLu. Swish outperformed ReLu in this problem. Various models are used for classification of patches. Histogram equalization is applied on the images. Learning rate is reduced with a factor 0.5, if the patience reaches 3. Number of images : 7917

Image Size : (224,224,3)

Epochs: 30

Class balancing: Random under sampling

End to End CNN model and resnet50 are used to compute the results.

- Image Size: (224,224,3)
- Optimizer: Adam
- Activation Funtion : Swish, Softmax
- Batch Size is 64

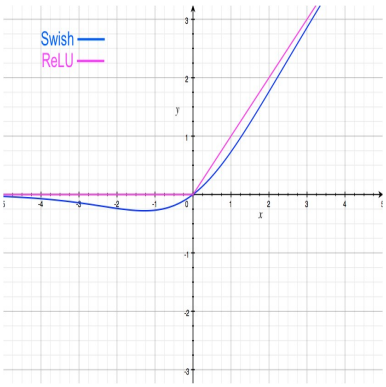


Figure 4.3: ReLu vs Swish

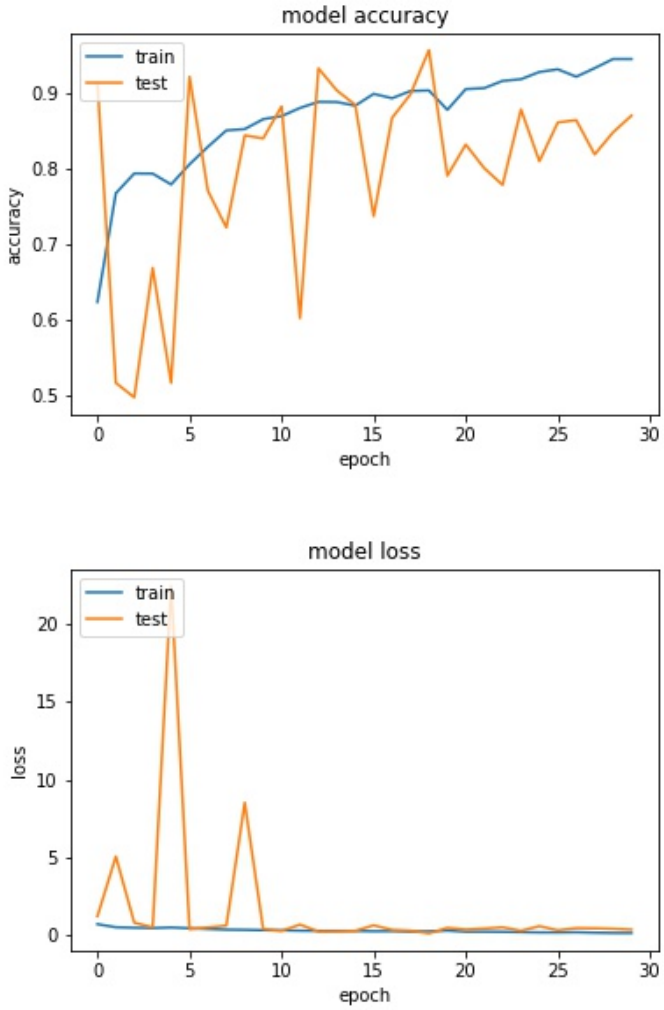
Table 4.4: Results of Histopathology after under Sampling

Model	Val Accuracy	AUC	Specificity	Sensitivity
End-End CNN	88.72	0.9624	0.91	0.9
ResNet50	92.145	0.9567	0.9804	0.9807

4.3.3 Result analysis on deep CNN model

Incorporating the Swish activation function and batch normalization on various models. We have incorporated 3 dense layers and batch normalization layers for classification. Features from these models are passed through other proposed layers for classification. Initially, models are trained on ImageNet dataset. Weights of these models have been taken as initial weights of the architecture. Further, these models are fine-tuned and trained with a breast image dataset. The last two layers of the model are updated and

Figure 4.4: End to End CNN model Graphs



all the other layers are frozen. We have incorporated a few other layers on top of pre-trained models. Features from these models are passed through other proposed layers i.e. dense and batch normalization layers for classification. From table 4.5, Resnet150,

Table 4.5: Results of Histopathology after under Sampling

model	Traacc(%)	Valacc (%)	sensi	speci	auc
VGG16	95.41	95.67	0.952	0.9519	0.9444
ResNet50	97.85	92.145	0.9804	0.9807	0.9567
InceptionResnet	99.76	89.12	0.8939	0.893	0.952
Resnet150	88	90.634	0.963	0.96	0.9674
Inception	97	90.53	0.905	0.9	0.9667
CNN	93.10	88.72	0.91	0.9	0.9624

Inception, and End to End CNN have better area under the curve.

4.3.4 Fusion Model

Parallel ensembling method is used With the previously trained multiple models. I have build a decision fusion model. The outcomes of these models are taken in various weights and finalized the decision. Using the fusion of decision probabilities from various models we have given obtained the classifica- tion results. They are tabulated in Table 4.6. Corresponding weights are given to the specific model are given in the table.

Table 4.6: Decision Fusion

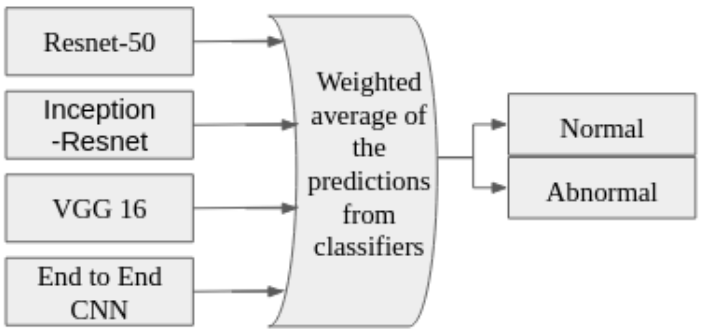


Table 4.7: Parallel Ensembling

CNN	Resnet150	Inception	VGG	Valacc (%)
0.5	**0.5**	**0**	**0**	**0.9566968781470292**
0.6	0.4	0	0	0.9164149043303121
0.4	0.6	0	0	0.9244712990936556
0.7	0.3	0	0	0.9234642497482377
0.3	0.7	0	0	0.9133937562940584
0.8	0.2	0	0	0.9244712990936556
0.2	0.8	0	0	0.9113796576032226
0.9	0.1	0	0	0.9234642497482377
0.1	0.9	0	0	0.9113796576032226
0.5	0	0	0.5	0.9355488418932527
0.6	0	0	0.4	0.945619335347432
0.4	0	0	0.6	0.972809667673716
0.7	0	0	0.3	0.9365558912386707
0.3	0	0	0.7	0.9778449144008057
0.8	0	0	0.2	0.9305135951661632
0.2	0	0	0.8	0.9788519637462235
0.9	0	0	0.1	0.9264853977844915
0.1	**0**	**0**	**0.9**	**0.9768378650553877**
0.5	0.5	0.5	0.5	0.8680765357502518
0.4	0.6	0.6	0.4	0.8429003021148036
0.4	0.6	0.4	0.6	0.8841893252769386
0.4	0.4	0.6	0.6	0.8499496475327291
0.6	0.6	0.4	0.4	0.877139979859013
0.6	0.4	0.6	0.4	0.8277945619335347
0.6	0.4	0.4	0.6	0.904330312185297
0.3	0.7	0.6	0.4	0.8580060422960725
0.3	0.7	0.7	0.3	0.8459214501510574
0.3	0.7	0.3	0.7	0.9073514602215509
0.7	0.3	0.4	0.6	0.8761329305135952
0.7	0.7	0.3	0.3	0.8831822759315207
0.6	0.7	0.4	0.3	0.865055387713998
0.7	0.4	0.7	0.3	0.81067472306143
0.3	0.3	0.3	1.1	0.9889224572004028
0.4	**0.3**	**0.3**	**1**	**0.9979859013091642**
0.4	0.4	0.3	0.9	0.9295065458207452
0.5	0.4	0.3	0.8	0.9204431017119838
0.5	0.5	0.3	0.7	0.9204431017119838

Chapter 5

CONCLUSION

5.1 Conclusion

In this final year project, I have explored and developed an diagnostic model for precise and accurate classification of breast malignancy. We give the biopsy patches which are taken from four different magnifying factors to our CNN, after we improve the data with pre-processing and noise reduction by applying the CLAHE enhancement method on patches.

The purposes of data sampling solves the problem of lack of data. and also helps the model to correctly fit in the data instead of over fitting. To explore and analyze the conv NN models which shows best performance on the two class classification task, we probed the models with various deep NN methods including inceptionV3, residual learning, VGG-16, Mobilenet and New CNN model. Among all of those models which are demonstrated, outcome tells that VGG, Resnet-150 gave the better result . Histopathology, biopsy images are pre-processed and classified using patch classifier. CNN model to classify Histopathology images provided the accuracy of 83% and AUC of 0.91. By solving the class imbalance issue in the histopathological image classification model the performance of the model is improved to 89% after solving class imbalance with under sampling.

End to end model is improved with 5% accuracy and 0.1 AUC with swish activation and batch normalization. Proposed to provide a fusion ensemble model with VGG, ResNet-

150, CNN end to end model and inception with the processed and sampled training set and the performance of parallel ensemble model Batch Normalization using swish activation function performed better with testing accuracy-98 to 99%. Proposed a data fusion model at the decision level. This improved the validation accuracy. Analysis of decision fusion with various models and weights.

Further Extract the regions from histopathology patches for further efficient detection and classification

REFERENCES

[1] X. Yu, C. Kang, D. S. Guttery, S. Kadry, Y. Chen, and Y. Zhang, “Resnet-scda-50 for breast abnormality classification,” IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2020.

[2] X. Shu, L. Zhang, Z. Wang, Q. Lv and Z. Yi, "Deep Neural Networks With Region-Based Pooling Structures for Mammographic Image Classification," in IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 2246-2255, June 2020, doi: 10.1109/TMI.2020.2968397.

[3] R. K. Samala, H. -P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter and K. H. Cha, "Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets," in IEEE Transactions on Medical Imaging, vol. 38, no. 3, pp. 686-696, March 2019, doi: 10.1109/TMI.2018.2870343.

[4] J. Arevalo, F. A. Gonzlez, R. Ramos-Polln, J. L. Oliveira, and M. A.Guevara Lopez, “Representation learning for mammography mass lesion classification with convolutional neural networks,” Computer Methods and Programs in Biomedicine, vol. 127, no. C, pp. 248–257, 2018.

[5] H. Chougrad, H. Zouaki, and O. Alheyane, “Multi-label transfer learning for the early diagnosis of breast cancer,” Neurocomputing, 2019.

[6] X. Zhang et al., "Classification of Whole Mammogram and Tomosynthesis Images Using Deep Convolutional Neural Networks," in IEEE Transactions on NanoBioscience, vol. 17, no. 3, pp. 237-242, July 2018, doi: 10.1109/TNB.2018.2845103.

[7] Dhungel, N., G. Carneiro, and A.P. Bradley. Fully automated classification of mammograms using deep residual neural networks. in Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. 2017. IEEE.

[8] N. Dhungel, G. Carneiro, and A. P. Bradley, “Automated mass detection in mammograms using cascaded deep learning and random forests,” in International Conference on Digital Image Computing: Techniques and Applications, 2016, pp. 1–8.

[9] J. Zheng et al., "3D Context-Aware Convolutional Neural Network for False Positive Reduction in Clustered Microcalcifications Detection," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2020.3003316.

[10] J. Zheng, H. Sun, S. Wu, K. Jiang, Y. Peng, X. Yang, F. Zhang, and M. Li, “3d context-aware convolutional neural network for false-positive reduction in clustered microcalcifications detection,” IEEE Journal of Biomedical and Health Informatics, 2020.

[11] N. Wu et al., "Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening," in IEEE Transactions on Medical Imaging, vol. 39, no. 4, pp. 1184-1194, April 2020, doi: 10.1109/TMI.2019.2945514.

[12] Prajit Ramachandran, Barret Zoph, Quoc V. Le SWISH :A SELF-GATED ACTIVATION FUNCTION

[13] Ioffe, S. and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. in International Conference on Machine Learning. 2015.

[14] 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, 2012.

[15] J Suckling et al (1994): The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series 1069 pp375-378.

[16] T. N. Cruz, T. M. Cruz, and W. P. Santos, “Detection and classification of mammary lesions using artificial neural networks and morphological wavelets,” IEEE Latin America Transactions, vol. 16, no. 3, pp. 926–932, March 2018.

[17] L. Fei-Fei, ImageNet: crowdsourcing, benchmarking and other cool things, CMU VASC Seminar, March, 2010.

[18] D. S. Jacob, R. Viswan, V. Manju, L. PadmaSuresh and S. Raj, "A Survey on Breast Cancer Prediction Using Data MiningTechniques," 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, 2018, pp. 256-258, doi: 10.1109/ICEDSS.2018.8544268.

[19] K. Loizidou, G. Skouroumouni, C. Nikolaou and C. Pitris, "An Automated Breast Micro-Calcification Detection and Classification Technique Using Temporal Subtraction of Mammograms," in IEEE Access, vol. 8, pp. 52785-52795, 2020, doi: 10.1109/ACCESS.2020.2980616.

[20] G. Carneiro, J. Nascimento and A. P. Bradley, "Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning," in IEEE Transactions on Medical Imaging, vol. 36, no. 11, pp. 2355-2365, Nov. 2017, doi: 10.1109/TMI.2017.2751523.

[21] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299–1312, May 2016.

[22] G. Quellec, M. Lamard, M. Cozic, G. Coatrieux, and G. Cazuguel, “Multiple-instance learning for anomaly detection in digital mammography,” IEEE Transactions on Medical Imaging, vol. 35, no. 7, pp. 1604–1614, July 2016.

[23] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He and J. Fan, "Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis," in IEEE Access, vol. 8, pp. 96946-96954, 2020, doi: 10.1109/ACCESS.2020.2993536.

[24] J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, and Y. Yang, “Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances,” IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 2, pp. 236–251, March 2009.

[25] H. Nasir Khan, A. R. Shahid, B. Raza, A. H. Dar, and H. Alquhayz, “Multi-view feature fusion based four views model for mammogram classification using convolutional neural network,” IEEE Access, vol. 7, pp. 165 724–165 733, 2019.

[26] Z. Chen, H. Strange, A. Oliver, E. R. E. Denton, C. Boggis, and R. Zwiggelaar, “Topological modeling and classification of mammographic microcalcification clusters,” IEEE Transactions on Biomedical Engineering, vol. 62, no. 4, pp. 1203–1214, 2015.

[27] Boulehmi, Hela, Mahersia, Hela, Hamrouni, Kamel, Sana, Boussetta, Najla, Mnif. (2013). Breast cancer detection: A review on mammograms analysis techniques. 2013 10th International Multi-Conference on Systems, Signals and Devices, SSD 2013. 1-6. 10.1109/SSD.2013.6563999.

[28] U. Ojha and S. Goel, "A study on prediction of breast cancer recurrence using data mining techniques," 2017 7th International Conference on Cloud Computing, Data Science and Engineering - Confluence, Noida, 2017, pp. 527-530, doi: 10.1109/CONFLUENCE.2017.7943207.

[29] C. Shah and A. G. Jivani, "Comparison of data mining classification algorithms for breast cancer prediction," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-4, doi: 10.1109/ICCCNT.2013.6726477.

[30] A. Oliver, A. Torrent, X. Llado, M. Tortajada, L. Tortajada, M. Sentis, J. Freixenet, and R. Zwiggelaar, “Automatic microcalcification and cluster detection for digital and digitised mammograms,” Know.-Based Syst., vol. 28, p. 68–75, Apr. 2012.

[31] N. Dhungel, G. Carneiro and A. P. Bradley, "Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests," 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Adelaide, SA, 2015, pp. 1-8, doi: 10.1109/DICTA.2015.7371234.

[32] W. Lotter, G. Sorensen, and D. D. Cox, “A multi-scale CNN and curriculum learning strategy for mammogram classification,” CoRR, vol. abs/1707.06978, 2017.

[33] Y.D. Zhang, S.-H. Wang, G. Liu, and J. Yang, “Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional fourier transform,” Advances in Mechanical Engineering, vol. 8, no. 2, p. 1687814016634243, 2016.

[34] J. E. Ball and L. M. Bruce, "Digital Mammographic Computer Aided Diagnosis (CAD) using Adaptive Level Set Segmentation," 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, 2007, pp. 4973-4978, doi: 10.1109/IEMBS.2007.4353457.

[35] H. Lei, S. Liu, A. Elazab, and B. Lei, “Attention-guided multi-branch convolutional neuralnetwork for mitosis detection from histopathological images,”IEEE Journal of Biomedicaland Health Informatics, pp. 1–1, 2020.

[36] H. Yang, J. Y. Kim, H. Kim, and S. P. Adhikari, “Guided soft attention network forclassification of breast cancer histopathology images,”IEEE Transactions on Medical Imaging,vol. 39, no. 5, pp. 1306–1315, 2020.

[37] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, “Deep features forbreast cancer histopathological image classification,” in2017 IEEE International Conferenceon Systems, Man, and Cybernetics (SMC), 2017, pp. 1868–1873.

[38] Y. Wang, H. Dou, X. Hu, L. Zhu, X. Yang, M. Xu, J. Qin, P. A. Heng, T. Wang, andD. Ni, “Deep attentive features for prostate segmentation in 3d transrectal ultrasound,”IEEETransactions on Medical Imaging, vol. 38, no. 12, pp. 2768–2778, 2019.

[39] J. Li, J. Zhang, Q. Sun, H. Zhang, J. Dong, C. Che, and Q. Zhang, “Breast cancerhistopathological image classification based on deep second-order pooling network,” in2020International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–7.

[40] Y. Yari, T. V. Nguyen, and H. T. Nguyen, “Deep learning applied for histological diagnosisof breast cancer,”IEEE Access, vol. 8, pp. 162 432–162 448, 2020.

[41] J. Liu, B. Xu, C. Zheng, Y. Gong, J. Garibaldi, D. Soria, A. Green, I. O. Ellis, W. Zou, andG. Qiu, "An end-to-end deep learning histochemical scoring system for breast cancer tma,"IEEE Transactions on Medical Imaging, vol. 38, no. 2, pp. 617–628, 2019.

[42] V. Gupta and A. Bhavsar, "Sequential modeling of deep features for breast cancer-histopathological image classification," in2018 IEEE/CVF Conference on Computer Visionand Pattern Recognition Workshops (CVPRW), 2018, pp. 2335–23 357.

[43] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathologicalimage classification using convolutional neural networks," in2016 International JointConference on Neural Networks (IJCNN), 2016, pp. 2560–2567.

[44] N. Bayramoglu, J. Kannala, and J. Heikkil a, "Deep learning for magnification independentbreast cancer histopathology image classification," in2016 23rd International Conference onPattern Recognition (ICPR), 2016, pp. 2440–2445.