Image Segmentation of digital Mammograms to identify abnormalities for breast cancer using Deep learning

P. PADMANABAN

Thesis Report

AUGUST 2022

Dedication

I dedicate my research work to my family and friends. A special feeling of gratitude to my organisation Danaher who is supporting me on pursuing this course. My Site Lead Vijayaraghavan Chari, manager Amit Ratnawat and Sofi Mary for the constant support and encouragement for pushing my limit. A special feeling of gratitude to my wife, son, and daughter as they understand that I need to spend more time on this research and motivated me throughout this program.

Acknowledgement

I wish to thank Upgrad and LJMU for coming up with a program for working professionals to upskill themselves on niche skills. A special thanks to Dr.Ahmed kaky with classroom sessions and clarifying the doubts. I would like to acknowledge and thank my mentor Kausthubh sakhare for guiding me to conduct my research and providing any assistance requested. Special thanks to upgrad mentor nayana for resolving the queries and ensuring the smooth flow for this program. And a special thanks for the lovely batchmates for helping each other and passing on the postive thoughts to each other all through program.

Abstract

Breast cancer is the most commonly occurring cancer type followed by cervical cancer. The number of cases is increasing every year across the globe. The diagnosis of the cancer cells even at a smaller size and during pre-growing stage will make the treatment effective and saves the patient life. There are several challenges present today in diagnosis and hence an automated reliable system is needed. The existing research of segmentation is very limited and has scope improvement in segmentation accuracy along with other opportunities.

The publicly available CBIS-DDSM dataset is used for training the model. The dataset is more reliable as the suspicious region has been marked by trained mammographers. The image pre-processing is handled extensively for the model to learn the patterns appropriately for abnormalities. Model with U-net architecture having ResNet transfer learning for encoder and custom CNN for decoder is used to segment the mask on the mammogram images. The best model has performed with a result of 0.8179 of Dice coefficient and 0.6943 of IOU. The results were overfitting with the test data. Notably the model can segment the abnormalities at the exact place but fails to predict the shape accurately. The end model can be further enhanced with advanced architectures along with new images generated through GAN.

TABLE OF CONTENTS

DEDICATION ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
LIST OF TABLES vii
LIST OF FIGURES viii
LIST OF ABBREVIATIONS ix
CHAPTER 1: INTRODUCTION 1
1.1 Background of the study 1
1.2 Problem statement 3
1.3 Aim and objectives 4
1.4 Research questions 4
1.5 Scope of the study 4
1.6 Significance of the study 4
1.7 Structure of the study 5

CHAPTER 2: LITERATURE REVIEW 6
2.1 Introduction 6
2.2 CAD system for breast cancer detection 6
2.3 DBT and ultrasound 7
2.4 Breast density signifiance 9
2.5 Transfer learning and its challenges 10
2.6 Datasets 14
2.7 Discussion 20
2.8 Conclusion 21

CHAPTER 3: RESEARCH METHODOLOGY 22
3.1 Introduction 22
3.2 Image segmentation 22
3.3 Dataset 22
3.4 Data pre-processing 22
3.5 Model 24
3.6 Evaluation 26
3.7 Hyperparameters for training 26
3.8 Required resources 27
3.9 Research methodology 27
3.10 Summary 27

CHAPTER 4: IMPLEMENTATION 22
4.1 Introduction 22
4.2 Dataset 22
4.3 EDA 22
4.4 Pre-processing 22
4.5 Model building 24
4.6 Summary 26

CHAPTER 5: RESULT EVALUATION 22
5.1 Introduction 22
5.2 Experiments 22
5.3 Summary 22

CHAPTER 6: CONCLUSIONS AND RECOMMENDATION 22
6.1 Introduction 22
6.2 Discussion and conclusion 22
6.3 Contributions 22
6.4 Future work 22

REFERENCES 28
APPENDIX A: RESEARCH PROPOSAL 32
APPENDIX B: CODE FOR DATA PRE-PROCESSING AND MODEL BUILDING 33
APPENDIX C: DATASET 33

LIST OF TABLES

Table 2.1 Methodology .... **Error! Bookmark not defined.**
Table 2.2 Pros and cons .... **Error! Bookmark not defined.**

Table 2.3 Datasets and modality .... **Error! Bookmark not defined.**

Table 2.4 Hardware and experimental setup .... **Error! Bookmark not defined.**

Table 2.5 Performance matrix .... **Error! Bookmark not defined.**

Table 2.6 Datasets and Modality .... **Error! Bookmark not defined.**

Table 2.7 Challenges and future directions .... **Error! Bookmark not defined.**

Table 3.1 Software requirements.. **Error! Bookmark not defined.**

Table 3.2 Hardware requirements .... **Error! Bookmark not defined.**

Table 5.1 Performance on changing learning Rate .... **Error! Bookmark not defined.**

Table 5.2 Overall performance for VGG16.............................. **Error! Bookmark not defined.**

Table 5.3 Regularizer ResNet... **Error! Bookmark not defined.**

Table 5.4 Regularizer without batch normalization ResNet..... **Error! Bookmark not defined.**

Table 5.5 ResNet experiment observations summary .... **Error! Bookmark not defined.**

LIST OF FIGURES

Figure 1_1 Breast cancer in US, india and china 10
Figure 1_2 Breast cancer late detection.................... 10
Figure 1_3 Late detection of breast cancer.................... 12
Figure 2_1 CC and MLO views of a patient from DDSM dataset.................... 10
Figure 2_2 Mammogram on left and DBT on right side.................... 10
Figure 2_3 Ultrasound image 12
Figure 2_4 Image segmentation workflow.................... 10
Figure 2_5 Faster R-CNN.................... 10
Figure 3_1 Mass vs calcification in CBIS-DDSM 12
Figure 3_2 U-Net architecture.................... 10
Figure 3_3 IOU metrics image 10
Figure 3_4 Dice coefficient 10
Figure 4_1 Downloaded folder structure.................... 10
Figure 4_2 Files count summary 12
Figure 4_3 Mammogram scans with masks 10
Figure 4_4 Mass abnormalities single vs multiple 10
Figure 2_3 Ultrasound image 12
Figure 2_4 Image segmentation workflow.................... 10
Figure 2_5 Faster R-CNN.................... 10
Figure 3_1 Mass vs calcification in CBIS-DDSM 12
Figure 3_2 U-Net architecture.................... 10
Figure 3_3 IOU metrics image 10

Figure Research proposal 10

LIST OF ABBREVIATIONS

CAD………….. Computer-Aided detection

IOU…................. Intersection over union

DBT…………….. Digital Breast Tomosynthesis

MRI ……………. .Magnetic Resonance Imaging

DDSM………….. Digital database for screening Mammography

CBIS-DDSM…. .. Curated Breast Imaging Subset of DDSM

CNN…………….. Convolutional Neural Network

AUC…………….. Area Under curve

MIAS….................MiniMammographic Database

VGG……………..Visual Geometry Group

SVM…………… Support Vector Machine

DICOM….............. Digital Imaging and Communication in Medicine

ROI…………….. .Region of Interest

NBIA………….....National biomedical imaging archive

CIA……………… Cancer Imaging Archive

CLAHE………… Contrast Limited Adaptive Histogram Equalization

YOLO………….. You Only Look Once

ARF-NET……… An Adaptive Receptive Field Network

ResNet…………. The Residual Neural network

GAN…………… Generative Adversial Network

CHAPTER 1

Introduction

1.1 Background the study

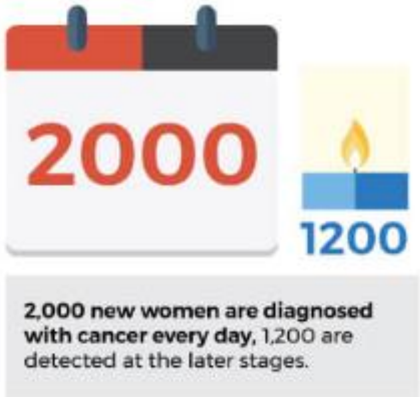
With constant lifestyle changes and the environmental exposures have started creating severe impact in mankind. Breast cancer has become most frequent tumour occurring in females with an increasing on number of cases every year. In 2020 USA alone has reported over 279000 cases with 15% death rate against other types of cancer(Cancer Facts & Figures 2020, n.d.). As there is no clear reason for the cause of breast cancer, the common suggestion by doctors across the globe is to make an early detection for the signs of abnormalities. There are clear evidence with the increase in survival rate when spotted early(Why is early diagnosis important?, n.d.).

1_1 Breast cancer in US, India and China (Breast cancer in India, n.d.)

The general process of diagnosis is to look for any abnormalities and if any suspicious found the patients are suggested for further tests. The process is currently "invasive, time-consuming and create unnecessary anxiety for patients if the results show that the tumour is benign. "(Yaker et al., 2021).

The histopathological images are collected by examining the biopsy or resection of the specimen. The most trusted method is examining these collected images by pathologist under microscope(Stathonikos et al., 2013). The advances in this field of digitalisation have led to increased number of histopathological images.

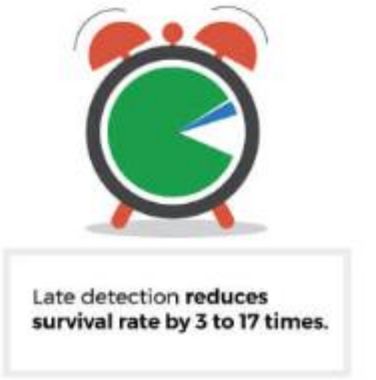
The amount of healthcare data generated for screening purpose has been increasing every year and the mammography is the key technique for breast cancer detection in identifying the anomalies. The results of the mammograms are currently evaluated by the Radiologist which has a chance of misdiagnosis(Yu et al., 2020) due to human error. The traditional algorithms have several short comings being unable to detect small masses and the accuracy is not as expected etc.



1_2 Breast cancer late detection (Breast cancer in India, n.d.)

For Computer aided design the deep learning is gaining momentum and has started creating impact on solving several healthcare problems specifically using computer vision technique. This study is focused on providing an additional assessment to the radiologist diagnosis and use it as a second opinion. To make Improvement in the identification of mass segmentation and reduce the false positive rate.

The medical image segmentation is very much needed in clinical decision, tumour growth and radiotherapy planning(Hoogi et al., 2017). And with mammography it is more sensitive in screening of breast cancer including the detection of high false positive rate. Still mammography is considered to create positive effect of reduction in breast cancer mortality.



1.3 Late detection of breast cancer(Breast cancer in India, n.d.)

There are number of studies/research performed in classification and segmentation of mammograms. They all target to solve different problems to classify presence of abnormalities, identify small mass, reduce false positive, quicker detection time. Various algorithms applied yielding to different results. In this research the primary focus will be to identify mass segmentation, reduce false positive and improve the state of art. This is primary aimed to make progress in research on mammograms with the early diagnosis which in turn lead to increase in life span of patients. The most difficult part of mammogram analysis is mass detection compared to micro-calcification due to its different shape and size(Oliver et al., 2010).

1.2 Problem statement

There is good amount of research has been initiated so far in understanding the medical images including mammogram, especially after the advancements in deep learning. There are significant amount applications targeted for image classification and segmentation techniques. The success of a deep learning algorithm in classification/segmentation task hugely relies on the extracted features from the available images for training. The successful model created for any computer vision challenge usually deals with millions of images for training, whereas for medical there are only few thousands of images available for medical domain.

The primary challenge is difficulty in gathering huge data with annotation as they are done manually by radiologist. And the community relies on limited availability of public data. On top of all these advancements it is still a challenge to detect mass at an initial stage by learning the object features. The visibility of the available mammograms is sometimes poor visibility and at times the low contrast posing a challenge for manual detection. Image enhancement is one of the techniques to overcome these misdiagnosis(Gardezi et al., 2019).

Segmentation in mammogram images is separating the anomaly region from the rest known as region of interest. The current state of art is the detections are with higher false positive rate. To reduce it there are several attempts made to be fully automatic CAD system which is at its initial stages. Another challenge is unavailability of good data with expert annotation. Data augmentation is one proposed solution to mimic similar not same images which sometimes lead to overfitting. The following literature survey will highlight the challenges in detect mass due to heterogeneous breast density.

1.3 Aim and objectives

The main aim of this research is to propose a model to diagnose the early detection of breast cancer from Mammogram images. The goal of this research is to improve the survival rate of the cancer patients through most effective and reliable methods.

- To investigate the mammogram images for presence of abnormalities in the form of mass segmentation
- To determine the optimum technique to reduce false positive in the mammogram images
- To propose a model which outputs a predicted mask from the mammogram if there are any abnormalities
- To evaluate the performance of the proposed semantic segmentation with improved IoU and Dice metrics.

1.4 Research question

How to detect mass abnormalities in Mammogram scans using image segmentation technique with an improvement in metric compared to the state of art.

1.5 Scope of the study

The mammogram images are analysed for segmenting the mass abnormalities in mammogram images. The images used for training and test are with mass abnormality presence and hence the classification of tumour is not in the scope. Also, the calcification abnormalities are not considered. This research focuses on preparing and evaluating model and the deployment part is not considered. The other abnormality such as calcification are also not considered

1.6 Significance of the study

Till date there are no concrete evidence for the occurrence of cancer cells(Sathyan et al., 2020a). Early diagnosis of breast cancer is the key factor in saving the life of the affected patients. Identifying the cells when they are smaller and reducing the false positive is also key challenge to Radiologist. "It is reported that most abnormalities missed by radiologists are related to cancerous masses"(procite, n.d.)

This study primarily aims to reduce false positive and improve the sensitivity from the existing work(Sathyan et al., 2020a). The U-Net architecture along with transfer learning will be used to identify the mass abnormalities in the mammograms. The improvement in metrics is encourages the radiologist to spend less time on more obvious case. The reduction in false positive will also reduce the panic among the patients and will reduce their subsequent visits.

1.7 Structure of the study

The structure of the thesis is as follows. Chapter 1 presents the background of the research in Mammogram, discuss the problem statement in 1.2. The aim and objectives are discussed in section 1.3. Section 1.4 presents the research question of this research aims to address. Followed by section 1.5 providing the scope of the study. Section 1.6 describes about the significance of the study.

Chapter 2 presents the necessary theoretical background and highlights the problem given in Chapter 1 by systematically reviewing the image segmentation done on mammogram images so far. Section 2.2 talks about the role of CAD system in diagnosis. Section 2.3 presents about alternative diagnosis method like DBT and ultrasound and their current state. Section 2.4 discuss the corelation of breast density with presence of tumour. Use of transfer learning in inevitable because of scarce data in medical domain which is discussed in detail in section 2.5. Section 2.6 details about the dataset used for the research. Section 2.7 concludes the research happened so far for breast detection. The summary of the chapter is given in section 2.7

Chapter 3 describes the procedure in selecting the data, pre-processing, and finalising the model based on the metrics. Section 3.1 introduces the research methodology used and 3.2 talks about the image segmentation technique . The dataset is described in detailed in section 3.3. The different pre-processing techniques for the images are explained in section 3.4. The section 3.5 explains the U-net model and the evaluation metrics is explained in section 3.6. The different hyperparameters used for tuning are explained in section 3.7 and the required resources are tabularised in section 3.8. The research methodology is presented in section 3.9 and summarised in 3.9.

Chapter 4 presents the actual implementation done for pre-processing and implementation of the model. Section 4.1 walks through the steps involved in the pre-processing pipeline and implementation of the model. The dataset is described with the structure of the dataset and steps done to pre-process ready in 4.2. Section 4.3 explores the data and presents the internal details of it. The pre-processing pipeline is described in section 4.4. The model building is explained in section 4.5 with the metrics used. Section 4.6 summarises the overall process and the end data.

Chapter 5 discussed the Results and evaluation techniques. Section 5.1 introduces the overall result achieved. The experiments performed with the model are detailed in section 5.2. The results of each model using VGG16 and ResNet are described in it. Section 5.3 summarises the achieved results.

Chapter 6 summarizes the conclusion and recommendations for future research. The interpretation of the result and contributions made are discussed in section 6.1. section 6.2 details the results achieved and the conclusion. The contributions made for the thesis are discussed in section 6.3. The future work is presented in section 6.4.

CHAPTER 2

Literature review

This section covers the progress of research work happened in the field of classification and image segmentation for mammograms. The CAD system is commonly used along with other technologies like Ultrasound and DBT. The scope of transfer learning in classification and image segmentation are discussed in detail.

2.1 Introduction:

The traditional CADx system has been in place for a decade and commonly used by Radiologist. The rapid development in the field of Deep learning has led paths for usage of x-ray picture of the breast region is also called as mammograms. There is a need to it as an additional diagnosis along with radiologist analysis and research has been initiated towards it. There also other alternatives such as Digital breast tomosynthesis and ultrasound has been tried so far. But as of today, the mammograms are the golden standard for breast cancer diagnosis.

2.2 CAD systems for breast cancer detection:

The systems Computer assisted diagnosis (CADx) and Computer assisted detection (CADe) are comprised together, commonly called as Computer-Assisted Diagnosis (CAD)(Dhull et al., 2018). These systems help in assisting doctors in the interpretation of medical images i.e., X-ray,MRI e.t.c.

These CAD systems is one of the reliable methods used in detection of breast cancer by screening the mammogram images. They generally help in classifying the tumour as malignant or benign (CAD-Wiki, n.d.) using a low dose X-Ray.

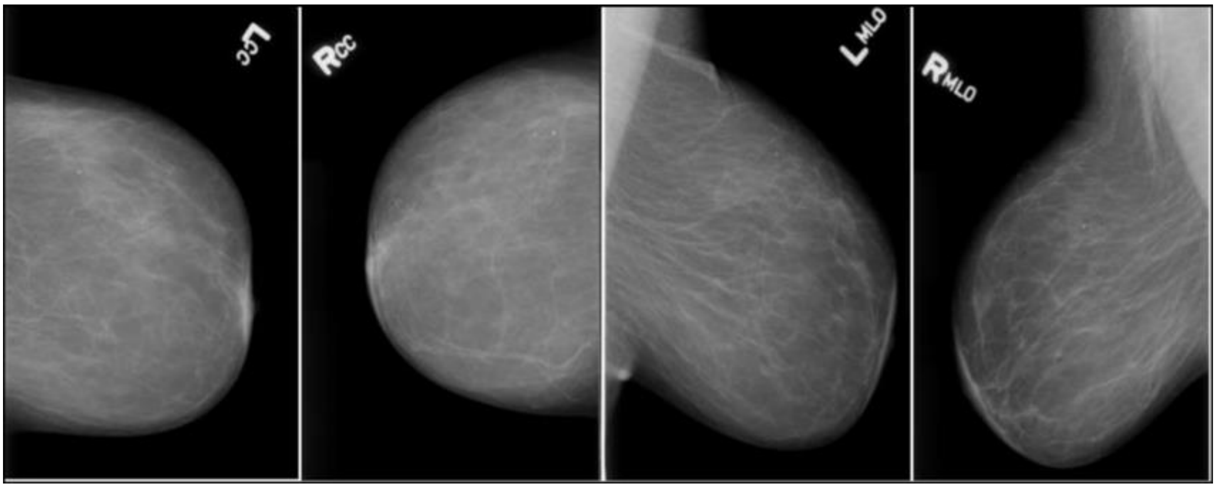
Though the CAD systems are used as a second opinion of what radiologist diagnose, they also help the medical specialist to overcome the shortcomings in detecting lesion as there are occasions where the image quality is distracted or the size of the tumour is small etc.(Giger et al., 2013)

Before the advancements in the field of Deep learning the important features of mammogram images were handpicked by the radiologist. This has a drawback of missing out the complex scenarios in the mammograms due to human error(Agarwal et al., 2019a). With Deep learning in picture the features important or necessary for diagnosis of tumour are detected by the systems with less human intervention.

The patients are screened at certain age interval or in case if they are experiencing any symptoms i.e., in the form of lump, nipple discharge, change in shape or size of the breast e.t.c. The x-ray images are taken in the mentioned area of concern and this process usually takes around 10 minutes.

The mammography images are 2-dimensional representation of 3 dimensional structures. There are generally of two views mentioned as on image 2_1:

The Craniocaudal (CC)
The Mediolateral oblique (MLO)



2_1 CC and MLO views of a patient from DDSM dataset

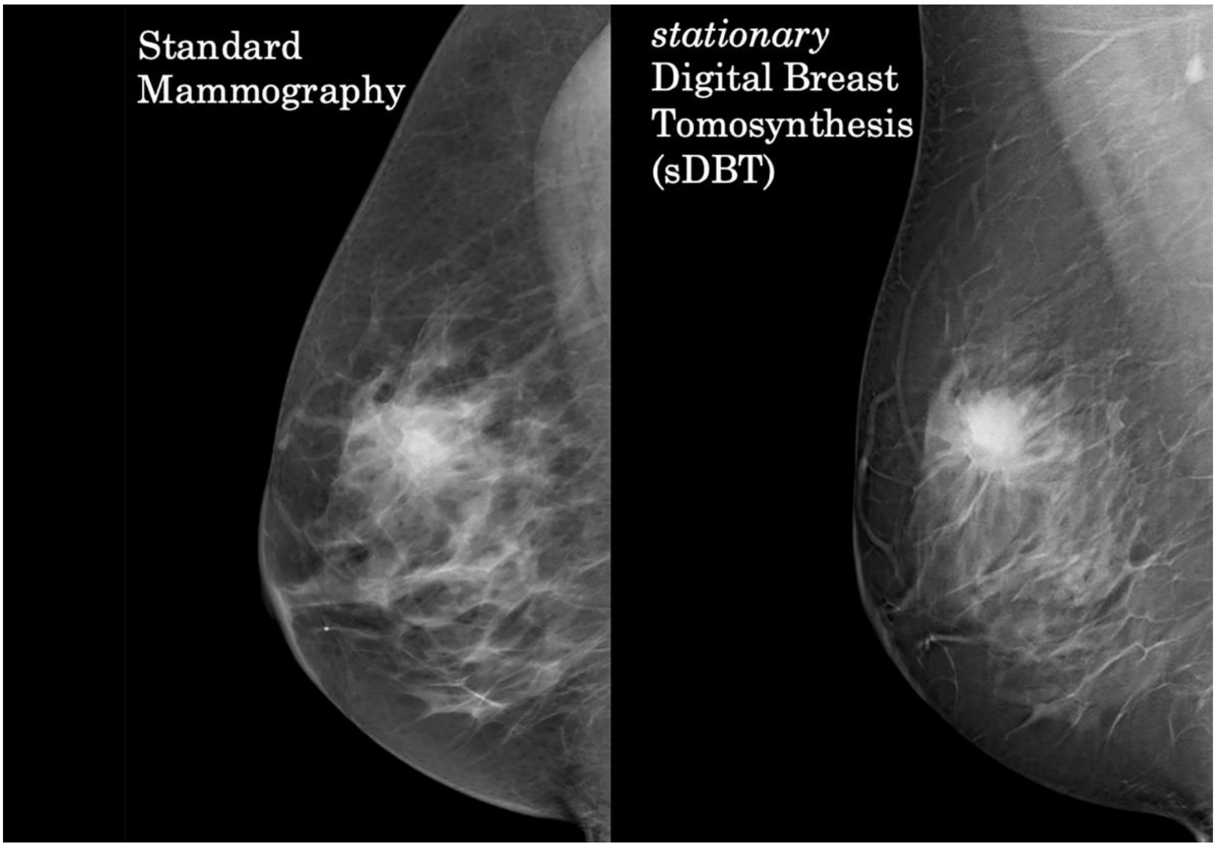
These are the images widely used for identification and diagnosis of breast cancer(Salama and Aly, 2021). These 2 views of different positions are taken to ensure the maximum number of tissues are included. At some cases additional images are also taken for more information but the care is also taken to limit the radiation exposure. This restriction is specially applied for people of age less than 40 on an average but based on local regulation the age may differ.

The presence of mass and calcification in X-ray images through examination indicates the presence of breast cancer(Sathyan et al., 2020b). The auto detection of these mass and calcification in X-ray images using CAD will be a supportive system to the radiologist. “Artifacts reduce the quality of mammograms and may mimic or obscure abnormalities and cause interpretation errors” (Geiser et al., 2011)for radiologist. There are also other factors affecting the interpretation as the newly trained radiologist who lacks experience, or the lack of skilled , experienced radiologist forcing the existing workforce with amount of workload beyond their capacity. All these factors may affect the evaluation of the mammograms(Ellenbogen et al., 2009).

2.3 DBT and ultrasound:

A convolutional neural network (CNN) is a type of artificial neural network used in image segmentation and processing that is specifically designed to process pixel data. It contains of mainly 2 main layers convolutional and pooling. The CNN receives the input with weights and biases which is then used to calculate the output of the neuron. The pooling layer reduces the size of the data by subsampling the output given by CNN.

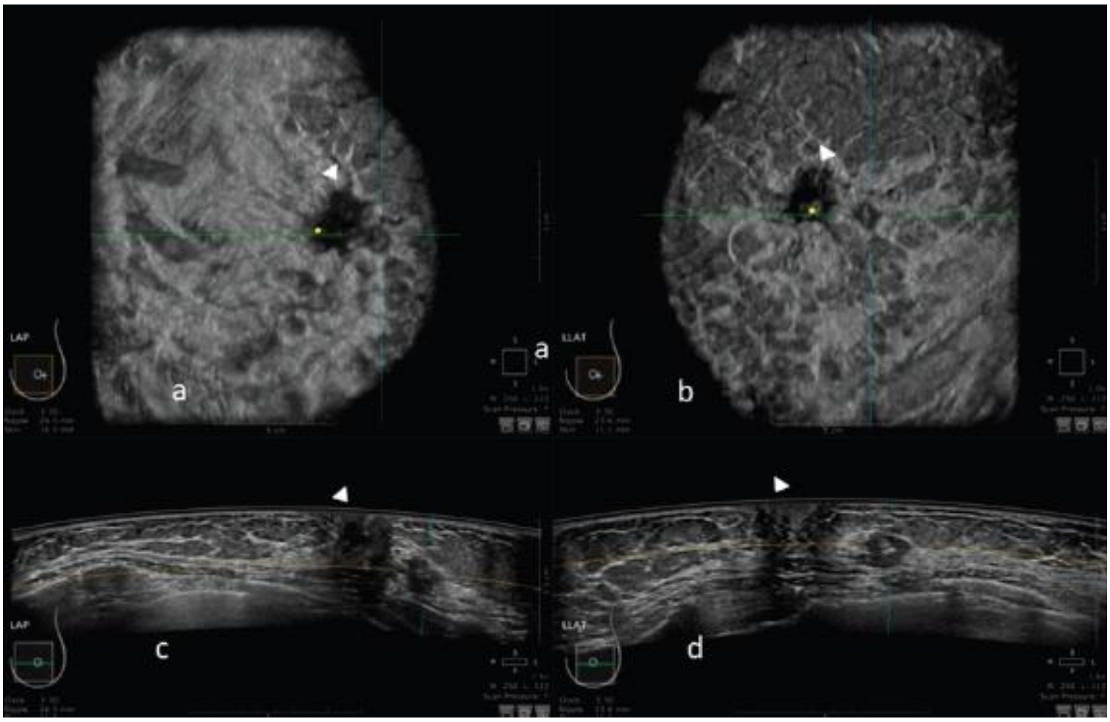
The recent advancement in the field of digital imaging has paved way for the technology Digital breast tomosynthesis(DBT) imaging system(Gur et al., 2009). This is introduced to reduce the recall rate in the diagnosis through mammogram images. The DBT as shown in figure mammogram on left and DBT on right can capture the series of images with different layers and provides better morphological analysis of mass with smaller background noise. Comparatively the mammogram images are a smaller fraction of size. The separation of layers in DBT provides an advantage in improving the reduction of false positive rate compared to mammogram images.



2_2 Mammogram on left and DBT on right side(Stationary digital breast tomosynthesis increases diagnostic accuracy, n.d.)

Along with reduction of recall rate there has significant improvement in cancer detection rate by 1-3% for every 100 screened women(Hooley et al., 2016). This is a more fast, practical, and simple solution for diagnosis over traditional mammogram examination. With all the advantages having said the medical industry still relies on mammogram images for cancer detection due to industry wide adaption and the low-cost factor(Agarwal et al., 2019a).

There are certain cases where mammogram is not feasible or affordable (Sood et al., 2019). Apart from x-ray, the breast ultrasound as shown in the ultrasound image is the most used diagnostic imaging technique using high frequency sound waves to produce images of internal body structure. This is also referred as Sonography (Giger et al., 2013). In case of biopsy, sonography is the preferred tool on guiding the follow up procedures. This is also used on places where mammography is unavailable in low-cost regions.



2_3 Ultrasound image(Breast ultrasound, n.d.)

2.4 Breast density significance:

Breast density is also considered for evaluation as it poses a risk factor for breast cancer.

The ratio of fibro glandular tissue in the breast to the amount of fatty tissue is called as Breast density(Giger et al., 2013). The amount of breast density evaluation is quiet challenging in Mammogram as they are 2 dimensional. Hence there is also a shifted interest towards new procedures.

Becker Anton et al trains a model on breast density and age matched control cohort to train a neural network on the Breast Cancer Digital Repository data set(Becker et al., 2017). The radiologists have verified the result Az of 0.79 on the testing set.

2.1 Methodology

Year	Reference	Pre-Processing Methods	Modelling-Techniques	Evaluation Methods
2021	(Salama and Aly, 2021)	Data augmentation	Inceptionv3 and U-Net	Accuracy, AUC, sensitivity, precision and F1 score
2019	(Sood et al., 2019)	Nil	Bivariate random effects model	Sensitivity specificity and diagnostic odds
2019	(Agarwal et al., 2019b)	Global thresholding to segment breast region and right breast are mirrored horizontally	InceptionV3with CBIS-DDSM dataset is selected	IOU with TPR
2019	(Tsochatzidis et al., 2019)	Image resizing as per the CNN model	AlexNet, VGG, GoogleNet	AUC
2015	(Dhungel et al., 2015)	Reduce image size to 40*40	Deep learning, Random forest	True positive rate and false positives per image
2017	(Lotter et al., 2017)	Train for abnormality detection	Multi scale CNN and curriculum learning strategy	AUROC – Area under the receiver operating characteristic curve
2018	(Ribli et al., 2018)	Vertical and horizontal flipping. JPG	Faster R-CNN	AUC

		images to PNG format		
2017	(Kooi et al., 2017)	Min-max scaling for patches. Interpolation. Pad images with zero	CNN and SVM	AUC
2018	(Chougrad et al., 2018)	Intensity normalisation. Resizing of the images	Inception V3	AUC

Preparing a model to train with many mammogram images is close to impossible especially in medical domain. When trained on smaller datasets they tend to overfit to the training data(Dhungel et al., 2015). The current solution to this common problem in ML world is by using transfer learning to train the data i.e., to use a pretrained model and fine tune them finally with the domain specific smaller dataset. There is also difference of opinion on this approach as Azizpour et al suggests that there should be similarity between the pretrained database with the test data(Azizpour et al., 2015). There is also evidence from research work by Tajbakhsh et al and many others on non-domain specific pretrained models when fine-tuned which has yielded better results.

Olaf et al has also contributed on breaking the misconception of the need of domain specific data in training especially in biomedical image segmentation as he has used U-Net and has achieved promising results. With the limited availability of domain specific data, he has proposed a network and training strategy using data augmentation to use them wisely and outperform the prior best methods(Ronneberger et al., 2015).

2.5 Transfer learning and its challenges:

Chougrad et al proposes an end-to-end deep learning convolutional neural network using transfer learning with pre trained weights(Chougrad et al., 2018) to classify mass lesions in mammography. The results of accuracy and AUC are 97.35% and 0.98 respectively and is a good score while using a smaller dataset like DDSM. To evaluate the system, the model is tested against MIAS and have got 98.23% accuracy with .99 AUC score. Similar studies have also pointed the need of using Transfer learning to use the trained model on different domains.

2.2 Pros and Cons

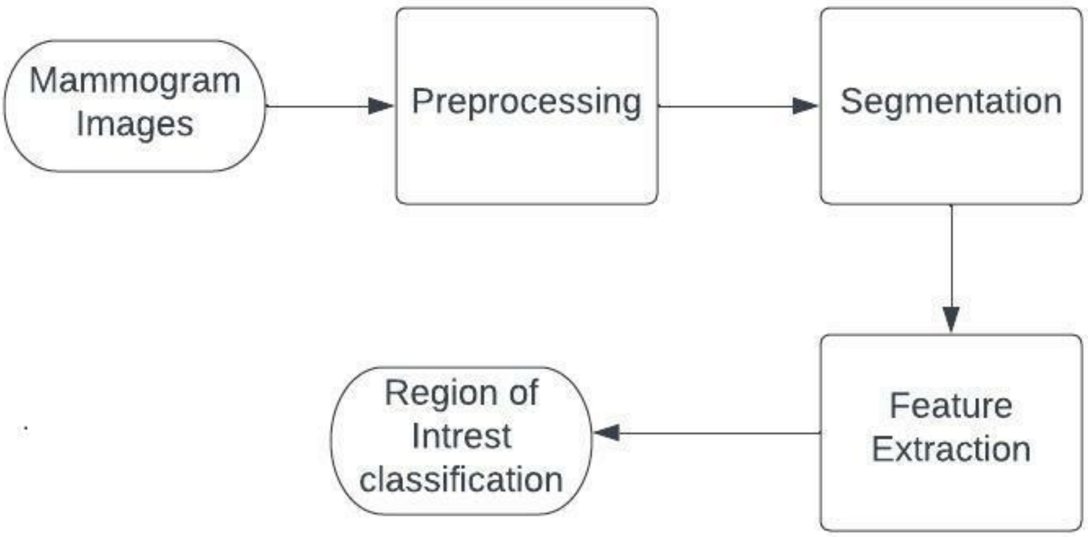
Year	Reference	Pros	Cons
2021	(Salama and Aly, 2021)	Proves no need of human intervention with pre or post processing or handcrafted features	
2019	(Sood et al., 2019)	Effective on small, invasive, node negative cancer in dense breast tissue. Improved sensitivity	Limited evidence on effectiveness of ultrasound as early detection tool.
2019	(Agarwal et al., 2019b)	Trained on 3 model and picked the best i.e., InceptionV3	Instead of classification of central pixel, the number of pixels within each patch needs to be used.
2019	(Tsochatzidis et al., 2019)	Explored usability of Transfer learning	Need for gathering large medical data is still there
2015	(Dhungel et al., 2015)	True positive rate is above 0.9	Dataset used is small
2017	(Lotter et al., 2017)	Effective result as AUROC is 0.92	Hand-drawn segmentation masks are used
2018	(Ribli et al., 2018)	AUC 0.95% score with only 0.3% FPM per image	Small dataset has been used
2017	(Kooi et al., 2017)	Uniqueness in the research to compare CNN with Radiologist	More false positive specially the benign abnormalities

2018	(Chougrad et al., 2018)	Shown effective use of transfer learning with good result and multiple databases	Deeper architecture were not used

The detection of mass in mammograms are considered to challenging as the abnormalities vary largely in texture, size, shape, boundary e.t.c(Dhungel et al., 2015) . Dhungel et al has proposed a solution to segment using Deep learning and the marked abnormalities are further processed with Random Forest for classification which has yielded a better test positive rate. As a note the test data is unseen and has produced visually accurate detection result without any false positives.

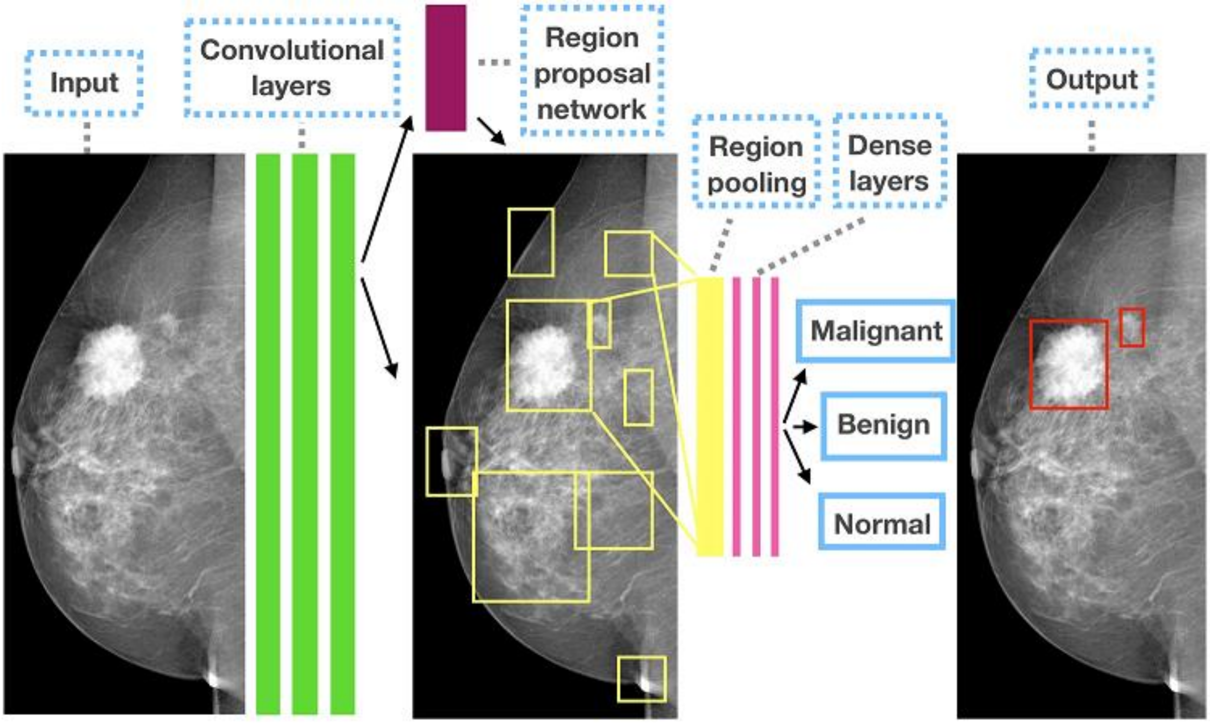
Even before segmentation, the classification of whether cancer is present or not is more important. And creating an effective model is more challenging and is termed as "needle in a hay stick" nature of the problem(Lotter et al., 2017). the downsizing of images has been successful in many datasets but in case of mammograms the small features are missed because of it. Hence the classification pipeline consists of several steps such as feature extraction, segmentation, and classification etc.

Lotter et al proposes classification proposes a multiscale CNN scanning with lesion-specific curriculum learning strategy. They are first trained for abnormality detection and then image level training. The choice of architecture and training scheme has led to a promising result compared to the transfer learning CNN models(Lotter et al., 2017)



2_4 Image segmentation workflow

Dezso Ribli et al proposes using Faster R-CNN as an object detection framework. “Faster R-CNN is based on a convolutional neural network with additional components for detecting, localising and classifying objects in an image”(Ribli et al., 2018). It is primarily trained using Region proposal network to detect and localise objects on the images using detection boxes. This model has achieved a highest AUC with 95 percentiles using INbreast dataset. This has led them to win second place on DREAM challenge with only 0.3 false positive marks per image.



2_5 Faster R-CNN (Ribli et al., 2018)

There were also attempts made on determining the importance of features such as location, patient information by T. Kooi et al. A head-to-head comparison is made between CNN and CAD system to show case the manually designed features are outperformed by CNN network(Kooi et al., 2017). To conclude, a final study is also performed to compare between a CNN network and radiologist predictions with an outcome the CNN learns from data without need of radiologist making development easier and faster(Kooi et al., 2017).

Richa Agarwal et al has proposed a model for mass detection in mammograms using deep convolutional neural networks with promising results(Agarwal et al., 2019b). The performance of the popular CNN architectures i.e., VGG16, ResNet50, InceptionV3 are compared and out of which InceptionV3 has outperformed in mass and non-mass classification on CBIS-DDSM dataset. The testing is done with INbreast database which is a public dataset. The results proves that the transfer learning models and fine-tuned on mammogram images (CBIS-DDSM) are performing good in classifying masses in mammograms.

There were experiments done on using models with pre trained weights and another with random weights. The prior approach performed better with best result as per Tsochatzidis et al(Tsochatzidis et al., 2019) . The images were trained with AlexNet, VGG and GoogleNet. The datasets were evaluated on DDSM-400 and CBIS-DDSM. The conclusion stresses the pain point of training from scratch as it is more time consuming. With scarcity of needed medical data for training, the Transfer learning approach proves to be a suitable option for mammogram studies.

A new automated CNN based framework for image segmentation is proposed by salma et al(Salama and Aly, 2021). InceptionV3, DenseNet121, ResNet50, VGG16 and MobileNetV2 are used to classify benign and malignant. U-Net is used for segmenting the breast region. With data augmentation applied to improve the data quantity, The U-Net architecture with Inception V3 transfer learning, it has achieved the best score with a computational time of 1.2134

2.3 Hardware and Experimental setup

Year	Reference	Technique	Hardware	RAM	Applications
2015	(Dhungel et al., 2015)		Intel(R) Core(TM) i5-2500k 3.30GHz CPU	8GB	
2019	(Agarwal et al., 2019b)	Adam optimizer with batch size 128	NVIDIA TitanX Pascal GPU	12 GB RAM	Keras-2 with Tensorflow

Data augmentation is one of the proposed approach when domain specific data is not available. Sebastien et al has investigated the benefit of data augmentation. The stage at which augmentation needs to be done plays a key role, if it is at feature generation or at classification stage i.e., dataspace or feature space(Wong et al., 2016). The solution provided is more robust than the DBSMOTE as it leads to overfitting when the newly created samples are close to existing cluster centre. The preferred approach is to perform data augmentation in dataspace if label preserving transforms are available.

Annop et al has proposed a solution (Sathyan et al., 2020a) using U-Net segmentation architecture to segment the boundaries of mass and the calcifications present in the mammograms. The architectures are modified to segment the mass and calcification with its own dataset The results are promising as most of the abnormal regions are correctly identified with fewer false positive and false negatives.

For radiologist sometime the small-scale masses are difficult to identify, and the miss could leads to serious consequences to the patients. Hui Yu et al has proposed a solution(Yu et al., 2020) based on model Dense Mask R-CNN which is suitable for capturing the low-level features which in turn helps in detection of the small-scale breast masses. The results have shown an improvement of the original Mask R-CNN with a shortcoming of redundant information generated.

2.4 Performance Metrics

Year	Reference	Sensitivity	Specificity	AUROC	AUC	Accuracy
2021	(Salama and Aly, 2021)	98.98	98.79		98.88	98.87
2019	(Sood et al., 2019)	80.1%	88.4%			
2019	(Agarwal et al., 2019b)	93%				84.16
2015	(Dhungel et al., 2015)	96%	80%			
2017	(Lotter et al., 2017)			0.92		
2018	(Ribli et al., 2018)				0.95	

2017	(Kooi et al., 2017)				.911	
2018	(Chougrad et al., 2018)				.99	

The unbalance class is an obstacle in achieving a good segmentation result. Juan chen et al has performed experiments(Chen et al., 2020) in solving this using multi-scale adversarial network for breast mass segmentation. Again, the U-Net has been the used for the segmentation part achieving satisfactory results. The results comes with a cost of additional training time.

Though High false positive rates are not a concern as the mammograms are again evaluated by the radiologist, Lugman Ahmed et al has proposed a solution focusing on the pre-processing(Ahmed et al., 2020) with the complete process to extract the patten. The helps to remove the noise in the data and improves the performances.

Ilhame Ait Lbachir et al has proposed a fully integrated CAD systems to mass detect and classify the mammograms(Lbachir et al., 2021). The images are pre-processed, segmented based on HRAK algorithm, false positives are reduced and finally classification is done through SVM. Though this has achieved great results there is a need for improvement on differentiating the spiculated masses with normal tissues.

Jihahui Zhao et al also proposes a solution of both segmentation and classification using YOLOv3 and Transfer learning(Zhao et al., 2021). There are 3 models i.e., general, mass and microcalcification are created and they all perform good at detecting the lesions. The only caveat is that it performs poor on low configured laptop.

Amine Gonca Toprak et al has worked on segmenting the CNN for detecting cancerous lessions in mammograms(Toprak et al., 2022). The publicly available CBIS-DDSM dataset is preprocessed and segmentation is performed using U-net and VGG16. The optimal threshold value is set as 115 and applied to model outputs to provide more accurate segmentation. The IoU and DICE metrics are 0.3345 and 0.4316 respectively. This research proposes having more complex model to improve on the existing metrics.

2.5 Problem and Purpose

Year	Reference	Problem	Purpose

2021	(Salama and Aly, 2021)	Aid radiologist in assisting for early detection and improve the efficiency of the system.	New and different strategy to segment and classify mammogram, explore powerful model
2019	(Sood et al., 2019)	Mammography is not always feasible, and it is expensive for low income countries	Analyze is ultrasound a viable breast cancer detection modality
2019	(Agarwal et al., 2019b)	Diagnose breast cancer at an early stage	Propose patch based CNN method for automated mass detection in full-field digital mammogram
2019	(Tsochatzidis et al., 2019)	Propose model for breast cancer diagnosis from mammograms with mass lesions	Compared using training model and random weigh assignment
2015	(Dhungel et al., 2015)	Mass detection is challenging for varied reasons of shape, size e.t.c	Improved Test positive rate
2017	(Lotter et al., 2017)	Tool to screen mammogram and use it as an early detection of breat cancer	Solve image classification for presence of cancer or not
2018	(Ribli et al., 2018)	Improvement of current CAD technologies	Propose a new CAD system for object detection framework
2017	(Kooi et al., 2017)	Head to head comparison of CNN and radiologist are done	To understand how reliable CNN networks are in diagnosis
2018	(Chougrad et al., 2018)	The datasets are smaller for training especially for medical domain. Use transfer learning to evaluate to predict if the mass lesions are benign or malignant	Evaluate the use of transfer learning for training the model and validate its effectiveness

Dina Abdelhafiz et al proposes U-Net model to segment Mass lesions in mammogram images by extracting both low level and high-level features(Abdelhafiz et al., 2020). This vanilla U-Net model has superior performance compared to existing models such as SegNet, Fast R-CNN etc. but with scope for improvement in the segmentation.

Mass detection and segmentation in digital mammograms plays a crucial role in early breast cancer detection and treatment. Yongye Su proposes a double shot model for mass detection and segmentation using YOLO and LOGO architectures(Su et al., 2022). It achieves true positive rate of 95.7%, mean average precision 65%, IOU at 64% on CBIS-DDSM dataset. The same trend has been observed in another public dataset INBreast also.

Identifying mammograms with small mass has always been a challenge because of the size of the receptive fields. Chunbo Xu proposes an adaptive receptive field network for breast mass segmentation in whole mammograms and ultrasound images(Xu et al., 2022). Thr ARF-Net is able to achieve the dice index of 85.75%,86.1% and 88.12% on CBIS-DDSM, INbreast and UDIAT benchmarks.

To improve the detection accuracy of breast masses Jianchen An et al proposes a target detection model D-Mask R-CNN network based on Mask R-CNN for breast mass detection(An et al., 2022). The improved model for detecting breast masses has reached 0.66 in the test set which is higher than that of the original mask R-CNN.

2.6 Dataset:

The Digital database for screening Mammography (DDSM) released in 1997 is one of the most important datasets among the research community for studying mammogram(Rose et al., 2006). It had few short comings such as Actual mammogram were converted to JPEG using non-standard lossless variant and the meta data is not readily usable format. The Curated Breast Imaging subset of DDSM (CBIS-DDSM) is an updated version of DDSM containing 2620 scanned film mammography studies(CAD-Wiki, n.d.).

2.6 Datasets and Modality

Year	Reference	Datasets
2021	(Salama and Aly, 2021)	MIAS,DDSM, CBIS-DDSM

2019	(Sood et al., 2019)	LMIC(Pivate data)
2019	(Agarwal et al., 2019b)	CBIS-DDSM, INbreast
2019	(Tsochatzidis et al., 2019)	DDSM-400 and CBIS-DDSM
2015	(Dhungel et al., 2015)	DDSM-BCRP and INbreast
2017	(Lotter et al., 2017)	DDSM
2018	(Ribli et al., 2018)	DDSM and INBreast
2017	(Kooi et al., 2017)	DDSM
2018	(Chougrad et al., 2018)	MIAS, DDSM, INbreast and BCDR

This has normal, benign, and malignant cases in decompressed DICOM format. The region of interest segmentation and the bounding boxes are curated and verified by certified pathologist. Hence the data is more accurate. The dataset is made public with good intention to continue the research for the mammogram community.

The mass which was not clearly seen are removed from the dataset. A lossless conversion is made from DDSM preserving all the information for CBIS-DDSM. The original DDSM files were distributed with set of tools which were difficult to refactor for use on modern systems. They were reimplemented in python to be truly cross-platform for today's users. A lesion segmentation algorithm was applied to supply much more accurate ROI for mass segmentation. The split of the dataset is done with 80-20% for mass and calcification separately.

2.7 Challenges and future directions

Year	Reference	Future scope

2021	(Salama and Aly, 2021)	Lack of tagged data availability
2019	(Sood et al., 2019)	Need investigation on the full potential of breast ultrasound
2017	(Lotter et al., 2017)	Experiment with unsuperviced methods to reduce on hand drawn segmentation
2018	(Ribli et al., 2018)	Low contrast INBreast images are used. Mamogram images are used as an input without any annotation
2017	(Kooi et al., 2017)	Reduce the false positive in the future.
2018	(Chougrad et al., 2018)	Use deeper architectures for training and use more challenging images

2.7 Discussion:

There are several reasons like need for early diagnosis, reducing the workload of a radiologist, identify when the abnormalities are of small e.t.c leads to a need for an automated diagnostic systems to interpret the images. There has been several research already involved in bringing change and the results are promising compared to few years and thanks to the development happening in the field of deep learning.

The several challenges need to overcome in making the automated system reliable. There should be in creating tagged data to make the model learn better. Many other architectures need to be tried on these available existing datasets to see if the false positive can be reduced.

Still there are open questions like reducing the false positive, achieving a good metrics in unseen data. There were also other options identified such as Ultrasound DBT if they can fill the gap what mammogram has. With community involvement and number of efforts happening we can expect a positive outcome in the coming years.

2.8 Conclusion:

Though there are several studies performed for classification , the research on segmentation is very limited and there is not a good metric achieved so far. There are also open questions on improving the segmentation performance, reducing time on running the models and inconsistent results on unseen data (Wang et al., 2020) . There are studies has shown for the lack of domain specific data as reason for not achieving good metrics on unseen data. When this trend changes, the auto mammogram examination can move as first screening task before the radiologist has a look on it.

CHAPTER 3

Research Methodology

3.1 Introduction:

In this chapter the datasets are explored, pre-processing steps before feeding to the model, deep learning architectures and finally the different metric to measure the mass detection are discussed in detail.

3.2 Image segmentation:

Image segmentation is a subdivision of computer vision that answers

a) What is in the image

b) Where is in the image

This is more a classification task which is done at pixel level to determine the class. The instance segmentation is a method to further identify the different instance of same class. For mammograms the focus is on image segmentation as instance segmentation not of interest. Each pixel in the image is classified if it is a tumour (Mass abnormality). This is more of one-hot encoding the class labels i.e., marking the pixels with 1 if it is tumour or else with zero

To detect the severity of mass abnormalities the radiologist investigates image for the regions containing the cancer cells. With neural networks we can use semantic segmentation to extract the suspicious region using segmentation technique. The primary purpose is to segregate the image into pixels, classify the pixels into unique categories and cluster the similar coherent pixels(Image segmentation, n.d.).

The scanned mammograms are in 2D format along with ground truth labels as binary masks. A model capable of extracting region out of the scans are created and the mammograms are used a s input. The model predicts the output which is compared with the ground truth.

3.3 Dataset:

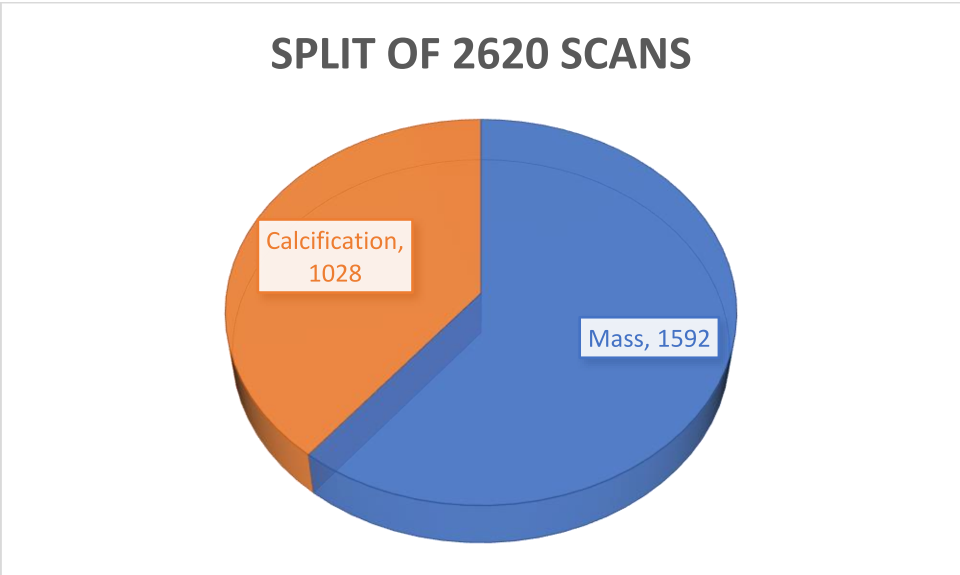
The dataset used for this task is CBIS-DDSM. It contains 2620 real world scanned mammogram images(CBIS-DDSM, n.d.) in Dicom format. DICOM is the common file format for storing and transferring medical images across hospitals. There is information such as PatientID, instance and few other metadata are embedded in the DCIM image. The region of interest annotation for abnormalities are selected, curated, and verified by trained mammographers. The mammograms are provided with their respective binary masks representing the abnormalities.

There are 2 scans MLO and CC are done for each patient along with region of interest. For some patients the mass abnormalities are more than 1. The data is split into train and test already to avoid class imbalance in them.

The dataset is available for download from wiki page of cancer imaging archive(CBIS-DDSM, n.d.). There is NBIA Data retriever available to download the imager, once installed the images can be downloaded. The images are available individual with mass and calcification separately.

The Folder name is similar to this e.g., Mass-Training_P_00911_LEFT_CC_1.

1. Mass represents the type if it is Mass or calcification abnormality
2. Training represents train or test data
3. P_00911 is the patient ID
4. LEFT represents if the scan has taken from left side or right side
5. CC represents if the orientation is CC or MLO
6. 1 is the first abnormality (There are cases where multiple abnormality for single scan is present)



3_1 Mass vs Calcification in CBIS-DDSM

The patient IDs re redundant as it is around 6671 the actual participants are 1566.

3.4 Data pre-processing:

The scanned images used in the dataset are not of high quality and they are with several kind of distortions meant to be corrected. The aim of this pre-processing step is to make the mammogram images ready for the models to be learned. There are several steps in the preprocessing which are followed to make it more effective, and the effectiveness of each depends on the combination of pre-processing step can be measured with the metric.

The mammogram images are not of same height and width, they need to be resized as almost all the images used to for pretrained model need to be of same size. There are white borders in few scans which might give an impression of tumor presence for the model. The images are taken in left and right direction, switching them to a single side will make the model to learn better.

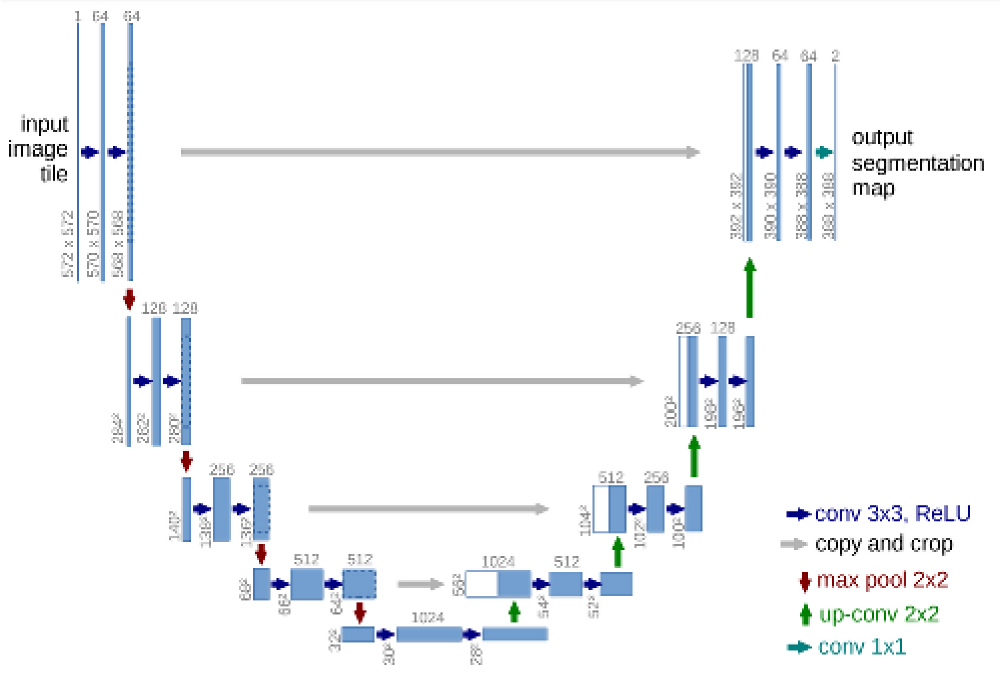
There are also markings present in the image which needs to be removed as there is possibility of model getting confused with the abnormality. The clear separation of breast with background is missing which needs to be fixed. And some images are observed to have poor contrast which fails to differentiate the breast tissues with the mass abnormalities.

The images will be having varied pixel values from 0-225 and this needs to be normalized for the model to learn. The images are rotated to one side for the ease of model to learn. The borders are cropped and resized to fit the model

The mammogram images contain white strips which creates a confusion among the actual cancer cell during segmentation of the cancer cells (Ahmed et al., 2020). The borders are cropped and normalization is performed. The multiple abnormalities need to be clubbed together as a single output image to feed into the model. There are only 1592 contains mass abnormalities out of which 71 contains more than 1 abnormality. Hence data augmentation applied to increase the variability with the existing images.

3.5 Model:

U-Net is a unique popular CNN architecture created for Biomedical image segmentation tasks created by olaf Ronneberger et al. in 2015. It has an end-to-end encoder-decoder network for semantic segmentation with unique Up-Down architecture with Contracting path and expansive path. It takes an input image and outputs a segmented image



3_2 U-Net architecture (Ronneberger et al., 2015)

The image is passed into the first half of the architecture, the encoder block with the pretrained VGG16 and ResNet model used as pretrained model. The convolution layer is stacked with maxpool layers at different levels. The up sampling can be achieved by several un pooling operations like transpose convolution. They learn the small, detailed features in the beginning and the images are converged. The classification of layers at the final step is excluded and the values are passed on to decoder block.

The decoder layer is the second half of the architecture containing up sampled images to increase the image size with regular convolution operations. The features learned by the encoder are passed onto the pixel space to get a dense classification. This is done to restore the image to the original size of the input image.

The **U-Net architecture** which "consists of a contracting path to capture context and a **symmetric** expanding path that enables precise localization."(Ronneberger et al., 2015). There are also other advance models evaluated over time with skip connection on FCN(Drozdzal et al., 2016).

Transfer learning can be applied to utilize the power of exiting models. Various experiments are performed to identify the hyperparameters such as no of epochs, learning rate, optimizer, Loss function e.t.c. Jegou et al. proposed the use of dense blocks stating, "the characteristics of DenseNets make them a very good fit for semantic segmentation as they naturally induce skip connections and multi-scale supervision."(Jegou et al., 2017)

The encoder part can be implemented using transfer learning by VGG16 or Resnet excluding top layer. The u-net architecture can be used to build the decider layer. The image augmentation can be performed by TensorFlow framework to improve the learning.

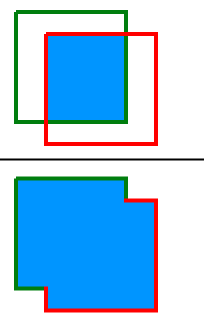
VGG16 is also known as ConvNet, consist of input, output, and various hidden layers. This is one of the popular transfer learning models for image classification learned with thousands of different categories with 92.7% accuracy. The downside is it takes lot of disk space, bandwidth and slow in training.

ResNet developed in 2015 introduced a novel residual architecture with skip connections and deeper networks than its previous popular networks. One downside of too many layers is that doing so makes the network more prone to overfit training data. The use of dropout , L2 regularization and batch normalisation is to avoid overfitting.

3.6 Evaluation:

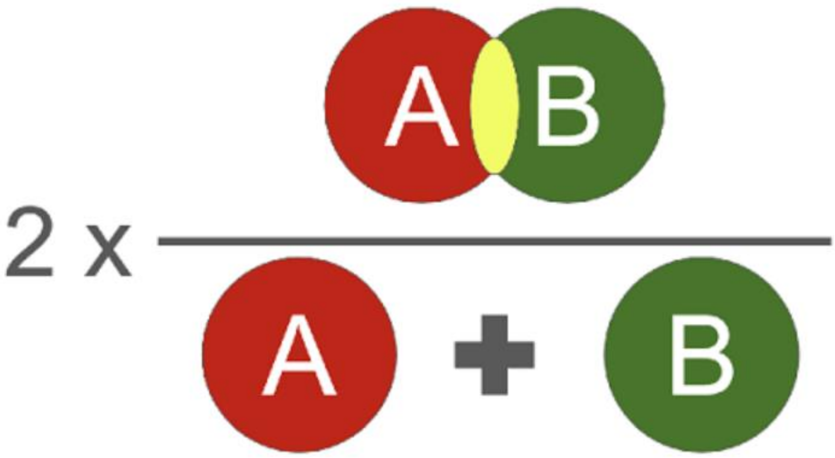
The results of a data augmentation are generally evaluated through IOU (Intersection over union) metrics. The metrics is more about evaluating the overlap of predicted with the true boundary, a measure of overlap. Higher the IoU has better result (preferably more than 0.5) vs lower IoU indicates poor results. In the numerator we compute the area of overlap and in denominator we compute the area of union

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



3_3 IOU metrics image (Padilla et al., 2020)

Dice coefficient is another metric for sematic segmentation. The intersection of actual and predicted pixels is counted and divided by sum of pixels from individual images. The result is multiplied by 2.



3_4 Dice coefficient (Dice and Dice loss, n.d.)

Pixel accuracy is another metric to measure correctly identified pixels. This is not of much use as the importance of prediction of image segmentation is on the masked area and hence it is not relevant. Binary crossentropy loss function is used as this is an identification of mask and non-mask areas.

3.7 Hyperparameters for training:

The model is experimented using below hyperparameter

- Optimiser
- Batch size
- Learning rate
- Dropout rate
- No. of epochs
- Loss function

Various combinations of these are used and the one producing the best IOU, Dice metrics in test data will be saved.

3.8 Required resources:

3.1 Software requirements

No	Software
1	NBIA Data Retriever
2	Python

3	AWS Sage maker
4	S2
5	VS code
6	Google colab

3.2 Hardware requirements

No	Hardware
1	Azure ML Studio
2	NVIDIA Tesla K80

3.9 Research methodology:

There will be good number of pre-processing techniques like resizing, normalisation will be done. Then the transfer learning can be applied to exclude the last layer and extract features. Probably VGG16, ResNet50 and few other models will be used for experimentation. As this is an image segmentation problem the U-Net model will be used to segment the abnormalities. Once we have the base model, we can perform additional pre-processing techniques and tuning of hyper parameters to explore better results. The IOU metrics will be used to find the best results.

4.0 Summary

The research is aimed to propose a segmentation model using U-net architecture and experimentations on pre-processing and hyperparameter tuning will be done. The focus is more on identifying the mass abnormalities even when they are small and reducing the false positive to reduce the anxiety on patients.

CHAPTER 4

Implementation

4.1 Introduction:

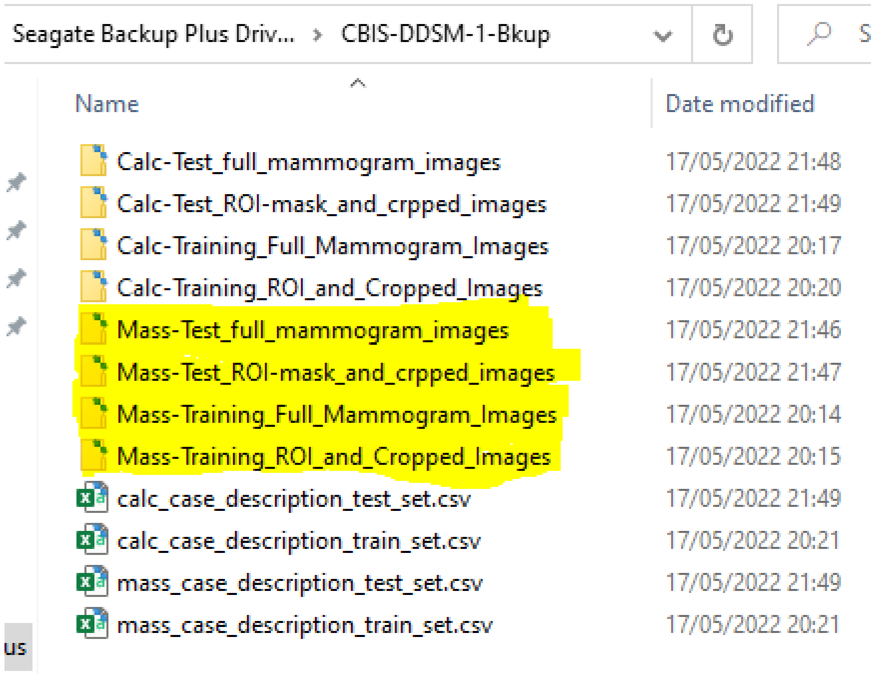
In this chapter the datasets are explored, pre-processing steps are performed as a pipeline. The files are reorganised for ease of modelling and finally feed to the model. The transfer learning using VGG16 and ResNet are used for deep learning. The different metric to measure the mass detection are discussed elaborately. The research work of Toprak A et al is taken as benchmark and aimed to improve on the existing metrics of (Toprak et al., 2022) and (Can you find the Tumour, n.d.).

4.2 Dataset:

The dataset is available in the wiki page of cancer imaging archive (CIA) for public download. The dataset includes the normal, benign and malignant images and the masks are handcrafted by trained mammographers. They are available in Dicom format with ROI segmentation and bounding box. The following pre-processing steps are already performed in the dataset by CBIS-DDSM:

1) The images where mass was unclear are removed
2) This is a facelift of DDSM dataset with image decompression
3) The images are corrected, and metadata is processed for the modern tolls to access it readily
4) The abnormalities in the images are cropped
5) As mentioned, the segmentation algorithm is applied for precise segmentation
6) Train-Test split is done for appropriate categories

The data is available for download through NBIA data retriever provided by CIA. The images are categorized and available in Mass and calcification split and further split on train and test data. The focus of the thesis is on Mass segmentation and the data related to calcification are not considered



4_1 Downloaded folder structure

Though the files are in separate folders for Full and mask, the names of the files are not descriptive of its type. As an initial step the name of the patient Id is retrieved from DCM image and its description is appended as per the type of the file. For example:

Full Image - Mass-Training_P_00001_LEFT_CC_FULL

Mask Image - Mass-Training_P_00001_LEFT_CC_MASK_1

Crop Image - Mass-Training_P_00001_LEFT_CC_CROP_1

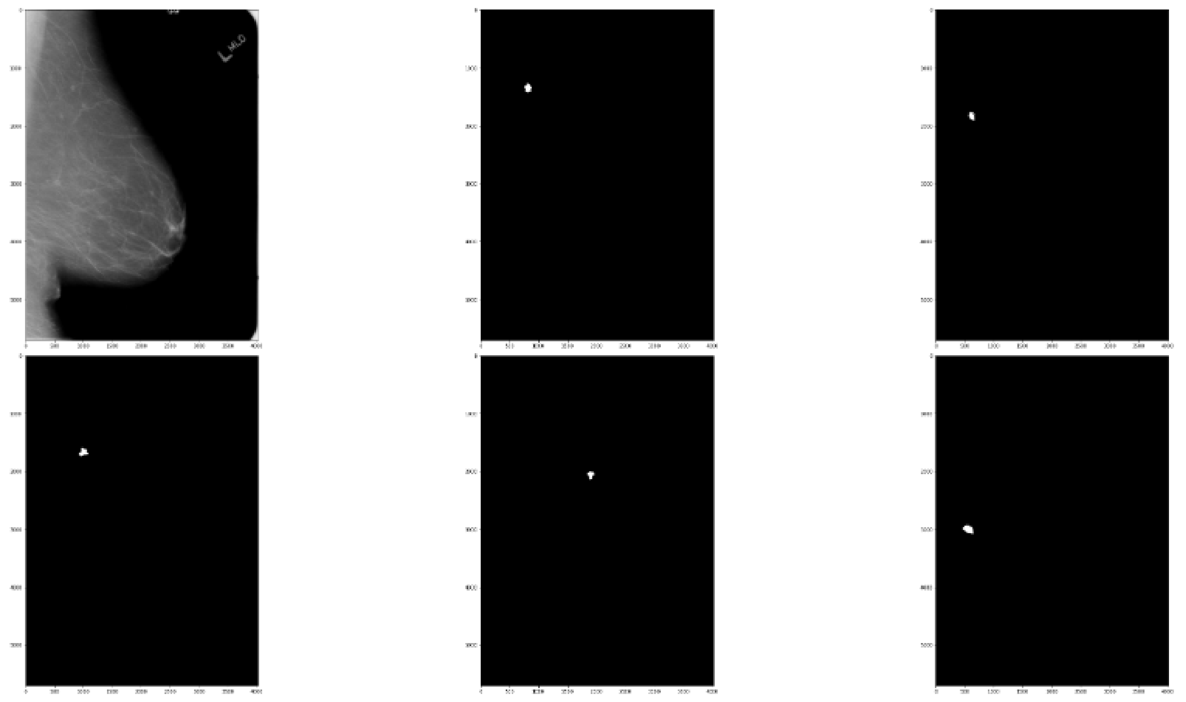
This naming convention helps to identify the file characterises by its name instead of location. The full Image is considered as an input and the mask image is the predictor. For this research the crop image files serves of no use and hence ignored for preprocessing. The images are processed and moved in a separate layer for further pre-processing. The total number of files were compared before and after renaming to ensure the integrity

4.3 EDA:

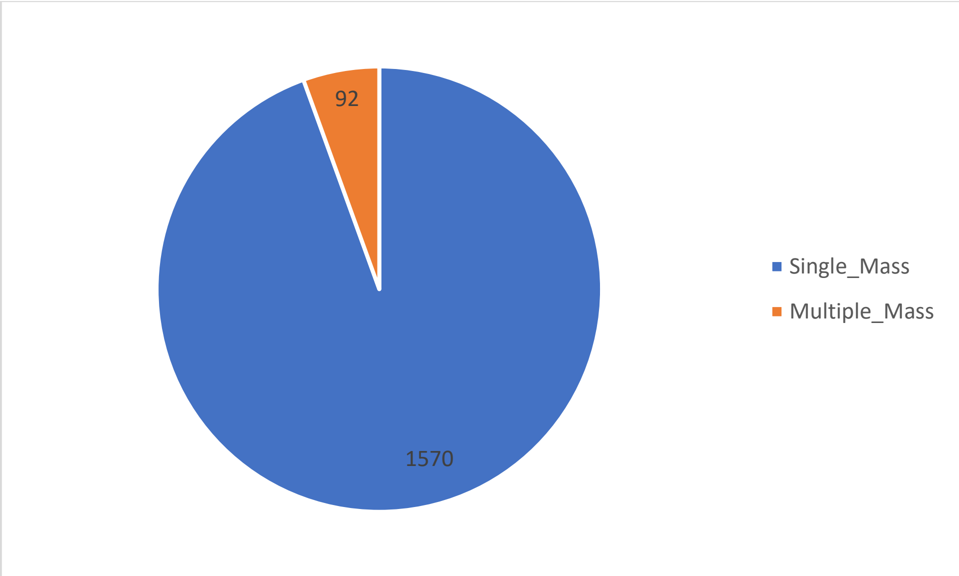
After splitting and recategorizing below are the total number of files for pre-processing.

```
Total number of files in Train: 1117
Total number of files in Test: 3867
-------------------------------------------------
Total number of files in Train for FULL images: 1231
Total number of files in Train for MASK images: 1318
-------------------------------------------------
Total number of files in Test for FULL images: 361
Total number of files in Test for MASK images: 378
```

As noticed the full and mask images are x and y respectively which is expected to equal in number for the model to process. The imbalance is due to the presence of multiple abnormalities present in a single image . Below is one example of multiple mass abnormalities in a single patient :

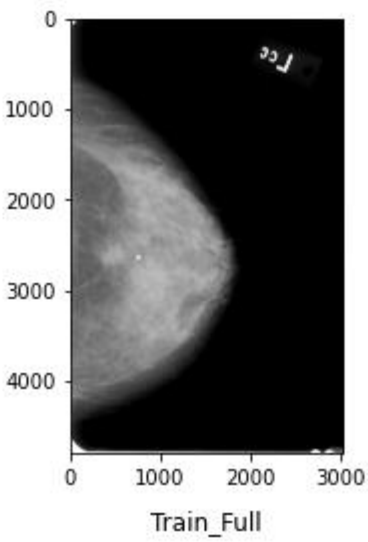


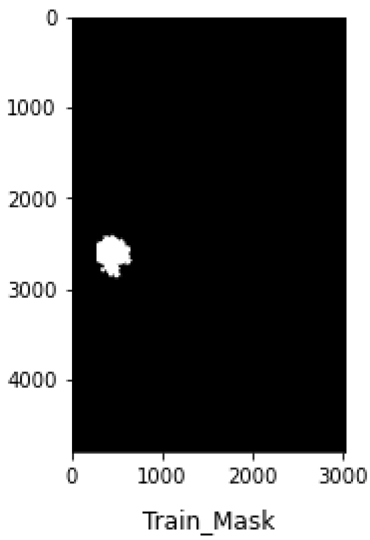
4_2 Mammogram scan with masks

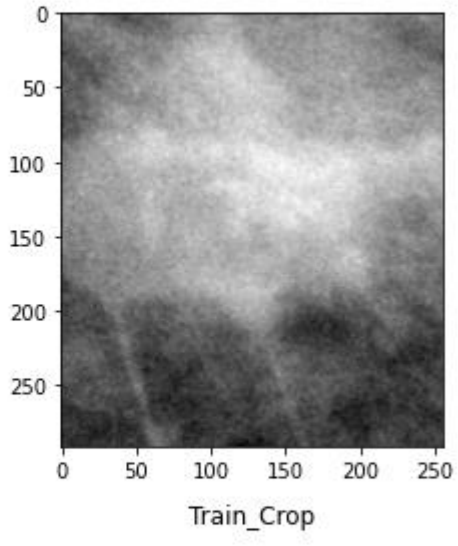


4_3 Mass abnormalities single vs multiple

As shown in the images mammogram full, mask and crop image are in left or right orientation and they are marked with in the images. The Left or right marker present at top corner of the image could confuse the model which needs to be removed. Also, some images in left and right will also be challenging for the model in learning the pattern, they need to be unified

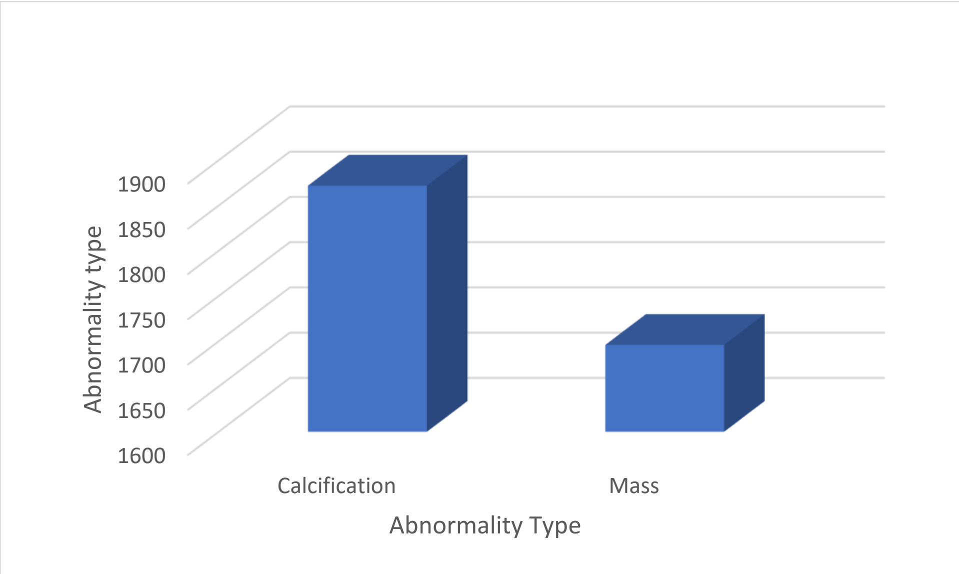






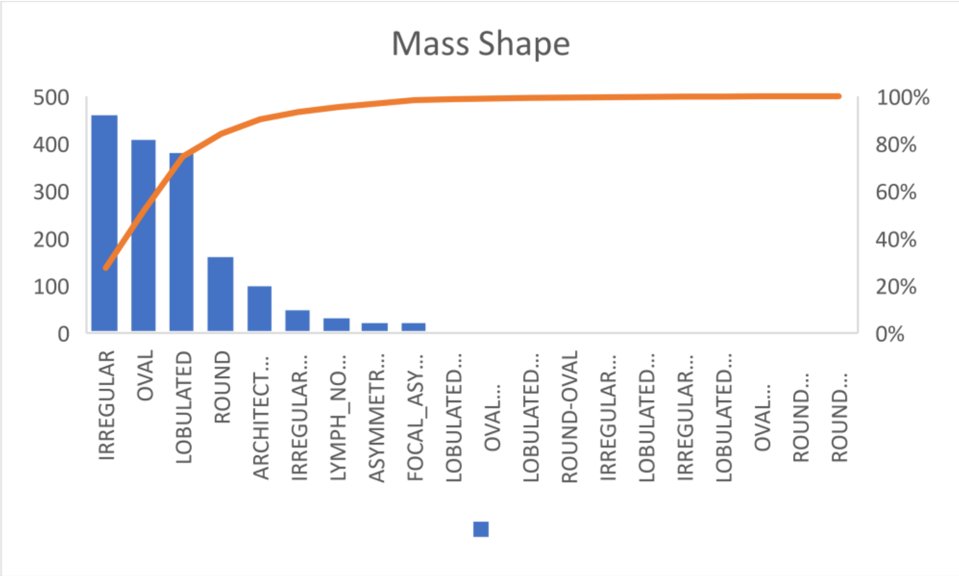
4_4 Mammogram Full, Mask and Crop

There are around 1872 calcification cases and 1696 Mass abnormalities overall.



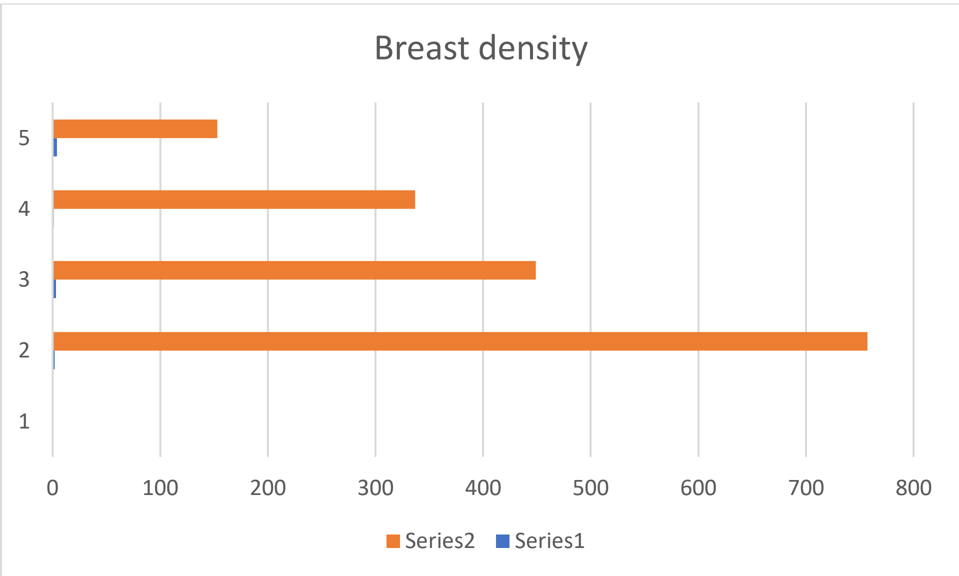
4_5 Abnormality types

The description file along with the dataset provided contains information about the available sub type of mass shape. Out of which Irregular, Oval and Lobulated are the most absorbed.



4_6 Mass shape types

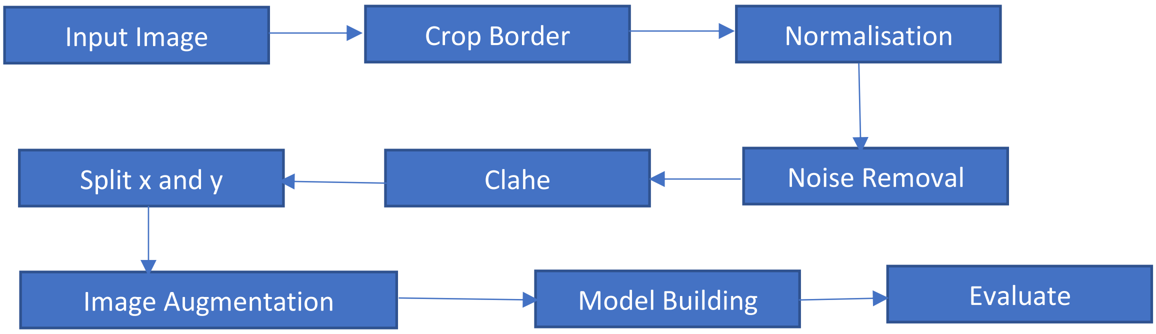
As discussed earlier there is a corelation between the mass abnormality and the breast density (Giger et al., 2013). , the different types of densities are available out of which the density above 2 is a possible abnormality.



4_7 Breast density

4.4 Pre-processing:

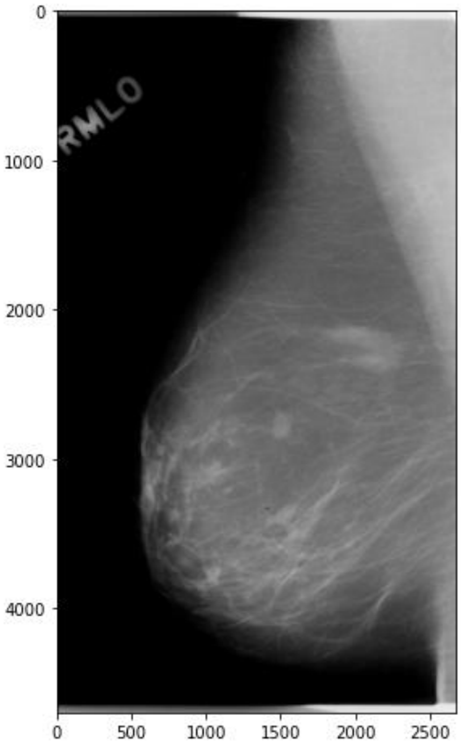
The images were already pre-processed by CIA before making available for the public. Still with the performed EDA there is a scope for pre-processing the image to be compatible for the model to process. The pre-processing steps are sequence of steps performed separately for Full and Mask data as all the steps are not applicable for Mask.



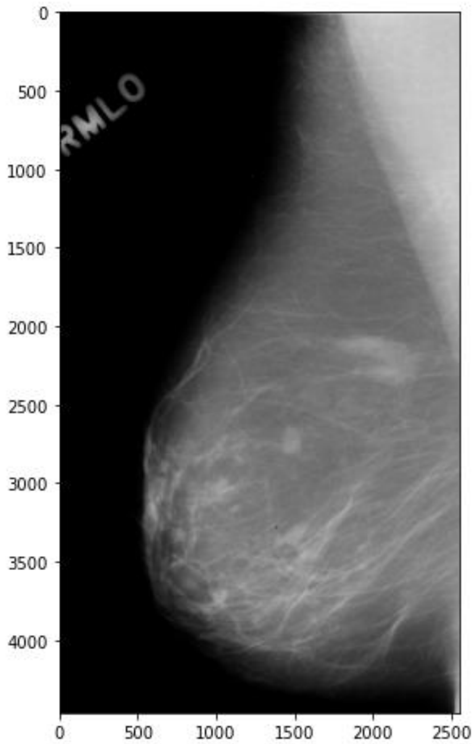
4_8 Research methodology

4.4.1 Crop images:

The mammogram images are expected to have the breast region and the abnormality in the white colour. In some cases, the images which are provided by CIA are having white patches in the border area. It is challenging to figure out only the images with white patches and remove them as it involves manual examination, and it has a good chance of miss. The images are hence cut with a default minimum value on all the four sides, provided it doesn't have any impact on the actual image. As shown in image cropping image the white border at the top is removed by cropping in the processed image



Before cropping Mass-Training_P_00871_LEFT_MLO



After cropping Mass-Training_P_00871_LEFT_MLO

4_9 Image cropping

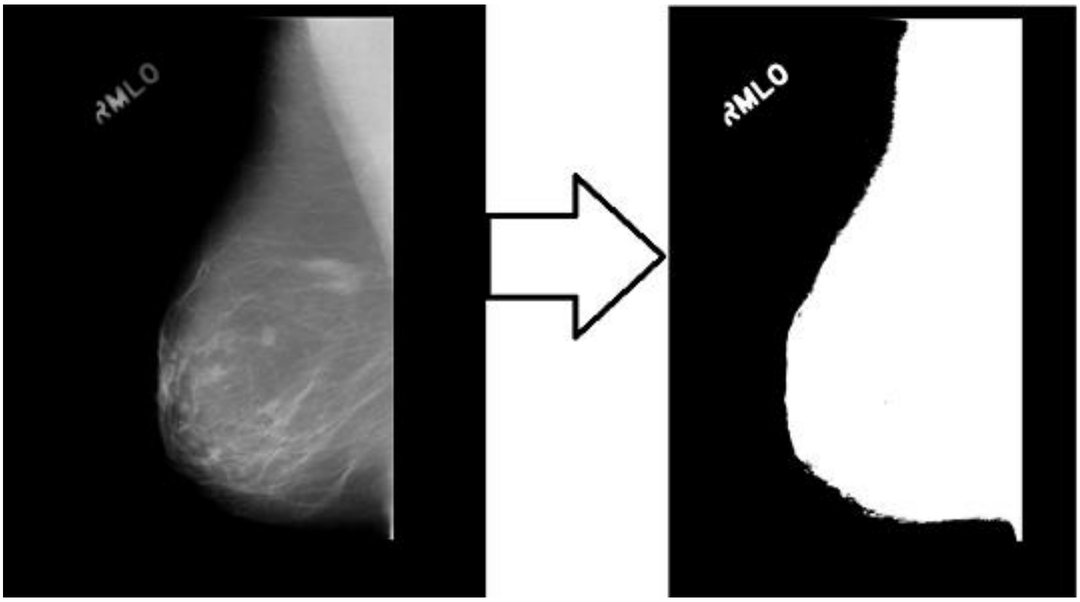
4.4.2 Normalization:

In image processing Normalization is an important step to ensure the distribution of the data is uniform. If the pixel values are at different scale, they don't tend to contribute to model fitting and will ending up with overfitting. the Min-Max Normalization is used i.e., The minimum and maximum value of a pixel will be 0-1.The only caveat to usage of Min-Max normalization is it doesn't work well on outliers. As the pixel values of an image are from 0-255 it is safe to use here.

4.4.3 Noise removal:

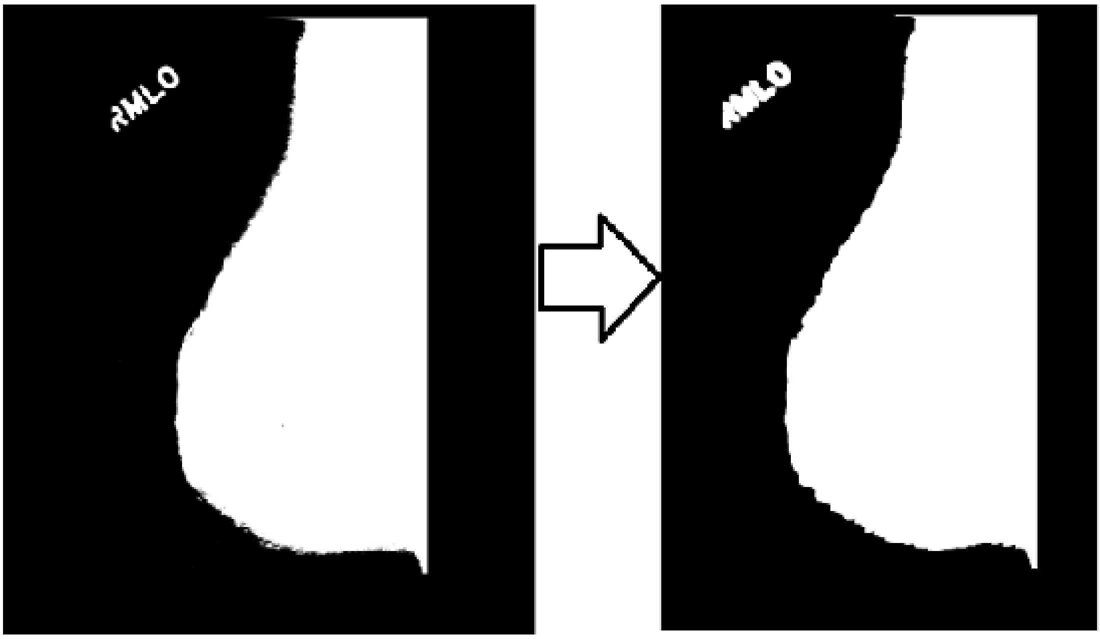
The focus area for the mammogram images is the breast area and the presence of tumour inside it. The white patches or letter markings present other than the focus region are distraction for the model and hence they need to be removed.

The first step towards noise removal is to create a mask for the breast region. This is done by checking for pixel greater than 0.1 with white shade. This may mark the non-breast region also as white.



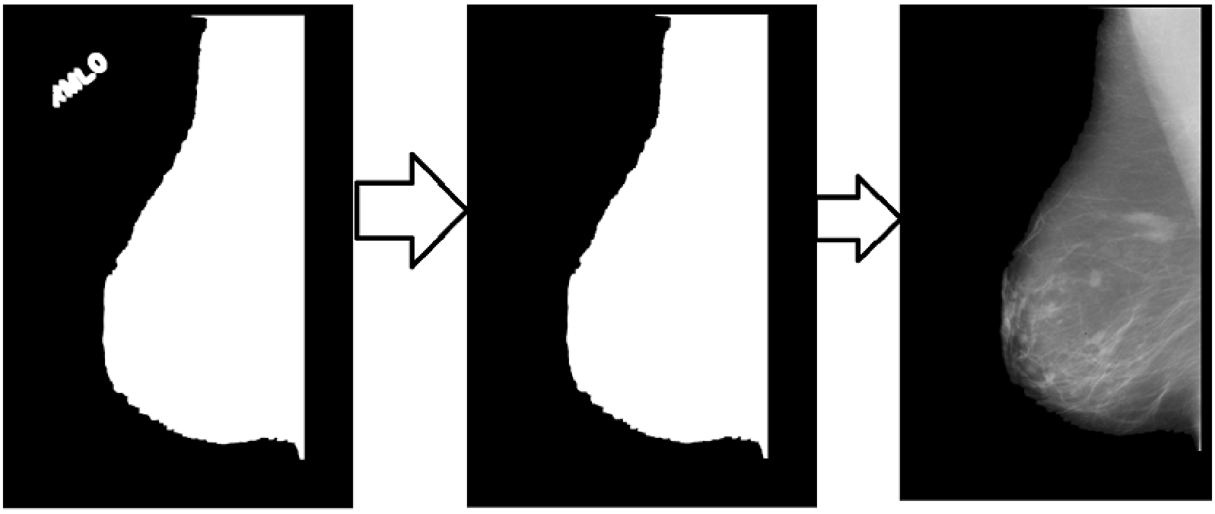
4_10 Mask creation

The images are then performed for morphological operations on its shape with kernel size of (23,23). This open operation is performed which slightly reduces the border and then enhances it better. With this there are only 2 segmented areas one with white and another with black.



4_11 Morphological operation

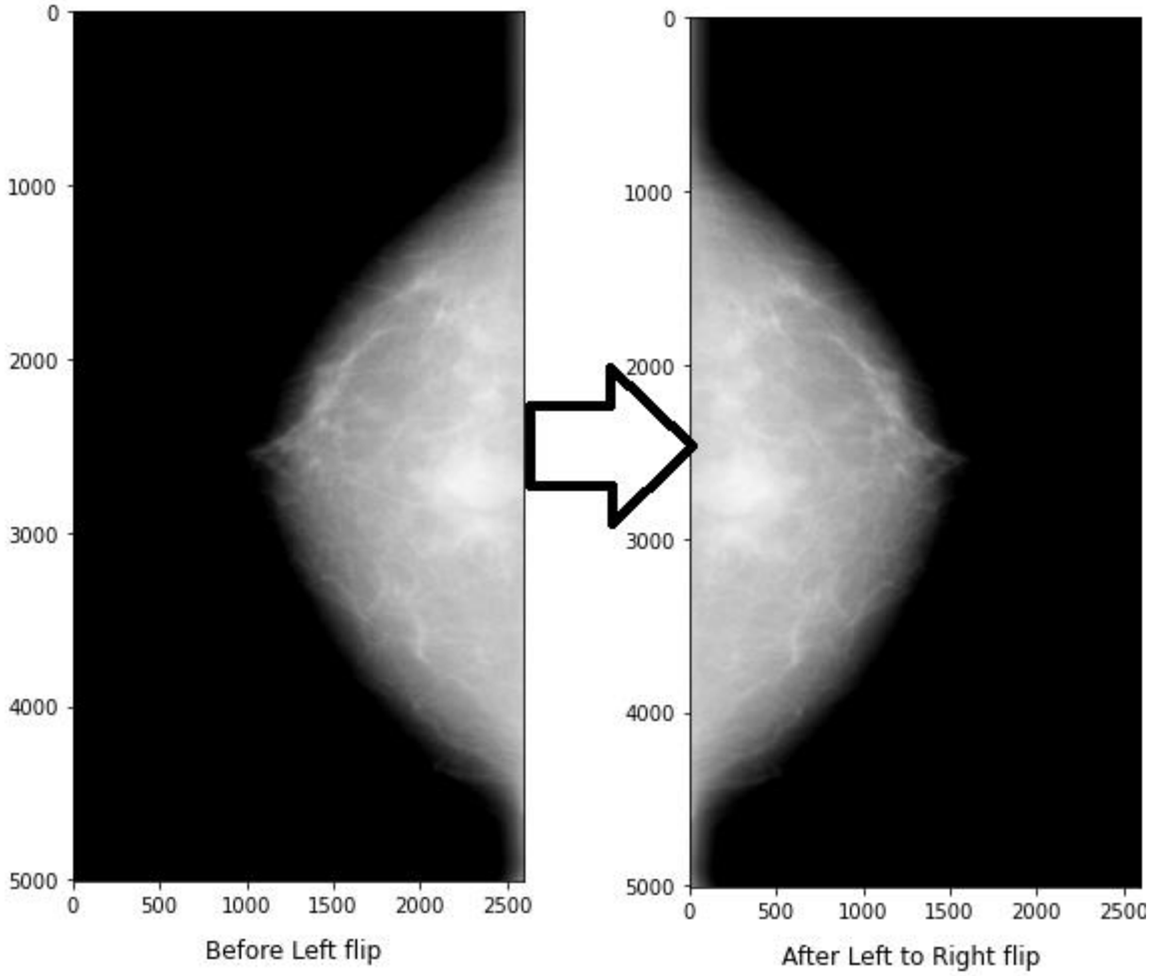
To remove one or more distortion , easy way is to find all the contours and sort them by order and retain only the biggest contour. Contours is a curve joining all the continuous points having same colour or intensity. The biggest contour in mammogram images is breast and the other distortions needs to be removed . This mask region is overlayed on the normalized image and other than white region is made black.



4_12 Noise removal

Out of 2348 images there are 1592 are left facing. Training a model with both directions could create confusion to the model. As the number of images are less for training and test, the images are shifted to left as it is the majority. The images divided in centre and the total pixels are summed. if the left side has more value the image doesn't need a flip and if on right side a flip is done. The same flip part is done for masks later to match the dependent and predictor variable on sync. This also helps on the padding step as all the images are on left, we confidently pad on right side.

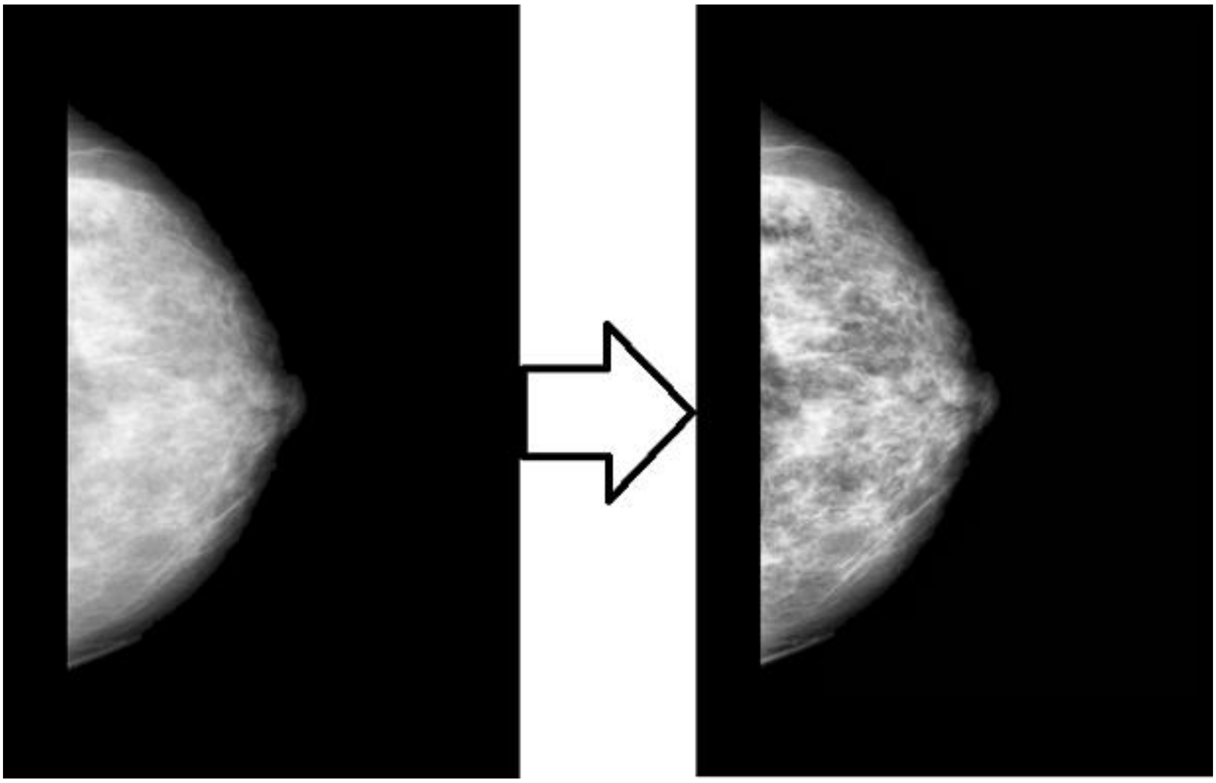
4.4.4 CLAHE:



4_13 Flip Mammograms

To eliminate unwanted characteristics and data repetition we use Normalization to change images pixel intensity. This in turn improves the contrast to help image segmentation by making the image less stressful, more normal and makes more visible. The image histograms pixel intensity can be updated by adjusting the global contrast of an image. This is a basic image processing technique known as Histogram equalization.

To achieve a high-quality output image the histogram equalization can be further improved by applying Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. This is a frequently used critical technique applied in scans to improve radiographs contrast which in turn helps in accurate diagnosis. The right-side image in (TBD) is clear and has improved global contrast of an image by spreading pixel intensities. All the images with different width and height are made uniform as the models in transfer learning has same column and rows.



4_14 CLAHE

4.4.5 Mask pre-processing:

The ground truth variable is the masked image. Though there were several pre-processing techniques used for images, the mask needs only 3 steps:

1) crop the border so that the white patches on corner can be trimmed off
2) The images for which were flipped from right to left, the corresponding masks also needs to be flipped
3) The masks also need to be padded of similar of its corresponding image

As the EDA represented 92 images with multiple pixels, there cannot be multiple predictor variable image for same input image. Passing different output variable for same input image would affect the learning of the model. Hence only those with multiple abnormalities for same patient are merged into single output variable.

The input and output were kept in same folder for ease of processing, based on the name they are placed in appropriate folders.

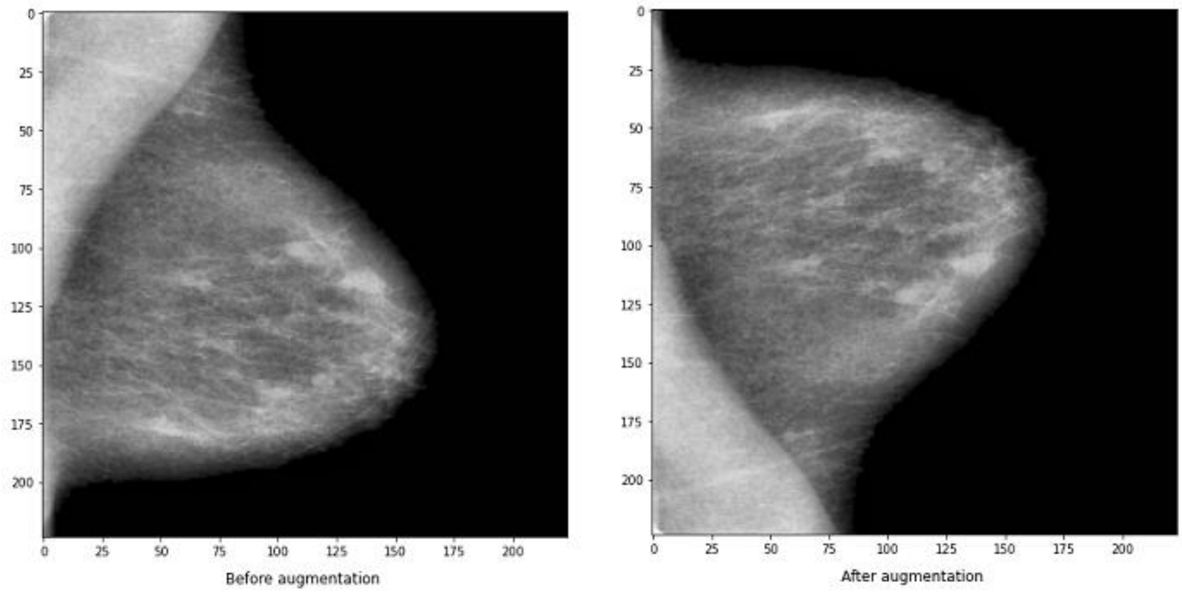
4.5 Model building:

The raw images are pre-processed sufficiently, and distortions are removed, made efficient for the segmentation model to learn further. The input images are converted into slice of array in the form of object using from_tensor_slices available in Dataset class. The images are read and resized and per the input size of the transfer learning model. The input variable is of 3 channel and the target variable is of 1 channel.

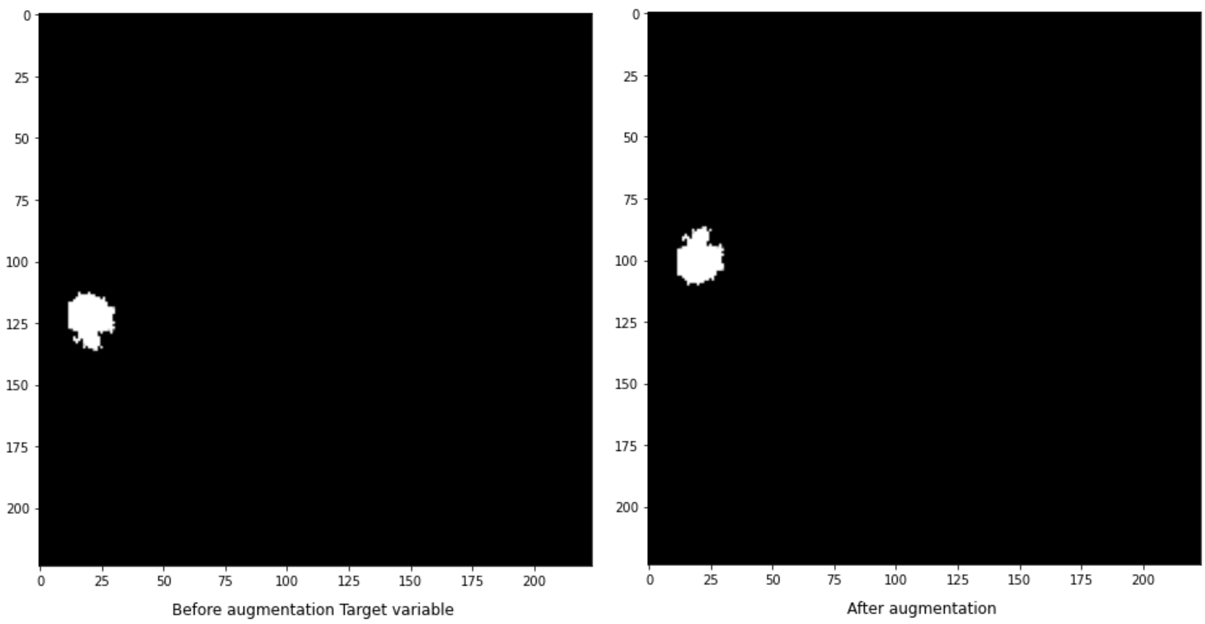
4.5.1 Image augmentation:

As the number images are relatively small in small, Image augmentation can be applied to rectify it. We can alter the exiting data to create new data for model processing. Though there are several augmentation techniques available, as the pre-processing is on medical images the choices of techniques are few for us to proceed. The below 2 techniques are used

1) Flip_up_down : The images and the mask are flipped up and down. This is done only for 50% of the images.

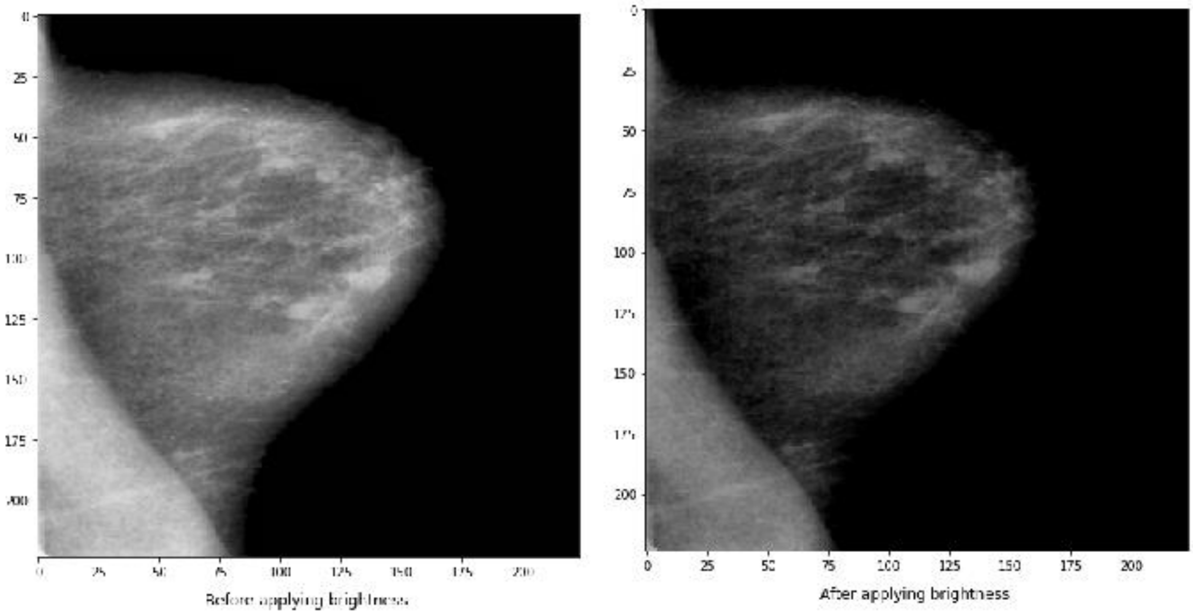


4_15 Image augmentation



4_16 Mask augmentation

2) Random_brightness: Improve the brightness of the image for input image.



4_17 Image brightness

4.5.2 VGG16:

The VGG16 model that was trained on different dataset is used for training the image segmentation. As the name suggest it is a convolutional neural network of 16 layer deep. This is trained with a dataset of 14 million images belonging to 1000 classes. This model expects the input image to be of size 256 * 256. The top layer is excluded with weights frozen and used as an encoder block of U-Net segmentation. The decoder block is built with the combination of convolution 2D Transpose and dropouts.

The final image is enlarged to the original size of 224*224*1. The convolution layers have Relu activation function, and the final layer has sigmoid activation function. As there are two classes of classification Binarycrossentropy is choosen. The Adam optimizer is used for adjusting the weights. The images are trained in batches for multiple epochs.

4.5.3 ResNet:

The Residual Neural network (ResNet) trained with very deep neural networks with 50, 101 and 152 weight layers. This solves the vanishing gradient problem that made the learning plateau or degrade when training very deep neural networks. The ResNet introduces skip connection that allow information to flow from earlier layers in the network to later layers, creating an alternate shortcut path for the gradient to flow. The images are resized to (512,512,3) to feed into the trained Resnet network.

4.5.4 UNET:

The U-Net architecture is split into Encoder and decoder block. The VGG16 model using transfer learning with weights frozen are used for the encoder block without the topmost classification layer. The features extracted out of this layer is used for skip connection and output of this encoder. This output layer behaves as the mediator for encoder and the decoder architecture

The decoder block is built manually which consist of 1*2 Transpose convolution layer and the skip connection is taken from the pretrained encoder. To reduce overfitting the L2 regularization is added in con 2d layer. The conv2d layers has activation function as "Relu" with padding "same". The best models are saved in hdf5 format. The early stopping is configured for IOU and dice metrics. The batch size and epochs are tried with different combinations.

4.5.5 Error function:

The intention of building a model is to evaluate how well it is performing with an unseen data. The simplest metric is Accuracy, which is not a good measure for image segmentation. This is because the focus of the prediction to find out the anomaly which is small in size. The accuracy metric will be focused on predicting both the abnormality and non-abnormality area.

The preferred metric for segmentation is IoU, which helps to measure the accuracy of the image segmentation. The result will clearly indicates the accuracy of the ground truth object with predicted segmented objects. If the value is close to 1 the prediction is exact match, if 0 there is no overlap and any value above 0.5 is a good result. The custom implementation IoU is used.

Dice coefficient is the 2 * area of overlap divided by total number of pixels in both the images. This also follows the same trend of 0 being no overlap and 1 as the best. That is, Dice score is just similar representation of IoU under the numerical sense. It's enough to only using one of them for model comparison.

4.6 Summary:

Overall, the effort has been spending mostly on preparing the data for the model and feeded. The best performing models based on the metrics are saved along with other data such as tensor logs, check points, csv logger, model history and model parameters. The training is stopped if there is no significant improvement in the metrics.

CHAPTER 5

Results and evaluation

5.1 Introduction:

The CNN network is a Blackbox. If it works well there is no defined reasons, but when it doesn't work as expected i.e., overfitting or underfitting based on the intuition and experience we can tune few hyperparameters, try a different model e.t.c., to get the desired results. The benchmark of this segmentation is around 0.2 and this research is aimed to improve on that baseline (Can you find the Tumour, n.d.) and (Toprak et al., 2022).

5.2 Experiments:

There are lot of pre-processing steps involved in preparing the image to be ready for the model to learn. Experimenting in the pre-processing pipeline is a tedious task as all the steps involved were performed was to clean the image. The experiments are done after pre-processing. The different experiments were performed on tuning the below hyperparameters

1) Learning rate and decay schedule
2) Dropout
3) Epoch
4) Batch size
5) L2 regularisation
6) Batch normalisation
7) Kernel regularizer
8) Bias regularizer
9) K-Fold cross validation

The stride, kernel size, padding is kept constant. The images are pre-processed as mentioned in chapter 3 research methodology and converted to .png format throughout the experiments.

5.2.1 VGG16:

The VGG16 is used for transfer learning for the encoder, and it accepts 224 as the image size and the channels are set to 3. To compare the metrics arrived at each experiment the seed values are unchanged. The metrics measured are IouMetric and dice score are evaluated. When the learning rate is decease to 0.00001 the training data was underfitting with dice matrix around 0.264 for train data.

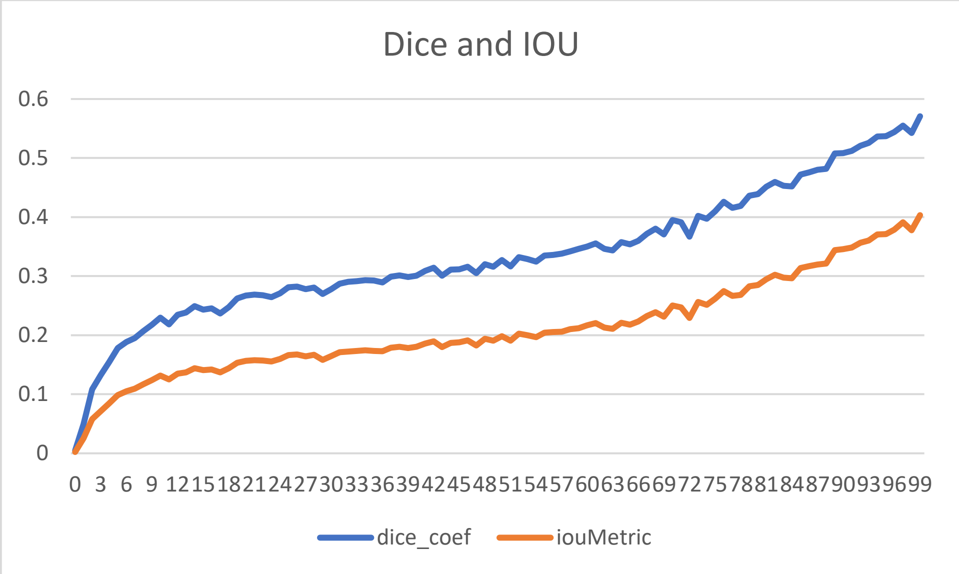
These initial values are set to batch size 10, 80 epochs, learning rate as 0.00001 and the model was trained and evaluated. The training took more time to converge, the loss reduced at a shallow rate and the model couldn't learn the patterns in the train data. It is as an underfitting model. As reducing LR didn't help, resetting it to 0.0001 and increasing the epochs ended up as an overfitting model. It failed to generalize the patterns and performed poorly with the test data. Model with learning rate 0.001 the gradients were oscillating around 0.35 the local minima and didn't improve. The learning rate 0.0001 is the optimal value for the model building.

There were multiple experiments performed and observed for its performance, to determine if there are any bottlenecks impacting its performance and identify the areas of improvements. The trained and evaluated model was overfitting and occasionally underfitting in multiple experiments. There is no scope for adding more layers as overfitting it increases the metric gap further between train and test. Nevertheless, the metrics of training is increasing, and validation loss was decreasing when number of epochs were increased. To resolve overfitting

- Collecting more data is not possible as these mammogram images are handcrafted by trained radiologist.
- Dropout layers are for switching off the neurons randomly during the training. This helps instead of remembering the details of the image the patterns are learnt. The dropout values are tried from 0.3 to 0.7 . At every iteration some neurons are randomly selected by the network and temporarily ignored. The ideal value found on experiments 0.5 as the results were unstable and oscillating when changed to other values.

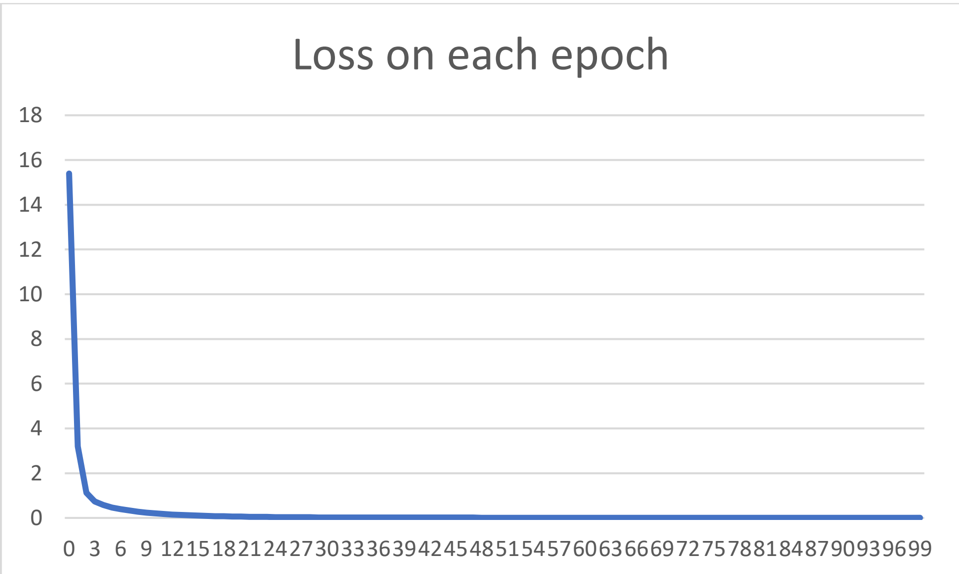
Regularization is one of the techniques to prevent overfitting in neural network and improve the accuracy of CNN model when an unseen test data is evaluated. The current model picks the noise in the training data during the training process and studies it intensively instead of learning the pattern. If the learned noise is unique to each mammogram doesn't pose a problem, and in case if they are not and using it for prediction results in overfitting. There are L1, L2 and dropout for regularization and dropout has not helped much with layers the model had.

L2 regularization popularly known as ridge regularization the weight parameters are encouraged to decrease but doesn't become zero. This helps to reduce the complexity of the network and prevent overfitting.0.01 to 0.001



5_1 Dice and IoU with L2

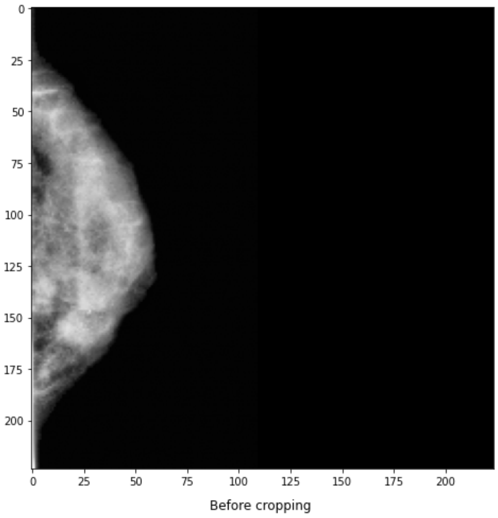
In the convolutional layer the kernel, bias and activity regularize are also tried for reducing overfitting. The kernel regularizer reduces the weight, bias regularizer reduces the bias and the activity regularizer reduces the layers output which will reduce the weight and adjust the bias. Making the weights not heavy the network instability can be prevented. Activity regularization is used to encourage a neural network to learn sparse feature representation of the training data being passed to the model. The kernel and bias with L2 regularizer have helped in controlling the overfitting to some extent



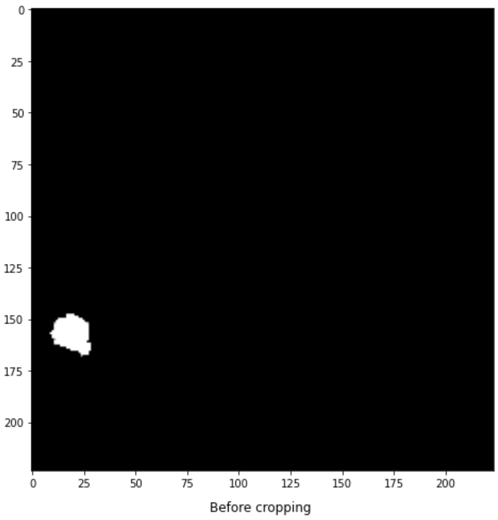
5_2 Loss on every epoch

To improve the IOU and dice metric the model can be validated before evaluating it on the test data. This process of deciding if the numerical results quantify hypothesized relationships between variables, are acceptable as descriptions of the data, is known as validation. The test images are around 361 and reserving them for validation is not a good choice. A portion of training data is used for validating to improve the to test the effectiveness of the model.

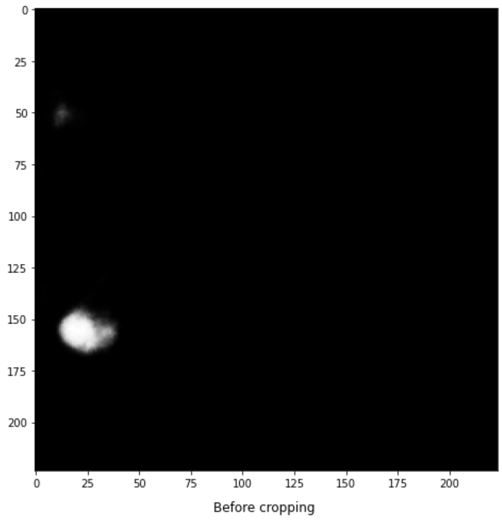
And the better way to reserve the train data and use it appropriately would be to K fold technique. As every observation from the original dataset will get a chance to appear in training and validation. The model was trained with k-fold of 6 and the best model was evaluated and found to be overfitting.



5_3 Input image VGG16



5_4 Masked image VGG16

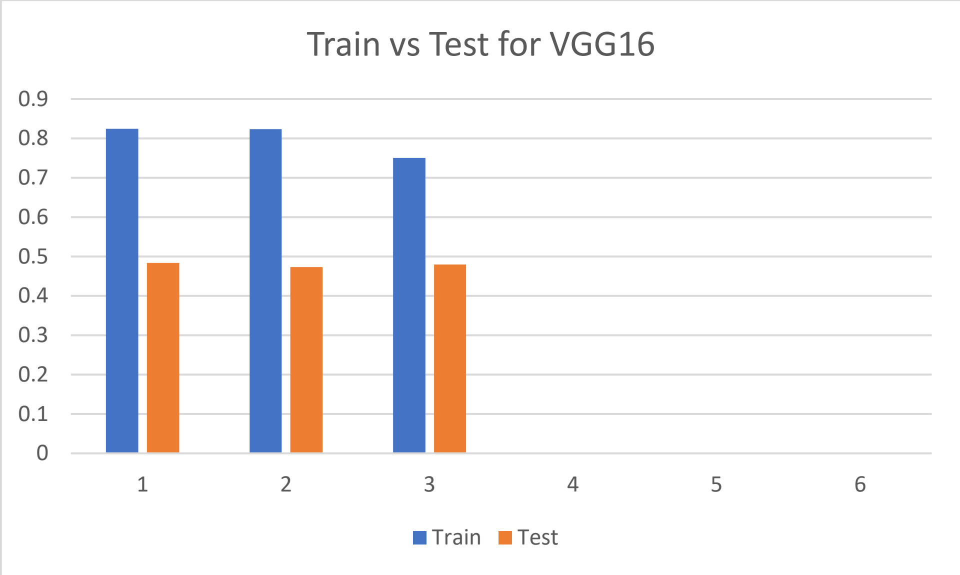


5_5 Predicted output VGG16

The Batch size varied from 1,10,20,50 for model training and the overfitting is observed when the batch size is above 10, hence 10 is kept as an optimal value. The epochs were tried from 50, 80, 100, 125, 150 and 175 there is not much learning happened after 80 epochs.

5.1 Performance on changing learning rate VGG16

S.No	Learning Rate	Train	Test
1	Initial	0.7879	0.34
2	0.01	0.4101	0.32
3	0.001	0.31	0.24
4	0.00001	0.264	Nan
5	0.0001	0.5708	0.34
6	Changed to 20 Batch	0.72	0.35



5_6 Train test comparison VGG16

The best model out of multiple experiments with VGG16 have a train of .7204 of dice metrics and IOU metric of 0.565 in training. This was trained for 100 epochs with a batch size of 20 and dropout of 0.5. The model was overfitting but still performed better with the unseen test data , achieving 0.3527 of dice metric and 0.2208 of IOU metric. Though there is a gap between train and test it is not much as expected. The segmentation of output mask has some noise captured as seen in the image. The loss has been observed at 16 on first epoch and has been decreasing consistently and was at 0.0075 at the end of 100 epoch.

5.2 Overall performance for VGG16

VGG16 Image Size (224,224, 3)					
		Train		**Test**	
	Parameters	IOU Metric	Dice Coeff	IOU Metric	Dice Coeff
Model 1	batch_size = 10 num_epochs = 80 Learning rate =0.0001	0.6519	0.7879	0.2201	0.3489
Model 2	Added Kernel and bias regularizer with 0.01	0.2628	0.4101	0.1951	0.3201
Model 3	Learning rate =0.00001	0.1546	0.2644	N/A	N/A
Model 4	Learning Rate =0.0001 Num_epochs=100	0.4033	0.5708	N/A	N/A
Model 5	Learning Rate =0.001 Num_epochs=50	0.1865	0.3108	0.1424	0.2459
Model 6	Num_epochs=100 Kernel regularizer = 1e-4 Without bias regularizer	0.1173	0.2077	0.0903	0.1643
Model 7	Learning Rate = 0.0001	0.3185	0.4754	0.1934	0.3147
Model 8	Kernel regularizer = 1e-5	0.5259	0.6853	0.245	0.382
Model 9	Activity regularizer = 1e-5	0.0034	0.0069	N/A	N/A

Model 10	Switched skip outputs (Connecting encoder to decoder) Removed activity regularizer with Kernel Regularizer	0.565	0.7204	0.2208	0.3527
Model 11	K fold cross validation	0.195	0.3223	N/A	N/A

5.2.2 ResNet:

As there was overfitting with VGG16 the same hyperparameter were retrained with U-Net architecture and ResNet was used for transfer learning. The experiments were performed for shorter epoch (around 40) initially to see if the model if overfitting. The ResNet model was made less complex compared to VGG16 to ensure prevention of overfitting. The initial results were shown as 0.56 for train data and the test with 0.21. The change of Resnet has not improved preventing overfitting initially.

```
Model: "ResNet50_U-Net"

=======================================================
Total params: 20,672,705
Trainable params: 20,640,193
Non-trainable params: 32,512
_______________________________________________________

Size of training set = 1231
Size of test set = 361
Number of epochs = 40
Batch size = 10
Number of training steps per epoch = 124
Number of test steps per epoch = 37
```

5_7 ResNet model summary

The L2 regularizer along with the dropout layer were tried with different values to reduce the overfitting. There are multiple regularizer which can be applied for convolution layer. The Kernel Regularizer is applied to reduce the penalty on the layer’s weights. The bias regularizer was applied to reduce the penalty of the bias layer. The activity regularizer is applied to reduce the penalty of the activation layer. These values are divided by the input batch size such that the relative weight between weight regularizer and activity regularizer doesn’t change the batch size.

The values of regularizer with value of 0.1 , 0.001, 0.0001 are applied for the both the convolution layers for each experiment. The values of dropout layers are kept 0.5. All these experiments had similar overfitting results observed. The train and test had a gap of 0.3 on an average as shown in 5.3 . The images are sent on batch on each epoch for training, based on the error the weights and bias are optimized and the optimal values are kept changing. This is an internal covariate shift.

5.3 L2 Regularizer ResNet

S.No	**L2 Regularizer (Kernel)**	**Train**	**Test**
1	No Regulariser	0.5665	0.21
2	L2 regularizer 0.001	0.46	0.1919
3	L2 regularizer 0.01	Nan	Nan
4	L2 regularizer 0.0001	0.53	0.26

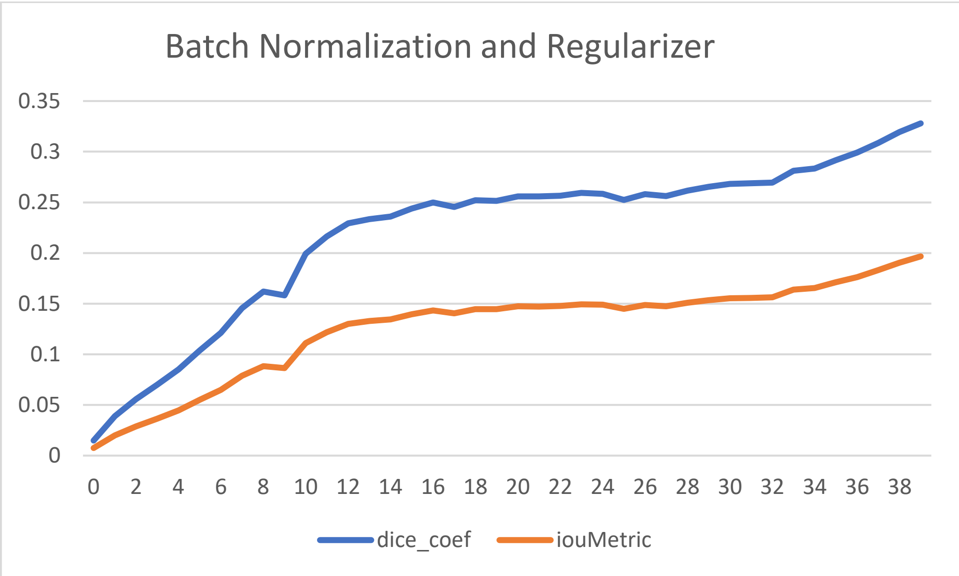
In these experiments initially batch normalization and L2 regularisations were used together. The weight decay from L2 Regularization on Conv2d layer is not making any impact. With batch normalization all the weights in hidden layer will be equally happy at the decayed value. This batch norm layer will revert the decay performed by L2 regularization. This is not applicable for the layers where normal regularization is applied. Hence the further trainings are performed without batch normalization and with L2 Regularization.

There are 2 convolution layers used in the decoder part and kernel regularization was added to later and experiments were performed. This is different than the previous experiments mentioned in TBD as the regularization was applied to both the convolution layers. The regularization was tried with values from 0.1 , 0.01 and 0.001. The training accuracy has drastically improved with this approach as penalizing the weights are done only for one layer. The overfitting was also observed with these changes.

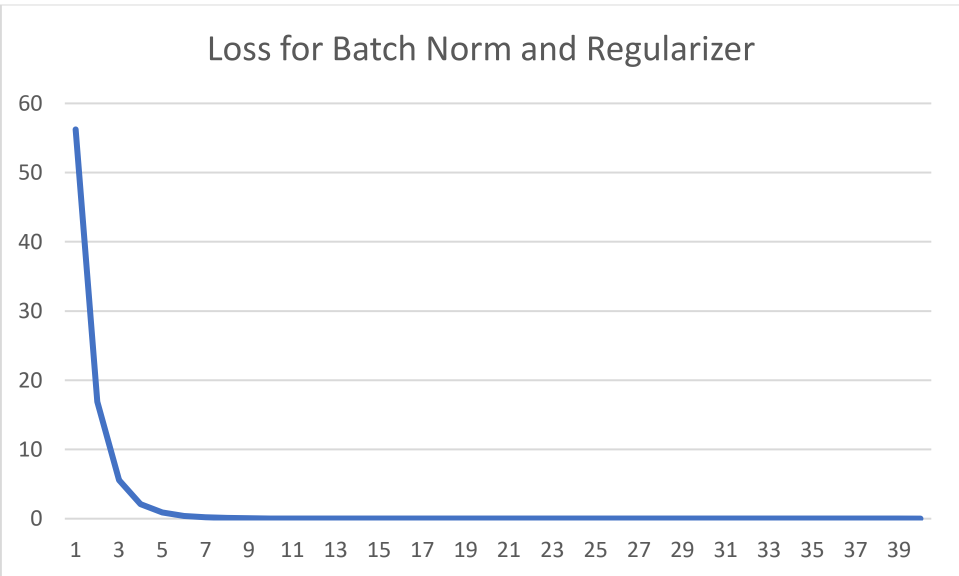
5.4 L2 Regularizer without batch normalization ResNet

S.No	**L2 Regularizer (Kernel)**	**Train**	**Test**
1	0.01	0.8242	0.484
2	0.001	0.8237	0.473
3	0.1	0.75	0.48

With the 2 convolution layers the first Conv2d is followed by Batch normalization and kernel_regularizer for the Conv2d with lambda value 0.1. The results were drastically reduced, and overfitting was observed with train 0.328 and test 0.161 for Dice coefficient. The IOU metric also at 0.19 for train and 0.16 for Test. The loss has been reducing the initial batches but towards the end the loss has not reduced, and they are mostly constant as shown in 5_9. The experiments without dropout and with Regularizer were also performed, there was slight drop on train and test. This could be because of the missing dropout the layers would have missed to learn the patterns.



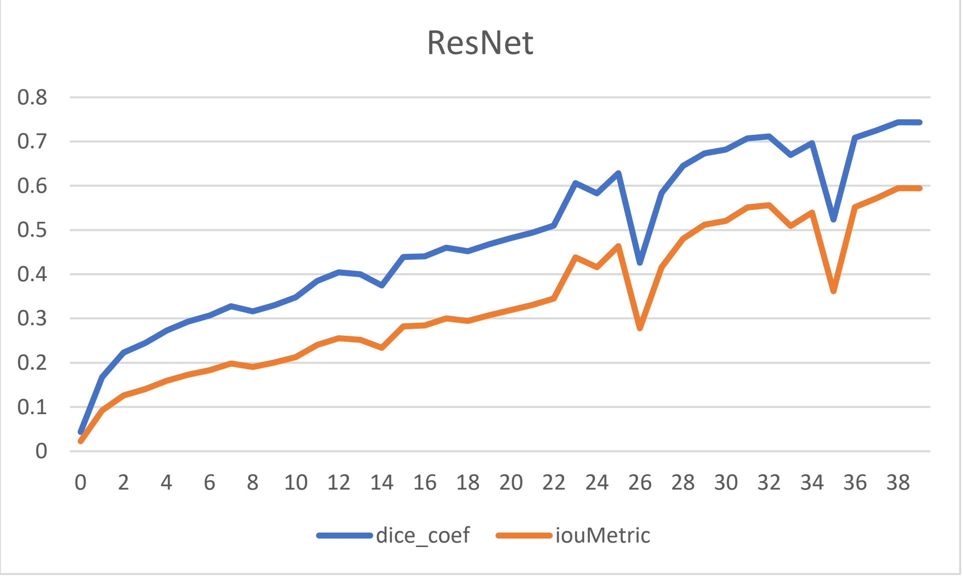
5_8 ResNet model Metric with Batch Norm and Regularization



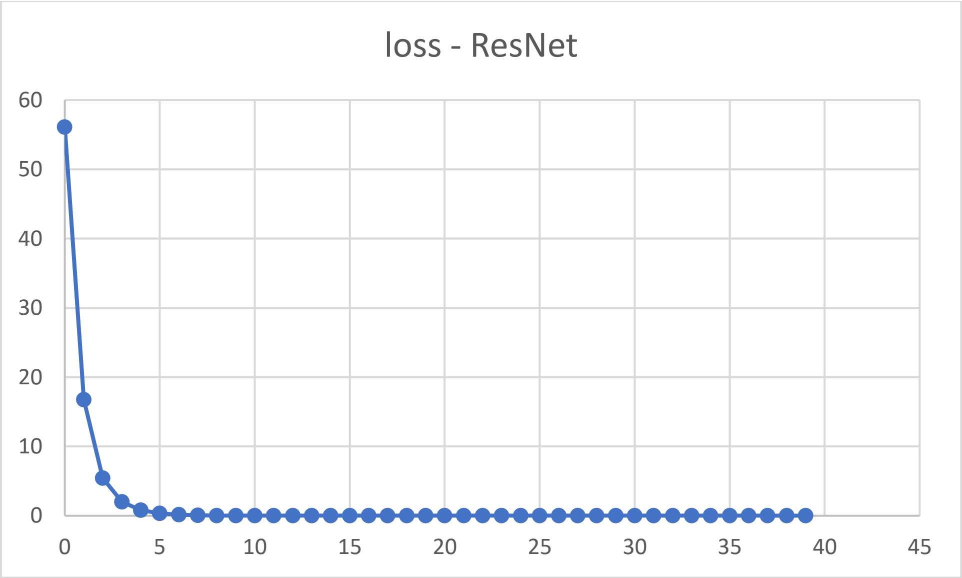
5_9 ResNet model loss with Batch Norm and Regularization

As all the possible hyperparameter tuning has been performed and the next possible solution is to perform image augmentation to resolve the overfitting problem. Since it is a medical image, a mammogram the angle position doesn't change much. There is less scope to use all available image augmentation technique as in real world of mammograms the options are limited.

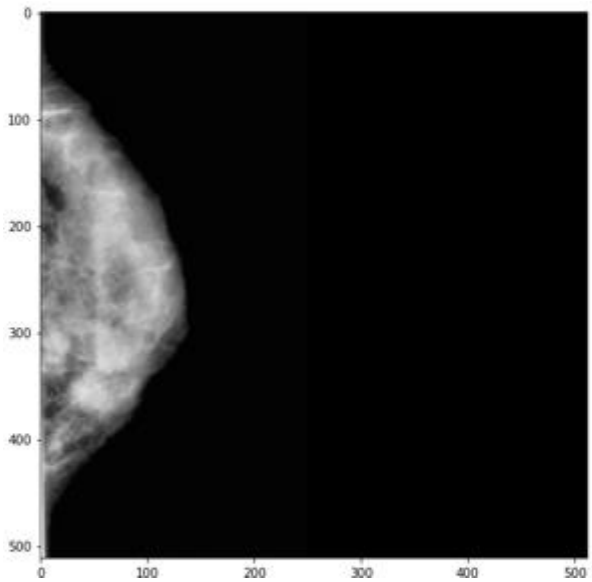
The brightness can be reduced with 0.3 to generate additional images. This helps to address the test images which are captured in different light conditions. The contrast of the images is also performed between 0.3 to 0.5. This augmentation step was not initially performed, and done later to see if there is any change in the result. This achieved a result of 0.81 in train dataset and 0.39 in test dataset for dice matrix. As this shows augmentation has also not helped in resolving the overfitting.



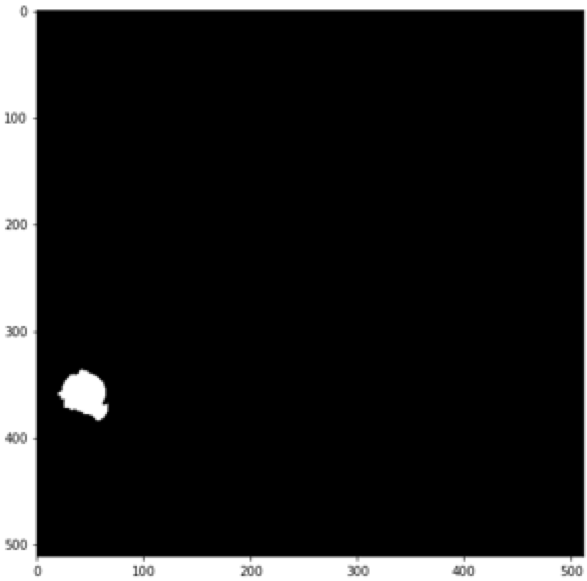
5_10 ResNet final model metric



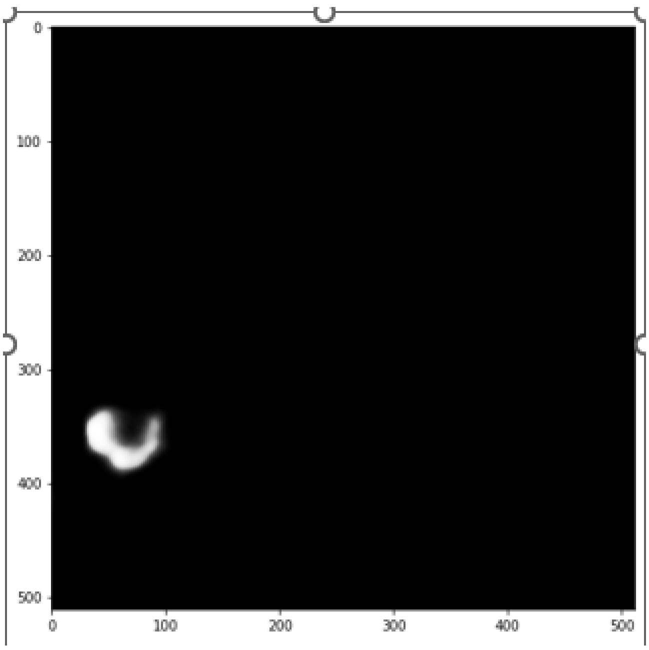
5_11 ResNet final model loss



5_12 ResNet Input image



5_13 Resnet Mask



5_14 Resnet predicted Mask

The predicted mask image is not of the same shape as the original handcrafted mask, but it is able to identify the area where the abnormalities is. The overfitting has created the problem of predicting the actual masked area though the training had good result.

The best model out multiple experiments with ResNet have a train of .8179 of dice metrics and IOU metric of .6943 in training. This was trained for 120 epochs with a batch size of 10 and dropout of 0.5. The model was overfitting but still performed better with the unseen test data , achieving 0.3956 of dice metric and 0.2591 of IOU metric. Though there is a gap between train and test it is not much as expected. The segmentation of output mask has some noise captured as seen in the image 5_14 .The loss has been observed at 16 on first epoch and has been decreasing consistently and was at 0.0026 at the end of 120 epoch.

5.5 ResNet experiment observations summary

ResNet Image Size (512,512, 3)					
		Train		Test	
	Parameters	IOU Metric	Dice Coeff	IOU Metric	Dice Coeff
Model 1	batch_size = 10 num_epochs = 50 Learning rate =0.001 No Regularizer	0.3995	0.5665	0.1262	0.2176
Model 2	num_epochs = 40 Added Droput 0.5 Bias Regularizer = 0.001	0.3044	0.4633	0.1077	0.1919
Model 3	Learning Rate = 0.0001 Bias Regularizer = 0.01	0.3714	0.5393	0.156	0.2644
Model 5	num_epochs = 50 Without batch normalisation and only kernel regularizer 0.01	0.7072	0.8242	0.3294	0.4847
Model 6	kernel regularizer 0.001	0.7183	0.8348	0.3193	0.473
Model 7	num_epochs = 40 kernel regularizer 0.1	0.6141	0.7592	0.3249	0.4814
Model 8	With Batch Normalisation on one conv2d	0.1967	0.328	0.0885	0.1611
Model 10	Without Batchnormalization and with dropout 0.5	0.5944	0.7432	0.2492	0.3901
Model 11	Image Augumentation	0.6189	0.7616	0.2542	0.3746
Model 12	Num_epochs = 120	0.6943	0.8179	0.2591	0.3956

5.3 Summary:

The pre-processed images were resized, normalized and image augmentation is performed and fed into the model for training. The U-net structure is kept constant and the hyperparameters were tuned . The transfer learning used in the encoder is tried with VGG16 and ResNet. With all the experiments the model has been overfitting. The L2 regularization is used at different level with different values which has not helped much in reducing it. Out of all the experiments performed Model 12 of ResNet has outperformed other model with an excellent train result of 0.8179 of Dice coefficient and 0.6943 of IoU.

CHAPTER 6

Conclusions and recommendation

6.1 Introduction:

In this section the interpretation of the results will be discussed and conclusion on the outcome of this research. The contributions made for this project will also be discussed and the future possibilities on improving the current result will be described to proceed further for future researchers.

6.2 Discussion and conclusion:

The early detection of breast cancer plays a significant role in reducing the risk of patients. There is much active research performed extensively and proposing new models for both classification and segmentation. The models are performing good with the train data are failing to detect or classify in unseen data.

CBIS-DDSM is one of the reliable datasets available for automating the anomaly detection in mammogram. As this is verified and curated by radiologist and released by the cancer imaging archive. The dataset is handled for class imbalance as it is split already into train and test with right mix of normal, benign, and malignant cases. This is very important activity as the size and shape of the tumour varies based on the type.

There was not much research performed till date with this dataset on solving Image segmentation to consider it as a benchmark result. The results available out of the research is also not significant to be used for production. The results of experiments performed with VGG16 are observed to be overfitting. The model performed fair when the learning rate was around 0.0001. There was not much of learning happened after certain epoch as the metrics have started flattened or reducing occasionally. As the number images used for training are less few images were reserved using K fold and validated. Nevertheless, the results were overfitting across all experiments of VGG16.

Resnet is much more advanced compared to VGG16 as the layer are much deeper and skip connection is used to resolve vanishing gradients. The experiments were performed replacing Resnet with VGG16 for transfer learning in Encoder block of U-Net architecture. The change in number of epochs has not made any significant changes in the result The optimum epoch number is observed to be around 40-50. Having both regularization and batch normalization in CNN architecture has made the model perform poorly.

The L2 regularization has not helped reducing the overfitting for both VGG16 and ResNet. The possible options of additional image augmentation didn’t improve the metrics. With several combination of hyper parameters on multiple experiments the model has tend to overfit. The possible methods need to be explored to improve the metrics from here.

6.3 Contributions:

This research has explored the significance of mammogram along with the other techniques used for breast cancer detection. The other diagnostic methods of DBT and ultrasound are evaluated and compared with mammograms. The advantages and disadvantages of these methods were discussed along with the evidence of why mammograms are being the industry accepted method for diagnosis. The usage of sonography in low-cost areas (compared to mammogram) has been discussed.

The density of the breast has always a relation with the abnormalities, and it was found during exploratory data analysis (4.7). The multiple framework and its result give a big picture for future research to choose the appropriate path for proceeding further. The CBIS-DDSM dataset has been discussed in detail and exploratory data analysis presented with details will pave way for the researcher to understand the data. The codebase created for preprocessing is very generic and could be reused for any kind of image preprocessing tasks.

For image segmentation the transfer learning has been used for multiple models. The benchmark model used as a starting point has achieved a IoU metric of 0.2 (Can you find the Tumour, n.d.) . This research has improved with better score of 0.6943 for the same dataset.

Also, the benchmark research has used the research has used only VGG16 architecture in encoder. In this research the Resnet is tried, and the results are shared for future researchers to understand the pattern of hyperparameter behavior when tuned.

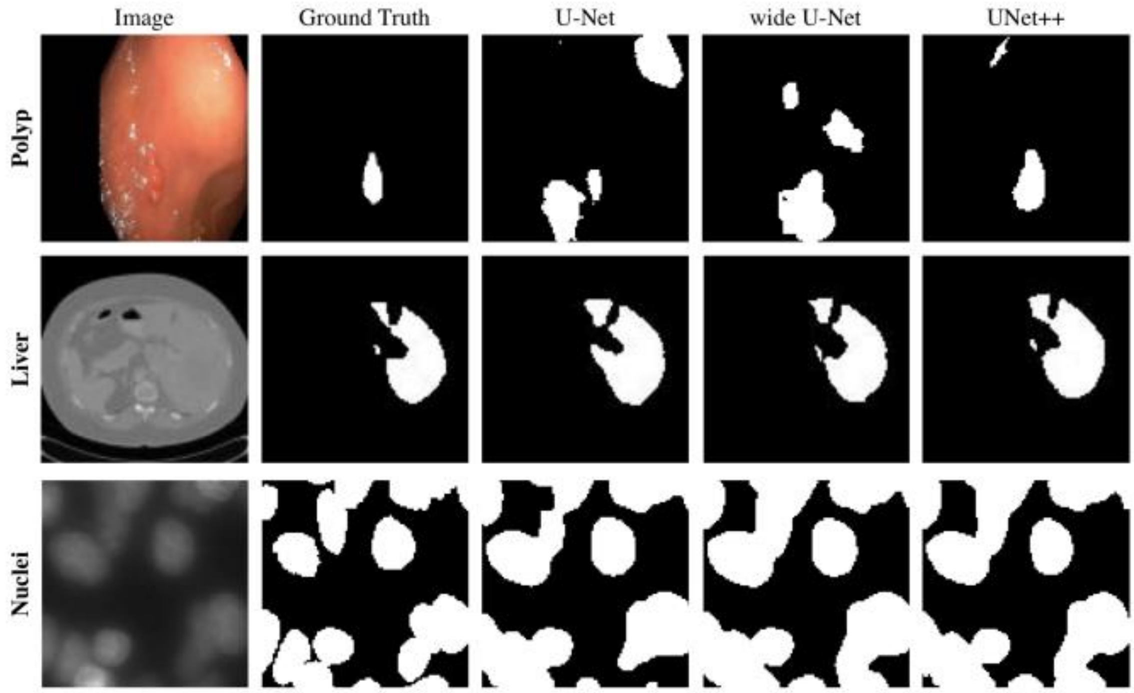
6.4 Future work:

The model has performed well in the trainset and better in the test data. The use of VGG16 and Resnet has been tried extensively in the encoder block of U-Net architecture. The future research can be tried with Inception v3 network for transfer learning in encoder layer.

Since the size of abnormalities are varying the experimenting with fixed kernel size is not achieving good results. The larger kernel performs well when the features are distributed across the large image. In case of mammogram the focus is on the breast region the smaller kernels are good at detecting the area specific features in a scanned image. This may also help in capturing salient features in varying levels.

Avoiding unnecessary areas while training model is equally important while focusing on the abnormalities in a mammogram image. The attention U-net achieves it using the attention gate(Siddique et al., 2021). On a repeated use this attention gate improves the performance of segmentation significantly with a simple model compared to U-Net. The U-Net can be replaced with attention U-Net and the same encoder with transfer learning can be tried for better results.

Using a fixed filter size could be a boon or a ban as tuning a correct filter size is cumbersome. The fixed filter size works perfect the features are of similar size. In case of mammogram the tumors could be of varying size. The use of deeper networks could partially solve it a cost of computation expensive. The Inception network provides a solution of using varying filter size on the same network. This has yielded a good result in Brain tumor segmentation for BraTS 2018 challenge(Chen et al., 2019).

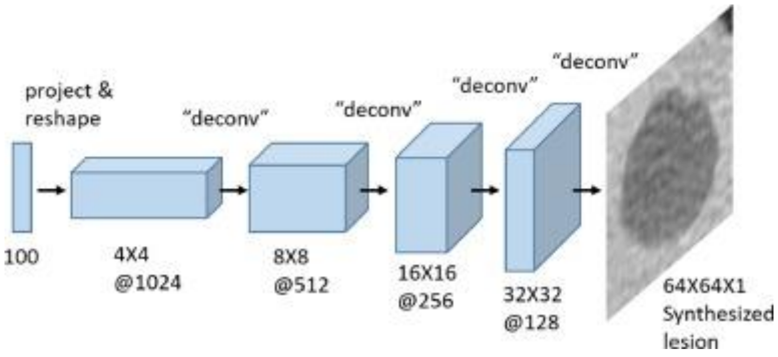


6.1 U-Net++ (Zhou et al., 2018)

U-net++ is an advanced version of U-net architecture inspired from Densenet. The more semantic information extracted by skip connections an intermediary grid between the contracting and expansive paths(Zhou et al., 2018). The U-net which is used for current thesis has the features contracting path are directly concatenated onto the corresponding layers in the expansive path. With U-net++ the skip connection unit receives all the feature maps from all previous layers along with the up sampled feature map from its immediate lower units. The U-net++ architecture can be replaced for U-net and the results can be improved from the current work.

The challenge with the current proposed model from this thesis is the limited availability of images for training. There were only 1231 images available for training and the option to generate images through augmentation is also limited as in reality the actual mammograms comes only in specific directions and color. If there are new images published by (CBIS-DDSM, n.d.) in future they can added to this training set and trained. There is a good possibility as the model can learn new features and generalize well.

The lack of sufficient data is a common challenge in the medical imaging field as the annotations must be performed by the experts, in mammogram it is the radiologist. This is a time consuming and costly process. A generative adversial network (GAN) can help in generating data with annotation like the training data.



6.2 GAN (Frid-Adar et al., 2018)

The classic GAN consist of 2 blocks, the generator and discriminator. The generator uses the CNN to generate images and corresponding masks whereas the discriminator uses the CNN to classify real images or masks from generated images. Same as NN it takes iterations to get photorealistic images. Additionally, the generated images are used for synthetic data augmentation and improve the performance for classification and segmentation(Frid-Adar et al., 2018) . This can easily solve the overfitting problem, provided the generated images are class imbalance free.

References

Abdelhafiz, D., Bi, J., Ammar, R., Yang, C. and Nabavi, S., (2020) Convolutional neural network for automated mass segmentation in mammography. *BMC Bioinformatics*, [online] 211, p.192. Available at: https://doi.org/10.1186/s12859-020-3521-y.

Agarwal, R., Diaz, O., Lladó, X., Yap, M.H. and Martí, R., (2019a) Automatic mass detection in mammograms using deep convolutional neural networks. *Journal of Medical Imaging*, [online] 63, pp.1–9. Available at: https://doi.org/10.1117/1.JMI.6.3.031409.

Agarwal, R., Diaz, O., Lladó, X., Yap, M.H. and Martí, R., (2019b) Automatic mass detection in mammograms using deep convolutional neural networks. *Journal of Medical Imaging*, [online] 63, pp.1–9. Available at: https://doi.org/10.1117/1.JMI.6.3.031409.

Ahmed, L., Iqbal, M.M., Aldabbas, H., Khalid, S., Saleem, Y. and Saeed, S., (2020) Images data practices for Semantic Segmentation of Breast Cancer using Deep Neural Network. *Journal of Ambient Intelligence and Humanized Computing*. [online] Available at: https://doi.org/10.1007/s12652-020-01680-1.

An, J., Yu, H., Bai, R., Li, J., Wang, Y. and Cao, R., (2022) Detection and Segmentation of Breast Masses Based on Multi-Layer Feature Fusion. *Methods*, [online] 202, pp.54–61. Available at: https://www.sciencedirect.com/science/article/pii/S1046202321001146.

Anon (n.d.) Breast cancer in India. *https://www.medanta.org/patient-education-blog/breast-cancer-statistics-in-india/*.

Anon (n.d.) Breast ultrasound. *https://ecancer.org/en/journal/article/1062-automatic-breast-ultrasound-state-of-the-art-and-future-perspectives*.

Anon (n.d.) *CAD-Wiki*.

Anon (n.d.) Cancer Facts & Figures 2020. *https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html*.

Anon (n.d.) *CBIS-DDSM*. https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM#2251662997944ee0e4a54b15a8479efafa2064dd.

Anon (n.d.) *Image segmentation*. https://en.wikipedia.org/wiki/Image_segmentation.

Anon (n.d.) procite.

Anon (n.d.) Stationary digital breast tomosynthesis increases diagnostic accuracy. *https://physicsworld.com/a/stationary-digital-breast-tomosynthesis-increases-diagnostic-accuracy/*.

Anon (n.d.) *Why is early diagnosis important?* https://www.cancerresearchuk.org/about-cancer/cancer-symptoms/why-is-early-diagnosis-important.

Azizpour, H., Razavian, A.S., Sullivan, J., Maki, A. and Carlsson, S., (2015) From generic to specific deep representations for visual recognition. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. pp.36–45.

Becker, A.S., Marcon, M., Ghafoor, S., Wurnig, M.C., Frauenfelder, T. and Boss, A., (2017) Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. *Investigative Radiology*, [online] 527. Available at: https://journals.lww.com/investigativeradiology/Fulltext/2017/07000/Deep_Learning_in_Mammography__Diagnostic_Accuracy.7.aspx.

Can you find the Tumour, (n.d.) *https://towardsdatascience.com/can-you-find-the-breast-tumours-part-1-of-3-1473ba685036*.

Chen, J., Chen, L., Wang, S. and Chen, P., (2020) A Novel Multi-Scale Adversarial Networks for Precise Segmentation of X-Ray Breast Mass. *IEEE Access*, 8, pp.103772–103781.

Chen, W., Liu, B., Peng, S., Sun, J. and Qiao, X., (2019) S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation. In: A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes and T. van Walsum, eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing, pp.358–368.

Chougrad, H., Zouaki, H. and Alheyane, O., (2018) Deep Convolutional Neural Networks for breast cancer screening. *Computer Methods and Programs in Biomedicine*, [online] 157, pp.19–30. Available at: https://www.sciencedirect.com/science/article/pii/S0169260717301451.

Dhull, A., Khanna, K., Singh, A. and Gupta, G., (2018) ACO Inspired Computer-aided Detection/Diagnosis (CADe/CADx) Model for Medical Data Classification. *Recent Patents on Computer Science*, 12.

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

Dice and Dice loss, (n.d.) *https://pycad.co/the-difference-between-dice-and-dice-loss/* .

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S. and Pal, C., (2016) The Importance of Skip Connections in Biomedical Image Segmentation. In: 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., *Deep Learning and Data Labeling for Medical Applications*. Cham: Springer International Publishing, pp.179–187.

Ellenbogen, P.H., Muroff, L.R. and Pressman, B.D., (2009) Challenges Facing the Mature Radiologist: A Guide. *Journal of the American College of Radiology*, [online] 66, pp.428–433. Available at: https://www.sciencedirect.com/science/article/pii/S1546144008006376.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H., (2018) GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, [online] 321, pp.321–331. Available at: https://www.sciencedirect.com/science/article/pii/S0925231218310749.

Gardezi, S.J.S., Elazab, A., Lei, B. and Wang, T., (2019) Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. *Journal of medical Internet research*, [online] 217, pp.e14464–e14464. Available at: https://pubmed.ncbi.nlm.nih.gov/31350843.

Geiser, W.R., Haygood, T.M., Santiago, L., Stephens, T., Thames, D. and Whitman, G.J., (2011) Challenges in Mammography: Part 1, Artifacts in Digital Mammography. *American Journal of Roentgenology*, [online] 1976, pp.W1023–W1030. Available at: https://doi.org/10.2214/AJR.10.7246.

Giger, M.L., Karssemeijer, N. and Schnabel, J.A., (2013) Breast Image Analysis for Risk Assessment, Detection, Diagnosis, and Treatment of Cancer. *Annual Review of Biomedical Engineering*, [online] 151, pp.327–357. Available at: https://doi.org/10.1146/annurev-bioeng-071812-152416.

Gur, D., Abrams, G.S., Chough, D.M., Ganott, M.A., Hakim, C.M., Perrin, R.L., Rathfon, G.Y., Sumkin, J.H., Zuley, M.L. and Bandos, A.I., (2009) Digital Breast Tomosynthesis: Observer Performance Study. *American Journal of Roentgenology*, [online] 1932, pp.586–591. Available at: https://doi.org/10.2214/AJR.08.2031.

Hoogi, A., Subramaniam, A., Veerapaneni, R. and Rubin, D.L., (2017) Adaptive Estimation of Active Contour Parameters Using Convolutional Neural Networks and Texture Analysis. *IEEE Transactions on Medical Imaging*, 363, pp.781–791.

Hooley, R.J., Durand, M.A. and Philpotts, L.E., (2016) Advances in Digital Breast Tomosynthesis. *American Journal of Roentgenology*, [online] 2082, pp.256–266. Available at: https://doi.org/10.2214/AJR.16.17127.

Jegou, S., Drozdzal, M., Vázquez, D., Romero, A. and Bengio, Y., (2017) *The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation*.

Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A. and Karssemeijer, N., (2017) Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis*, [online] 35, pp.303–312. Available at: https://www.sciencedirect.com/science/article/pii/S1361841516301244.

Lbachir, I.A., Daoudi, I. and Tallal, S., (2021) Automatic computer-aided diagnosis system for mass detection and classification in mammography. *Multimedia Tools and Applications*, [online] 806, pp.9493–9525. Available at: https://doi.org/10.1007/s11042-020-09991-3.

Lotter, W., Sorensen, G. and Cox, D., (2017) A Multi-Scale CNN and Curriculum Learning Strategy for Mammogram Classification. [online] Available at: https://arxiv.org/abs/1707.06978 [Accessed 16 Jun. 2022].

Oliver, A., Freixenet, J., Martí, J., Pérez, E., Pont, J., Denton, E.R.E. and Zwiggelaar, R., (2010) A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis*, [online] 142, pp.87–110. Available at: https://www.sciencedirect.com/science/article/pii/S1361841509001492.

Padilla, R., Netto, S.L. and Silva, E.A.B. da, (2020) A Survey on Performance Metrics for Object-Detection Algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. pp.237–242.

Ribli, D., Horváth, A., Unger, Z., Pollner, P. and Csabai, I., (2018) Detecting and classifying lesions in mammograms with Deep Learning. *Scientific Reports*, [online] 81, p.4165. Available at: https://doi.org/10.1038/s41598-018-22437-z.

Ronneberger, O., Fischer, P. and Brox, T., (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N. Navab, J. Hornegger, W.M. Wells and A.F. Frangi, eds., *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, pp.234–241.

Rose, C., Turi, D., Williams, A., Wolstencroft, K. and Taylor, C., (2006) *Web Services for the DDSM and Digital Mammography Research*.

Salama, W.M. and Aly, M.H., (2021) Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alexandria Engineering Journal*, [online] 605, pp.4701–4709. Available at: https://www.sciencedirect.com/science/article/pii/S1110016821002027.

Sathyan, A., Martis, D. and Cohen, K., (2020a) Mass and Calcification Detection from Digital Mammograms Using UNets. In: *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. pp.229–232.

Sathyan, A., Martis, D. and Cohen, K., (2020b) Mass and Calcification Detection from Digital Mammograms Using UNets. In: *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. pp.229–232.

Siddique, N., Paheding, S., Elkin, C.P. and Devabhaktuni, V., (2021) U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access*, 9, pp.82031–82057.

Sood, R., Rositch, A.F., Shakoor, D., Ambinder, E., Pool, K.-L., Pollack, E., Mollura, D.J., Mullen, L.A. and Harvey, S.C., (2019) Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis. *Journal of Global Oncology*, [online] 5, pp.1–17. Available at: https://doi.org/10.1200/JGO.19.00127.

Stathonikos, N., Veta, M., Huisman, A. and van Diest, P.J., (2013) Going fully digital: Perspective of a Dutch academic pathology lab. *Journal of Pathology Informatics*, [online] 41, p.15. Available at: https://www.sciencedirect.com/science/article/pii/S2153353922006368.

Su, Y., Liu, Q., Xie, W. and Hu, P., (2022) YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms. *Computer Methods and Programs in Biomedicine*, [online] 221, p.106903. Available at: https://www.sciencedirect.com/science/article/pii/S0169260722002851.

Toprak, A., Unay, D., Cerit, T. and Üğüdücü, K., (2022) *Cancerous Lesion Segmentation for Early Detection of Breast Cancer by using CNN*.

Tsochatzidis, L., Costaridou, L. and Pratikakis, I., (2019) Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. *Journal of Imaging*, [online] 53. Available at: https://www.mdpi.com/2313-433X/5/3/37.

Wang, X., Liang, G., Zhang, Y., Blanton, H., Bessinger, Z. and Jacobs, N., (2020) Inconsistent Performance of Deep Learning Models on Mammogram Classification. *Journal of the American College of Radiology*, [online] 176, pp.796–803. Available at: https://www.sciencedirect.com/science/article/pii/S1546144020300284.

Wong, S.C., Gatt, A., Stamatescu, V. and McDonnell, M.D., (2016) Understanding Data Augmentation for Classification: When to Warp? In: *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. pp.1–6.

Xu, C., Qi, Y., Wang, Y., Lou, M., Pi, J. and Ma, Y., (2022) ARF-Net: An Adaptive Receptive Field Network for breast mass segmentation in whole mammograms and ultrasound images. *Biomedical Signal Processing and Control*, [online] 71, p.103178. Available at: https://www.sciencedirect.com/science/article/pii/S1746809421007758.

Yaker, A.H., Bouzar-Benlabiod, L., Menendez, A. and Paillet, G., (2021) Mammograms Classification with NeuroMem® Chip for Breast Cancer Detection. In: *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*. pp.236–239.

Yu, H., Bai, R., An, J. and Cao, R., (2020) Deep Learning-Based Fully Automated Detection and Segmentation of Breast Mass. In: *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. pp.293–298.

Zhao, J., Chen, T. and Cai, B., (2021) A computer-aided diagnostic system for mammograms based on YOLOv3. *Multimedia Tools and Applications*. [online] Available at: https://doi.org/10.1007/s11042-021-10505-y.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J., (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J.M.R.S. Tavares, A. Bradley, J.P. Papa, V. Belagiannis, J.C.

Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan and A. Madabhushi, eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer International Publishing, pp.3–11.

Appendix A: Research Proposal

Image Segmentation of digital Mammograms to identify abnormalities for breast cancer using Deep learning

P. Padmanaban

Research Proposal

MAY 2022

1. Abstract

Breast cancer is the most commonly occurring cancer type followed by cervical cancer. The trend is increasing year on year in both urban and rural India. The diagnosis of the cancer cells even at a smaller size and during pre-growing stage will make the treatment effective and saves the patient life. The existing models has scope improvement in detection accuracy and other improvement opportunities. The publicly available CBIS-DDSM dataset will be used for training the model. The dataset is more reliable as the suspicious region has been marked by trained mammographers. The image pre-processing will be handled appropriately to reduce the false positive results. U-net will be used to generate the mask on the suspicious region to segment mass. The end model can be used as an additional analysis to support radiologist to identify hard to detect tumours.

Table of content

1. Background……………………………………………………………………………5
2. Problem statement and Related works………………………………………………..6
3. Research questions…………………………………………………………………8
4. Aim and objectives…………………………………………………………………8
5. Significance of the study……………………………………………………………8
6. Scope of the study…………………………………………………………………..8
7. Research Methodology……………………………………………………………..9
8. Required resources………………………………………………………………....11
9. Research plan………………………………………………………………………12
10. References…………………………………………………………………………13

LIST OF FIGURES

Figure 7.1 U-Net Architecture 10
Figure 7.2 IOU metrics image 10
Figure 9.1 Gantt chart 12

2. Background

With constant lifestyle changes and the environmental exposures have started creating severe impact in mankind. Breast cancer has become most frequent tumour occurring in females with an increasing on number of cases every year. As there is no clear reason for the cause of breast cancer, the common suggestion by doctors across the globe is to make an early detection for the signs of abnormalities. There are clear evidence with the increase in survival rate when spotted early(Why is early diagnosis important?, n.d.).

The general process is to look for any abnormalities and if any suspicious found the patients are suggested for further tests. The process is currently "invasive, time-consuming and create unnecessary anxiety for patients if the results show that the tumour is benign. "(Yaker et al., 2021).

The amount of healthcare data generated for screening purpose has been increasing. The mammography is the key technique for breast cancer detection in identifying the anomalies. The results of the mammograms are currently evaluated by the Radiologist with a chance of misdiagnosis(Yu et al., 2020). The traditional algorithms have several short comings like will be unable to detect small masses, the accuracy is not as expected etc.

For Computer aided design the deep learning is gaining momentum and has started creating impact on solving several healthcare problems specifically using computer vision technique. This study is focused on creating an additional assessment to what radiologist do today. Improve the identification of mass segmentation and reduce the false positive rate.

There are number of studies/research performed in classification and segmentation of mammograms. They all target to solve different problems to identify small mass, reduce false positive, quicker detection time. Various algorithms applied yielding to different results. In this research the primary focus will be identify small mass segmentation, reduce false positive and improve the metrics. This is aimed to help with the early diagnosis and increase the life span of patients.

3 Problem Statement

The early detection of any breast mass will provide the doctors a high chance of saving the patients. The screening interval depends purely on age factor of the people and there are recommendations from American cancer society to undergo screening every 1-2 years as per guidelines(Smith et al., 2018).There are certain types of breast mass and micro calcification which are an indicator of cancer development.

The radiologist identifies the abnormalities in the form of mass or calcification in the mammogram images. "Artifacts reduce the quality of mammograms and may mimic or obscure abnormalities and cause interpretation errors" (Geiser et al., 2011)for radiologist. There are also other factors related to newly trained radiologist or the amount of workload faced by them may affect the evaluation of the mammograms(Ellenbogen et al., 2009)

The automated detection of these abnormalities could assist the radiologist in the diagnosis. There are around 10% of miss which looks normal though they are diagnosed with breast cancer at later stage(Seely and Alhassan, 2018) . These second opinion could reduce the number of false positives.

There has been good number of attempts in the past for an Automated mass and calcification detection from mammograms. The attempts are generally of 2 categories

- Classification the type of cancer i.e., benign and malignant
- Segment the mass and calcification abnormalities

There are few public datasets available annotated by trained professional which are specialized in detecting the mass and calcification abnormalities. With the recent development in the field of deep learning architecture they have yielded better results than the traditional computer aided designs.

Annop et al has proposed a solution (Sathyan et al., 2020) using U-Net segmentation architecture to segment the boundaries of mass and the calcifications present in the mammograms. There are separate architectures for mass and the calcification with its own dataset The results are promising as most of the abnormal regions are correctly identified with few false positive and false negatives.

For radiologist sometime the small-scale masses are also a challenge and the miss could lead to serious consequences. Hui Yu et al has proposed a solution(Yu et al., 2020) based on model Dense Mask R-CNN suitable for capturing the low level features leading to the detection of the small scale breast masses. The results have shown an improvement of the original Mask R-CNN with a shortcoming of redundant information generated.

The unbalance class is an obstacle in achieving a good segmentation result. Juan chen et al has performed experiments(Chen et al., 2020) in solving this using multi-scale adversarial network for breast mass segmentation. Again the U-Net has been the choice for the segmentation part achieving satisfactory results which comes with a cost of additional training time.

Though High false positive rates are not a concern as the mammograms are again evaluated by the radiologist, Lugman Ahmed et al has proposed a solution focusing on the preprocessing(Ahmed et al., 2020) with the complete process to extract the patten. The helps to remove the noise in the data and improves the performances.

Ilhame Ait Lbachir et al has proposed a fully integrated CAD systems to mass detect and classify the mammograms(Lbachir et al., 2021). The images are pre-processed, segmented based on HRAK algorithm , false positives are reduced and finally classification is done through SVM. Though this has achieved grate results there needs to be improvement on differentiating the spiculated masses with normal tissues.

Jihahui Zhao et al also proposes a solution of both segmentation and classification using YOLOv3 and Transfer learning(Zhao et al., 2021). There are 3 models i.e., general, mass and microcalcification are create and they all perform good at detecting the lesions. The only caveat is that it performs poor on low configured laptop.

Dina Abdelhafiz et al proposes U-Net model to segment Mass lesions in mammogram images by extracting both low level and high-level features(Abdelhafiz et al., 2020). This vanilla U-Net model has superior performance compared to existing models such as SegNet, Fast R-CNN e.t.c but with scope for improvement in the segmentation.

Though there are several studies performed for both classification and segmentation of mammogram images, there are still open questions on improving the segmentation performance, reducing time on running the models and inconsistent results on unseen data (Wang et al., 2020)

3. Research Question

How to detect mass abnormalities in Mammogram scans using image segmentation technique

4. Aim and Objectives

The main aim of this research is to propose a model to diagnose the early detection of breast cancer from Mammogram images. The goal of this research is to improve the survival rate of the cancer patients through most effective and reliable methods.

- To investigate the mammogram images for presence of abnormalities in the form of mass
- To determine the optimum technique to reduce false positive in the mammogram images
- To propose a model which outputs a predicted mask from the mammogram if there are any abnormalities
- To evaluate the performance of the proposed semantic segmentation

5. Significance of the study

Till date there are no concrete evidence for the occurrence of cancer cells(Sathyan et al., 2020). Early diagnosis of breast cancer is the key factor in saving the life of the affected patients. Identifying the cells when they are smaller and reducing the false positive is also key challenge to Radiologist.

This study primarily aims to reduce false positive and improve the sensitivity from the existing work(Sathyan et al., 2020). The U-Net architecture will be used to identify the mass abnormalities in the mammograms. The improvement in accuracy will encourage radiologist to spend less time on more obvious case. The reduction in false positive will reduce the panic among the patients and will reduce their subsequent visits.

6. Scope of the study

The mammogram images are analysed for semantic segmentation and the interest area for mass abnormalities. The classification of tumour presence or not in the scope. Also, the calcification abnormalities are not considered. This research focuses on preparing model and not on the deployment part.

7. Research Methodology

7.1 Image segmentation:

To detect the mass abnormalities the radiologist investigates image for the regions containing the cancer cells. With neural networks we can use semantic segmentation to extract the suspicious region using segmentation technique. The primary purpose of it is to segregate the image into pixels, classify the pixels into unique categories and cluster the similar coherent pixels(Image segmentation, n.d.).

7.2 Dataset:

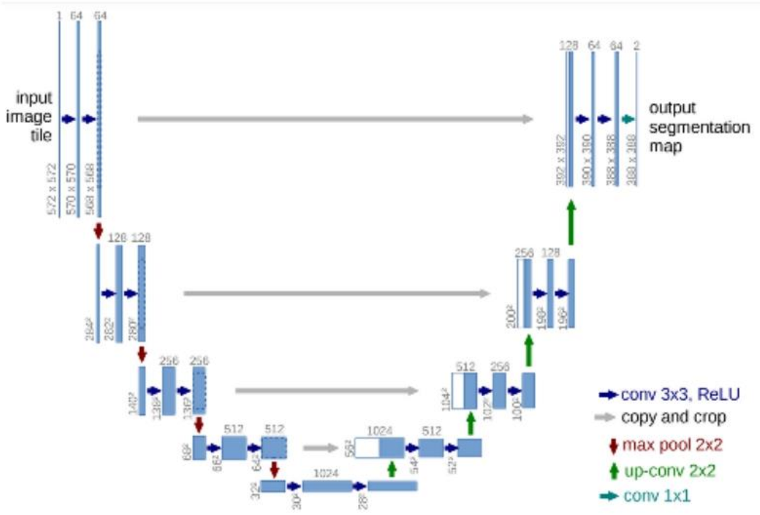
The dataset used for this task is CBIS-DDSM. It contains 2620 real world scanned mammogram images(CBIS-DDSM, n.d.) in Dicom format. The region of interest annotation for abnormalities are selected, curated, and verified by trained mammographers.

7.3 Data Pre-processing:

The mammogram images contain white strips which creates a confusion among the actual cancer cell during segmentation of the cancer cells (Ahmed et al., 2020). The borders are cropped, normalisation done. There are only 1592 contains mass abnormalities out of which 71 contains more than 1 abnormality. Hence data augmentation applied to increase the variability with the existing images.

7.4 Model:

U-Net is an unique popular CNN architecture created for Biomedical image segmentation tasks created by olaf Ronneberger et al. in 2015. It has an end-to-end encoder-decoder network for semantic segmentation with unique Up-Down architecture with Contracting path and expansive path. It takes an input and outputs a segmented image

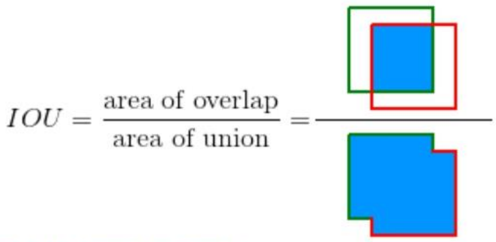


7.1 U-Net architecture (Ronneberger et al., 2015)

Transfer learning can be applied to utilize the power of exiting models. Various experiments are performed to identify the hyperparameters such as no of epochs, learning rate , optimizer, Loss function e.t.c

7.5 **Evaluation:**

The results of a data augmentation are generally evaluated through IOU (Intersection over union) metrics. The metrics is more about evaluating the overlap of predicted with the true boundary, a measure of overlap. Higher the IoU has better result (preferably more than 0.5) vs lower IoU indicates poor results.



7.2 IOU metrics image (Padilla et al., 2020)

8. Required Resources

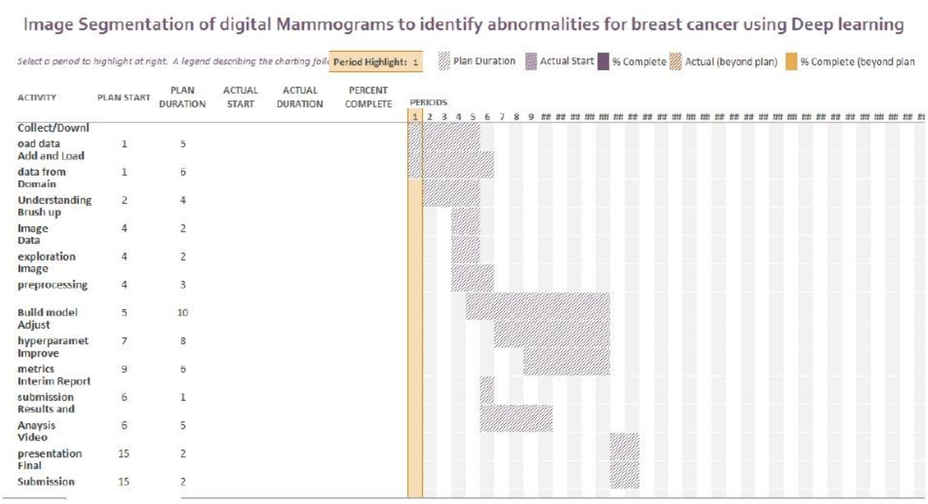
8.1 Software:

- NBIA Data Retriever
- PyTorch
- U-Net
- Azure ML Studio

8.2 Hardware:

- Azure ML Studio
- NVIDIA Tesla K80

9 Research Plan



9.1 Gantt chart

10 References

Abdelhafiz, D., Bi, J., Ammar, R., Yang, C. and Nabavi, S., (2020) Convolutional neural network for automated mass segmentation in mammography. *BMC Bioinformatics*, [online] 211, p.192. Available at: https://doi.org/10.1186/s12859-020-3521-y.

Ahmed, L., Iqbal, M.M., Aldabbas, H., Khalid, S., Saleem, Y. and Saeed, S., (2020) Images data practices for Semantic Segmentation of Breast Cancer using Deep Neural Network. *Journal of Ambient Intelligence and Humanized Computing*. [online] Available at: https://doi.org/10.1007/s12652-020-01680-1.

Anon (n.d.) *CBIS-DDSM*. https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM#2251662997944ee0e4a54b15a8479efafa2064dd.

Anon (n.d.) *Image segmentation*. https://en.wikipedia.org/wiki/Image_segmentation.

Anon (n.d.) *Why is early diagnosis important?* https://www.cancerresearchuk.org/about-cancer/cancer-symptoms/why-is-early-diagnosis-important.

Chen, J., Chen, L., Wang, S. and Chen, P., (2020) A Novel Multi-Scale Adversarial Networks for Precise Segmentation of X-Ray Breast Mass. *IEEE Access*, 8, pp.103772–103781.

Ellenbogen, P.H., Muroff, L.R. and Pressman, B.D., (2009) Challenges Facing the Mature Radiologist: A Guide. *Journal of the American College of Radiology*, [online] 66, pp.428–433. Available at: https://www.sciencedirect.com/science/article/pii/S1546144008006376.

Geiser, W.R., Haygood, T.M., Santiago, L., Stephens, T., Thames, D. and Whitman, G.J., (2011) Challenges in Mammography: Part 1, Artifacts in Digital Mammography. *American Journal of Roentgenology*, [online] 1976, pp.W1023–W1030. Available at: https://doi.org/10.2214/AJR.10.7246.

Lbachir, I.A., Daoudi, I. and Tallal, S., (2021) Automatic computer-aided diagnosis system for mass detection and classification in mammography. *Multimedia Tools and Applications*, [online] 806, pp.9493–9525. Available at: https://doi.org/10.1007/s11042-020-09991-3.

Padilla, R., Netto, S.L. and Silva, E.A.B. da, (2020) A Survey on Performance Metrics for Object-Detection Algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. pp.237–242.

Ronneberger, O., Fischer, P. and Brox, T., (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N. Navab, J. Hornegger, W.M. Wells and A.F. Frangi, eds., *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, pp.234–241.

Sathyan, A., Martis, D. and Cohen, K., (2020) Mass and Calcification Detection from Digital Mammograms Using UNets. In: *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. pp.229–232.

Seely, J.M. and Alhassan, T., (2018) Screening for breast cancer in 2018-what should we be doing today? *Current oncology (Toronto, Ont.)*, [online] 25Suppl 1, pp.S115–S124. Available at: https://pubmed.ncbi.nlm.nih.gov/29910654.

Smith, R.A., Andrews, K.S., Brooks, D., Fedewa, S.A., Manassaram-Baptiste, D., Saslow, D., Brawley, O.W. and Wender, R.C., (2018) Cancer screening in the United States, 2018: A review of current

American Cancer Society guidelines and current issues in cancer screening. *CA: A Cancer Journal for Clinicians*, [online] 684, pp.297–316. Available at: https://doi.org/10.3322/caac.21446.

Wang, X., Liang, G., Zhang, Y., Blanton, H., Bessinger, Z. and Jacobs, N., (2020) Inconsistent Performance of Deep Learning Models on Mammogram Classification. *Journal of the American College of Radiology*, [online] 176, pp.796–803. Available at: https://www.sciencedirect.com/science/article/pii/S1546144020300284.

Yaker, A.H., Bouzar-Benlabiod, L., Menendez, A. and Paillet, G., (2021) Mammograms Classification with NeuroMem® Chip for Breast Cancer Detection. In: *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*. pp.236–239.

Yu, H., Bai, R., An, J. and Cao, R., (2020) Deep Learning-Based Fully Automated Detection and Segmentation of Breast Mass. In: *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. pp.293–298.

Zhao, J., Chen, T. and Cai, B., (2021) A computer-aided diagnostic system for mammograms based on YOLOv3. *Multimedia Tools and Applications*. [online] Available at: https://doi.org/10.1007/s11042-021-10505-y.

Appendix B: Code for Data pre-processing and Model building

The code and its description can be found in https://github.com/Padmanaabhah/CBIS-DDSM-Thesis/tree/master

Appendix C: Dataset

The dataset and its description can be found in
https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM