Breast Cancer Detection and Segmentation using Deep Learning

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Engineering

By

Anne Pinto	**53**
Minita Joshee	**23**
Amaan Nizam	**46**
Abhiram Pillai	**52**

Supervisor: **Dr. Satishkumar Chavan**

Department of Electronics and Telecommunication

Don Bosco Institute of Technology

University of Mumbai

2020 - 2021

Don Bosco Institute of Technology

(Affiliated to the University of Mumbai)
Premier Automobiles Road, Kurla, Mumbai - 400070)

CERTIFICATE

This is to certify that the project entitled **BREAST CANCER DETECTION AND SEGMENTATION USING DEEP LEARNING** is a bonafide work of

Anne Pinto	**53**
Minita Joshee	**23**
Amaan Nizam	**46**
Abhiram Pillai	**52**

submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Undergraduate"** in **"Bachelor of Engineering"**. This approval does not necessarily endorse or accept every statement made, opinion expressed or conclusion drawn as recorded in the report. It only signifies the acceptance of the report for the purpose for which it has been submitted.

Dr. Satishkumar Chavan
Project guide

Dr. Ashwini Kotrashetti
HOD,EXTC Department

Dr. Prasanna Nambiar
Principal

Don Bosco Institute of Technology

(Affiliated to the University of Mumbai)
Premier Automobiles Road, Kurla, Mumbai - 400070)

Project Report Approval for B.E.

This project report entitled **BREAST CANCER DETECTION AND SEGMENTATION USING DEEP LEARNING** by **Anne Pinto, Minita Joshee, Amaan Nizam and Abhiram Pillai** is approved for the degree of Bachelor of Engineering in Electronics and Telecommunication.

Examiners

1.

2.

Date: 29/05/2021
Place: **Kurla, Mumbai**

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Anne Pinto **Minita Joshee**

Amaan Nizam **Abhiram Pillai**

Date: 29/05/2021

ACKNOWLEDGEMENT

A project is a teamwork which involves the contribution of many people. We would like to thank everyone who have contributed by taking interest in our work and motivating us all the way through.

We wish to express our sincere gratitude to our Project Guide, Dr. Satishkumar Chavan, for his enthusiasm, patience, insightful comments, helpful information and unceasing ideas that have helped us tremendously at all times. His immense knowledge, profound experience and professional expertise has enabled us to complete our project successfully. Without his support and guidance, this project would not have been possible.

We would like to thank Dr. Prasanna Nambiar, Principal, Dr. Ashwini Kotrashetti, Head of Department, Dr. Sudhakar Mande, Project Mentor and Ms. Freda Carvalho, Project Co-ordinator for their continuous valuable guidance, support and their precious time in every possible way throughout our project activity.

We wish to express our special gratitude to Mr. Rishikesh Tambe for his guidance on U-Net Pipeline. Without his assistance we wouldn't have completed this project on time.

We also wish to express our deepest thanks to our parents and siblings for their unwavering support and encouragement throughout our project.

Date: 29/05/2021

ABSTRACT

One of the most prevalent cancers among women is breast cancer. Accurate diagnosis of breast cancer at an early stage can reduce the mortality associated with this disease. Diagnosis can be done with screening mammography. The main challenge of screening mammography is that it is associated with high risk of false positives and false negatives.

Moreover, the large number of mammographic images generated day by day has brought a huge workload on radiologists and also increased the rate of misinterpretation. The advancement in technology, usage of more and more deep learning models, and availability of huge datasets resulted into developing better algorithms and models for breast cancer detection and segmentation. However, the accuracy and robustness of system is not achieved upto the standard expected. Therefore, there is a need to develop an automated system for detection and segmentation that can significantly relieve the pressure on radiologists and improve the diagnosis accuracy.

In this work, we propose a CNN based breast cancer detection system which classify tumour into benign and malignant. Further, we also implement the segmentation of tumour present in Cranial-Caudal and Medial-lateral oblique (CC and MLO) views of the mammograms using U-Net. The datasets used in this work are DDSM and MIAS which are popular benchmark datasets which are freely available with large number of images, appropriate to develop a good CNN model. The comparative analysis of different classification models shows the highest accuracy of classification is achieved with GoogleNet model on the DDSM dataset and VGG16 on the MIAS dataset. Also, the segmentation algorithm effectively extracts the lesion with removal the unwanted area like pectoral muscle. The segmentation results are much promising and may be helpful to radiologists for deciding the effective therapy to manage the patient.

Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence

ANN Artificial Neural Network

CCN Convolutional Neural Network

CC Cranial Caudal

DDSM Digital Database for Screening Mammography

DL Deep Learning

ILSVRC ImageNet Large Scale Visual Recognition Challenge

MIAS Mammographic Image Analysis Society

MLO Mediolateral Oblique

MRI Magnetic Resonant Imaging

VGG Visual Geometry Group

Chapter 1

Introduction

Breast cancer has the second highest mortality rate in women next to lung cancer [1]. According to the global cancer statics, the number of new cases in 2018 was estimated to be 18,078,957 and deaths 9,555,027 (52.85%) globally. The cases of breast cancer amounts to 2,088,849 (11.55%) and the deaths is estimated to be 626,679 (6.56%). 60% of the deaths occur in low income developing countries like Ethiopia [2] [3] [4] [5]. If the cancer is detected early, it increases expectancy of survival rate/mortality of patient. Existence of breast cancer can be indicated through many abnormalities such as masses, micro-calcifications and areas of symmetry and distortion. Among these, masses are the most representative and common lesion type. . However, masses can be easily hidden by overlapping breast tissues, making it difficult to detect them. Masses can be of two types, namely, undetected and misindentified. An undetected mass is a false negative, which delays a patient diagnosis until the next screening. A misidentified mass is a false positive, which leads to additional tests including re-screening and biopsy, causing unnecessary anxiety and pain to patients [6]. Many mammographic density ratings, ranging from manual categorical (e.g. BI-RADS) to automatic continuous scores, have been suggested. Radiologists characterized the mammographic appearance in the early years by a series of intuitive yet poorly defined patterns of breast tissue that have been shown to contribute to the risk of breast cancer. As obtained by Cumulus-like thresholding, the present gold standard is semi-automated continuous ratings.The strength threshold is set by the radiologist in Cumulus to distinguish radio dense (white appearing) from fatty (dark appearing) tissue. The computer then calculates the percentage of dense to total breast area known as the mammographic density percentage (PMD). User-assisted thresholding, however is subjective and time consuming, and is therefore not appropriate for high epidemiological thresholds. Moreover, doctors experience tiredness and fatigue going through 50-60 mammograms every day. A trend has been towards completely automating PMD scoring, but most of these methods depend on handcrafted features with many parameters that needs to be monitored. Hence, there arises the need for a more robust, fast, accurate, and efficient noninvasive cancer detection system. an automated

system is required for achieving error-free detection of breast cancer using mammogram.

1.1 Breast Anatomy

As we know human beings are mammals and every mammal has a mammary gland, breasts are supposed to be the mammary glands in humans. They are specialized organs on the anterior of the chest and are found to be more developed in females than in males because primary function in females is lactation. They produce milk for the infant.

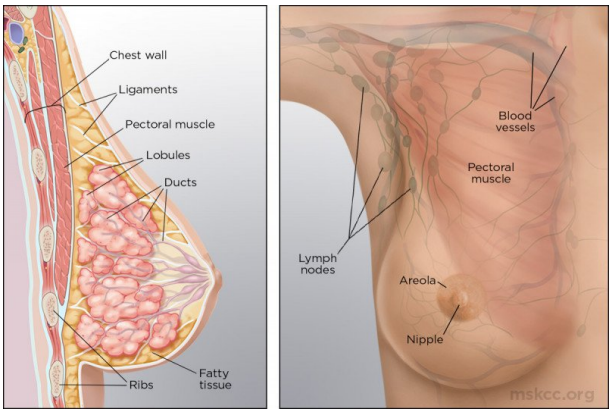


Figure 1.1: Structural anatomy of the breast

As shown in Figure 1.1, breasts are mainly made up of different kinds of tissues, glands, ducts and lymph nodes. Fatty, Connective (Glandular) or Fibrous and Glandular tissues consist of lobules and ducts. Fibrous tissues include Ligaments and scar tissues. Fatty tissues fills gap between glandular and fibrous tissue. Milk glands (lobules) that produce and supply milk. They are the sack like structures. Gland that produces mil there are 15-20 glands present which are known as lobes each of which has smaller lobules and they are arranged in clusters. Special ducts that transfer milk from the lobules to the nipple. These are thin tubes that carry milk from the lobes to the nipple and are located in the middle of the areola. Breast cancers can form in lobes and ducts. Fat fills space between lobes and ducts. The nipple is the centre of the areola. The nipple is surrounded by lobes in a radial manner.Areola (pink/brown pigmented region surrounding the nipple) is the area surrounding the nipple. Lymph nodes and Lymph are small bean shaped organs that help fight infection. A colourless fluid called lymph is produced by them which contain WBCs (lymphocytes).

1.2 Breast Cancer

The most common medical conditions found in breasts is cancer. As depicted in Figure 1.2, breast cancer can be of two types i.e. Malignant (Cancerous) and Benign (Non-Cancerous) conditions. Cancerous medical conditions can include formation of lump, discharge from nipple pitting, change in shape etc. Benign conditions include Fibrosis, Simple cyst formation. Cancers can also be differentiated as Ductal or Lobular. Further the cancers can be classified into In situ or Invasive Carcinoma and Proliferative or Non-proliferative.

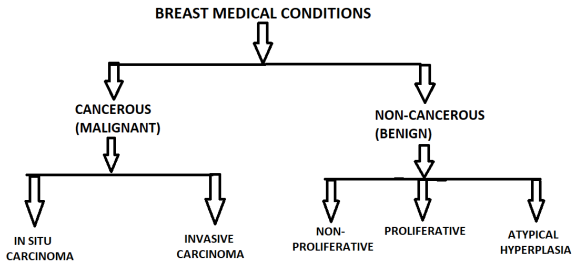


Figure 1.2: Classification of Breast Cancer

There are two main types of breast cancer namely: Ductal and Lobular. Figure 1.3 represents the two types of cancers and where exactly does the cancer form.

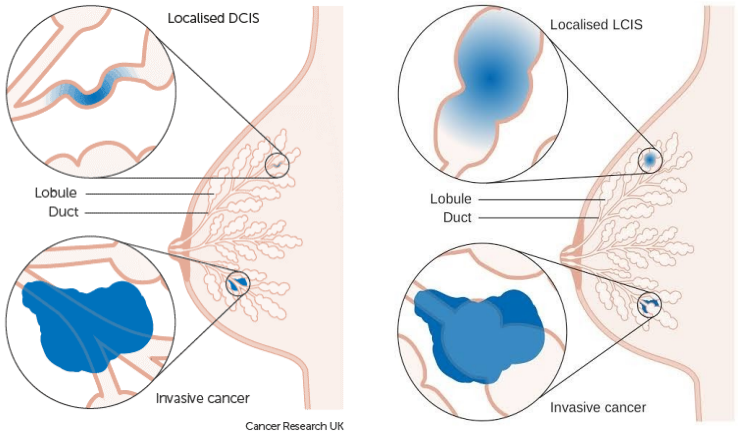


Figure 1.3: Ductal and Lobular Carcinoma In-situ

Ductal Carcinoma is a common type of breast cancer that starts in cells that

line the mil ducts, which carry milk to the nipple. It is divided into 2 types:

Ductal Carcinoma In-situ also known as intraductal carcinoma accounts for one of every five new breast cancer diagnoses. It's an uncontrolled growth of cells within the breast. It is non-invasive meaning it hasn't grown into the breast tissue outside of the ducts. The phrase in-situ means "it's in the original place". It's a stage 0 breast cancer. Invasive Ductal Carcinoma accounts for about 80% of all invasive breast cancers in women and 90% in men. It begins in the cells of the milk duct, then it grows into the duct walls and into the surrounding breast tissue. It can spread to the other parts of the body as well.

Lobular Carcinoma is another type of breast cancer that is commonly observed in women. It is also divided in 2 types:

Lobular Carcinoma In-situ isn't cancer, but it is thought to be a sign that breast cancer may develop. With this condition, there are abnormal cells in multiple lobules of the breasts, but these cells rarely spread to other parts of breasts. Even though this condition doesn't spread, it's important to keep an eye on it. Between 20% to 40% of women with this condition will develop a separate invasive breast cancer that will grow outside its original location within the next years. Invasive Lobular Carcinoma is the second most common form of invasive breast cancer. It begins in one of the breast lobules, then spreads to other parts of the breast. It is most likely to be found in both breasts. It can spread to other areas of the body.

Breast cancer can be symptomatic or asymptomatic. Symptoms for the most common breast cancer include a breast lump or tissue thickening, breast pain, swelling in most part of breast, changes in appearance of the skin near the breast. There are cases where no symptoms can be presented but the only way to detect is through a mammogram or an ultrasound.

To locate the region of abnormality, breasts are divided into four quadrants as shown in figure 1.4. Vertical quadrants are divided as inner and outer portions whereas horizontal quadrants are divided as upper and lower portions. Physicians sometimes describe the location of a breast tumour.

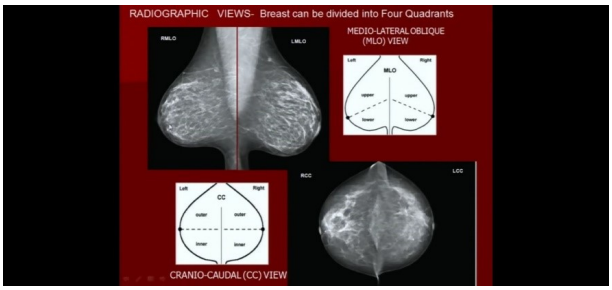


Figure 1.4: Quadrants of the breast

There are various screening techniques available for the detection of breast cancer. The most common being mammography. To detect early stage breast cancer, the individual patient's risk profile can serve as a good guideline for the

entire screening process. During mammography, the key factor is the position of the breast. There are two most common views which are CC and MLO as depicted in Figure 1.5 and Figure 1.6.

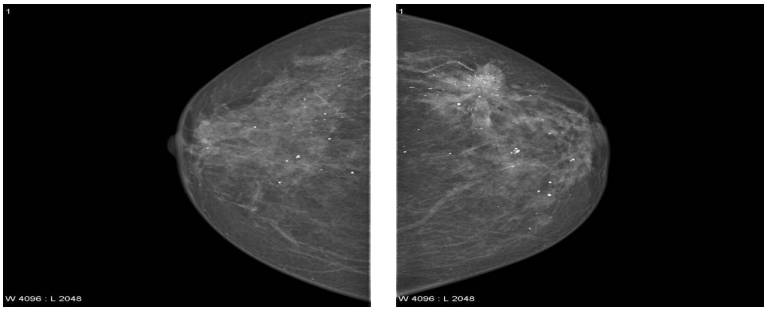


Figure 1.5: CC view in a mammogram

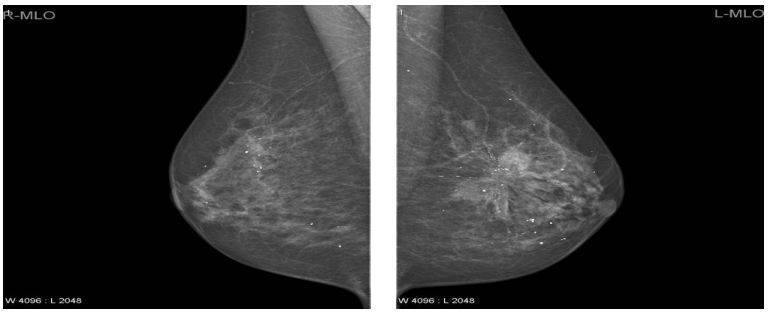


Figure 1.6: MLO view in a mammogram

Though the most common, mammography has its limitations which include low sensitivity in dense breasts. For this reason, ultrasound or MRI can be valuable adjunctive in clinical examinations. A combination of mammography, ultrasound, and MRI has shown to give very good results in detecting ductal carcinoma in situ and invasive cancer. Their combined use increases the likelihood of detecting breast cancer, and reduces the rates of false detections. However, further improvements of these medical imaging techniques are needed in order to overcome current limitations and to increase their effectiveness in detecting breast cancer.

1.3 Breast Cancer Statistics

According to global cancer statistics 2020, GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 180 countries [3]. Figure shows the distribution of incidence and deaths in both sexes due to various cancers.

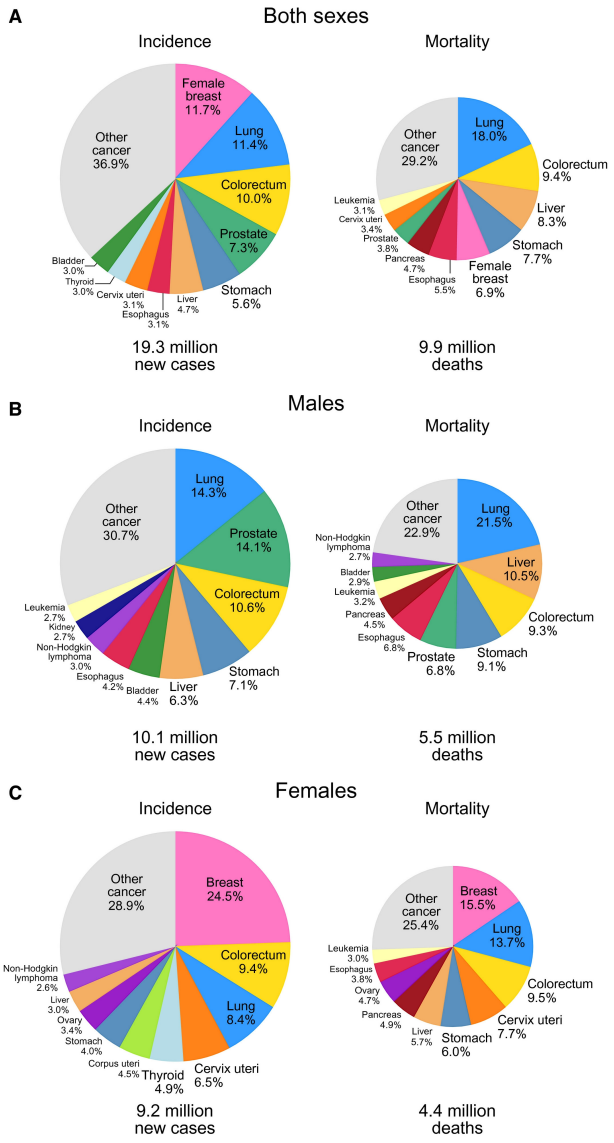


Figure 1.7: Global Cancer statistics

In the year 2020, Breast cancer among females has overshadowed lung cancer as the most commonly found cancer with a total estimate of 2.3 million cases i.e 11.7% overall. Breast cancer is a heterogeneous and hormone-dependent cancer that accounts for around 22.9% of all female cancers and is the second most commonly diagnosed cancer in women in developing countries [7]. Over the years, about one in 8 women, and one in 1000 men develop invasive breast

cancer [8].

1.4 Motivation

The motivation to complete a demanding project in this fascinating area of study was the driving force behind this project. The perfect combination of data-science and medical field helped us in building this project. It was appealing to learn about a new area of computing which includes AI and DL. Our project will help doctors speed up and automate the process of detecting a normal and abnormal mammogram and further obtaining the ground truth for the abnormal mammogram. It will also help them to prioritise the mammograms according to the level of severity.

1.5 Problem Statement

To develop an automated system for detection and segmentation of tumours using mammograms in Cranial-Caudal and Medial-lateral oblique (CC and MLO) views using Deep Learning Techniques.

1.6 Objectives

The main objectives of our project revolve around the classification of tumors as Benign and Malignant and then to detect location of the tumour by segmenting the lesions in an abnormal mammogram. Finally, to evaluate the performance of various Deep Learning approaches for detection and segmentation of tumours.

1.7 Outcomes

Our project will be able to develop a software tool which automatically detects cancerous tissues. The software will also be able to identify location of tumour and it will act an an assisting tool to radiologists to classify or choose the abnormal mammogram and prioritise based on level of concern.

Chapter 2

Literature Survey

2.1 Introduction

Mammography can be used as a noninvasive method for screening purposes and assisting the radiologists in prognosis and treatment. Bray et al. [3] have evaluated a status report of the various types of cancer that are diagnosed in both sexes combined and mainly among females, breast cancer is the most commonly detected cancer and the major cause of mortality . Lung cancer is the most common cancer (11.6% of the total cases) and the leading cause of cancer death(18.4% of the total cancer deaths), closely followed by female breast cancer (11.6%), prostate cancer (7.1%), and colorectal cancer (6.1%) for incidence and colorectal cancer (9.2%), stomach cancer (8.2%), and liver cancer (8.2%) for mortality. Islam et al. [9] have explored the different breast cancer imaging techniques such as mammography, magnetic resonance imaging (MRI), and ultrasound for breast cancer diagnosis and compared their effectiveness, advantages, and disadvantages for detecting early-stage breast cancer. Singh et al. [10] says that error-free diagnosis of breast cancer at an early stage can reduce the mortality associated with the disease. Thermographic image and usage of artificial neural networks have improved the accuracy of thermography in early diagnosis of breast abnormality. They evaluated the role of image thermography in early breast cancer detection. [11] introduced a multistage system for detecting MCs in mammograms. They used a back-propagation neural network to find candidate calcification regions first, cleaned network output to remove thin elongated structures and used a measure of local density for final classification. [12] applied a two-level algorithm for the detection of MCs using diverse-AdaboostSVM. Six features (four wavelet plus two gray level features) were computed for neural network to detect candidate MC pixels. As a result, 25 features from candidate MCs were extracted and further reduced with geometric linear discriminant analysis (GLDA). The classifier was built with diverse Adaboost SVM.

A mass in mammogram is defined as a space-occupying lesion seen in more

than one projection [13]. The general procedure for detecting mass is first to detect suspicious regions, then extract shape and texture features, and finally detect mass regions through classification or removing false positive regions [14]. [15] used texture features to distinguish mass and non-mass regions. [16] used an adaptive density-weighted contrast enhancement filter to obtain potential masses and used Laplacian Gaussian for edge detection. Morphological features were extracted for classifying normal and mass ROIs. [17] first segmented the boundary of ROI using an edge-based approach and then computed geometric and shape features. Neural networks were trained to distinguish true mass from normal regions.

Since the 1990's CNN's applications can be traced in the field of medical diagnostics just when calcification where detected in digital mammogram. Due to their outstanding performance CNN's have attracted significant attention for feature extraction. There was a breakthrough in 2008 for image classification. Krizhevsky et al. [18] trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. Simonyan et al. [19] investigated the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting.Increasing depth using an architecture with very small (3x3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. Selvanthi et al. [1] deployed the deep learning techniques such as convolutional neural network,sparse autoencoder, and stacked sparse autoencoder. The performance of these techniques is analyzed and compared with the existing methods. From the analysis, it is observed that the stacked sparse autoencoder performs better compared to other methods.

2.2 Breast Cancer Detection Problems

A patient's risk profile is a good criterion for selecting the screening techniques to detect early stage breast cancer. One of the prominent medical imaging techniques for this purpose is X-ray Mammography. One of the major constraints of X-ray Mammography is its low sensitivity in dense breasts making it difficult to interpret. Dense breasts, which are common in younger women are more likely to develop breast cancer and the sensitivity of Mammography in these can be as low as 30%-40% [9]. Another disadvantage of X-ray Mammography is the exposure of patients to the X-ray radiations, which may induce cancerous cells. Furthermore, the Mammography screening process can be uncomfortable because the breast has to be compressed against flat surfaces to improve image quality. A more effective and comfortable alternative for detection of tumours in dense breasts is an Ultrasound. MRI is usually suggested for women with a high risk of developing cancer. An integrated use of these three screening techniques has shown to enhance the likelihood of detecting breast cancer and simultaneously reduces the rates of false detection. Additionally, this integrated use is remarkable in detecting ductal carcinoma insitu and invasive cancer [9].

2.3 Breast Cancer Segmentation

The goal of semantic segmentation is to label each pixel of the image with a particular or corresponding class of what is being represented in the image. The output will be a high resolution image in which pixel level image classification is done. In semantic segmentation the expected outputs are not only labels and bounded box (BB) parameters but a high resolution image.

Alfonso et al. [20] proposed a work which explores the use of characterization features extracted based on breast-mass contours obtained by automated segmentation methods, for the classification of masses in mammograms according to their diagnosis. Abdul Malek et al. [21] presented a combined method to do segmentation for micro calcifications. They applied the Region growing and Boundary segmentation. M. Kavitha et al. [22] attempted to segment the pectoral muscle and then tried to diagnose the lump. Jumaat et al. [23] tried segmentation of masses in the mammogram by application of active contouring methods. For the segmentation of the desired area, R. Subash Chandra Boss et al. [24] applied the Fuzzy C-mean clustering mean algorithm. Samir M. et al. [25] applied the double thresholding method with enhanced and a deeper approach to segment the masses. Xiaoming Liu et al. [26] studied the mass segmentation with respect to the Level Set segmentation and shape analysis. Abdelhameed et al. [27] proposed a segmentation method for breast cancer from thermal images based on Chaotic Salp Swarm Algorithm (CSSA). In the proposed method, a segmentation algorithm is formulated using the quick-shift method for superpixels extraction whose parameters are optimized by CSSA.

2.4 Technologies used for Segmentation

U-Net is a popular end-to-end encoder-decoder network for semantic segmentation that is originally invented for bio-medical image segmentation tasks and for pixel level image classification. The architecture has U shaped connections that connect multi-channel feature map which form a contracting path to capture all the high resolution features and a symmetric expanding path that enables accurate localization [28]. Kallenberg et al. [29] proposed a model wherein breast density segmentation and mammography risk scoring were done by utilizing unlabeled data and using unsupervised DL technique. Singh et al. [30] proposed an efficient model to detect and segment the cancerous regions in mammogram images using a Max-Mean and Least-Variance technique to detect the ROI and segmenting the image. Punita et al. [31] proposed an automated breast cancer detection method using a swarm optimization technique called Dragonfly Optimization. The texture details are extracted using GLCM and GLRM techniques from segmented images and fed into a Feed Forward Neural Network (FFNN).

Chapter 3

Methodology

3.1 Block Diagram - Detection

The overall plan followed throughout the work is depicted in 3.1.

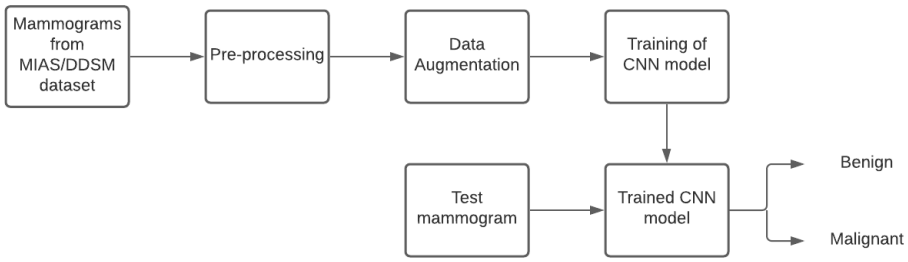


Figure 3.1: Block Diagram - Detection

To start with the project, we worked on a simple image classification algorithm using CNN because CNNs can be thought of automatic feature extractors from the image. It effectively uses adjacent pixel information to down-sample the image first by convolution and then uses prediction layer at the end. After studying the four classification models, we were left to decide whether to go for supervised or unsupervised learning model [32].

In a supervised learning model, the algorithm learns on a labeled dataset, providing an answer key that the algorithm can use to evaluate its accuracy on training data, where as an unsupervised model, in contrast, provides unlabeled data that the algorithm tries to make sense of by extracting features and patterns on its own.

There are two main areas where supervised learning is useful: classification problems and regression problems.

Classification problems ask the algorithm to predict a discrete value whether the image belongs to one of the many mentioned classes i.e identifying the input data as the member of a particular class or group. In a training dataset of animal images, that would mean each photo was pre-labeled as one defined class or the other. The algorithm is then evaluated by how accurately it can classify new images of other similar classes of data.

Finally, a test image was used to check the performance of the model. This trained model classifies the mammogram as benign or malignant. To compare and evaluate the efficiency of our trained models we also calculated Precision, recall and F1 score.

3.1.1 Pre-processing

To remove noise and enhance the quality of these images the dataset undergoes preprocessing. Based on the results obtained, pre-processing of dataset can be done to reduce the noise and convert the raw data into understandable format. It helps to remove unwanted information from the image and allows users to have more valuable datasets. Existing datasets are always incomplete and data cannot be sent to a training model directly. If done so, it always would cause errors. Thus, data pre-processing plays a major role in any classification algorithm.

3.1.2 Data Augmentation

The problem with small datasets is that models trained with them do not generalize well data from the validation and test set. Hence, these models suffer from the problem of overfitting. To increase the diversity of the training data which includes rescaling, rotation, horizontal flip etc, the training set performs data augmentation. Data augmentation is another way we can reduce overfitting on models, where we increase the amount of training data using information only in our training data [33]. Data augmentation refers to increasing the amount of data available by changing various parameters of the image. The amount of data can be increased by rotating the image, changing the dimensions, flipping the image. It is an effective way to increase the efficiency of the model and reduce the losses while training.

3.2 Block Diagram - Segmentation

Segmentation block diagram is shown in the figure 3.2.

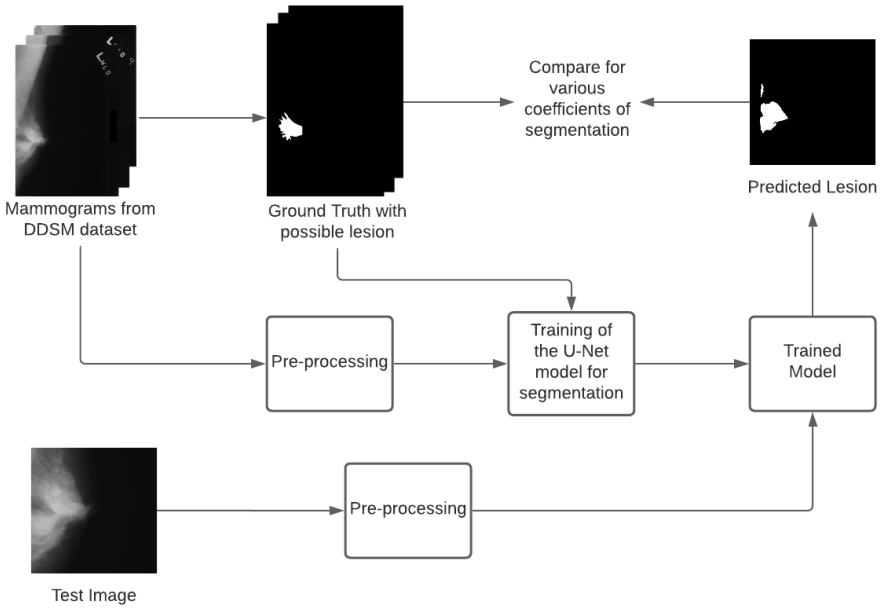


Figure 3.2: Block Diagram - Segmentation

After applying CNN for detection now it's time to explore more about its potential. Image segmentation using U-NET architecture will help us identify the area of possible lesion present in the mammogram.

Let us understand this block diagram, firstly we created a dataset of ground truth images with possible lesions from the existing mammograms in the DDSM dataset which was given as an input to the U-NET model for training.

Mammograms from the DDSM dataset performed pre-processing and was also given as input for training along with the groundtruth images. After training, we used a test image which performed pre-processing and was used to test our trained model which predicted the area of possible lesion present in the mammogram.

We then compared our predicted lesion with the ground truth image by evaluating various parameters such as dice similarity coefficient and jaccard index.

3.3 Deep Neural Networks

Deep Neural Networks (DNNs) have seen a wide range of applications in classification problems, natural language processing, automated driven cars etc. A deep neural network basically comprises of inputs, neurons with different weights, layers such as Max pooling, Convolution, Average, Fully Connected, etc which we will further understand in section 3.6, dropout, and the output. More precisely, DNNs can be understood as neural networks with more number of layers and increased number of neurons. The are basically ANNs with hidden layers. There is no limit to the number of layers or number of neurons in a DNN, rather it is said "the more the better", but practically speaking, it has been impossible to train a neural network with large number of hidden layers.

3.4 Deep Learning

One of the key aspects in most machine learning methods is the way data is represented, that is, which features to use. If the features used are badly chosen, the method will fail regardless of its quality. There are two main ways of obtaining features, manually choosing them such as physiological values in medical applications or automatically generating them, an approach known as representation learning. The latter has proven to be more effective in problems such as computer vision, as it is very difficult for us humans to know what makes an image distinguishable. The main characteristic of Deep Learning is that it is capable of making abstractions, by building complex concepts from simpler ones. Given an image, it is capable of learning concepts such as cars, cats or humans by combining sets of basic features, such as corners or edges. This process is done through successive "layers" that increase the complexity of the learned concepts. The idea of depth in Deep Learning comes precisely from these abstraction levels. Each layer gets as input the output of previous one and it uses to learn higher level features.

3.5 Convolutional Neural Networks

In the past few years, Deep Learning has become an area of interest to the researchers. CNN overcomes the limitations of traditional machine learning approaches. It is a deep learning approach that is widely used for solving complex problems. Machine learning promises to reduce the efforts by making the machines to learn themselves through past experiences [34] using three approaches of learning namely, learning under supervision, without supervision and semi-supervised learning [35]. Deep belief networks, recurrent neural networks, convolution neural networks etc. are different deep learning architectures. Convolution Neural Network, often called ConvNet, has deep feed-forward architecture and has astonishing ability to generalize in a better way as compared to networks with fully connected layers [36].

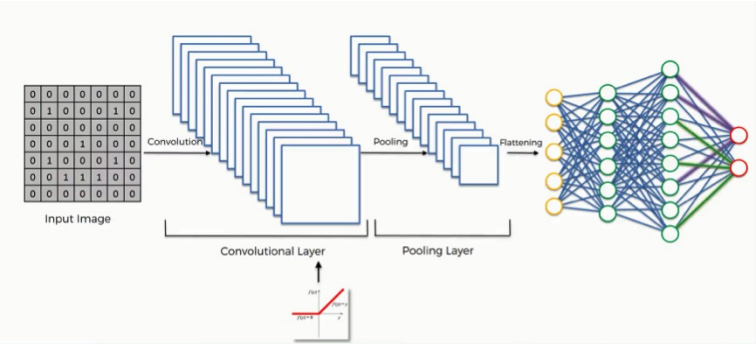


Figure 3.3: Convolutional Neural Network

[37] describes CNN as the concept of hierarchical feature detectors in biologically inspired man. There are many considerable reasons why CNN is considered above other classical models. First, the key interest for applying CNN lies in the idea of using concept of weight sharing, due to which the number of parameters that needs training is substantially reduced, resulting in improved generalization. Due to lesser parameters, CNN can be trained smoothly and does not suffer overfitting. Secondly, the classification stage is incorporated with feature extraction stage, both uses learning process. Thirdly, it is much more difficult to implement large networks using general models of ANN than implementing in CNN.

3.6 Layer types

3.6.1 Convolutional Layer

A Convolutional Layer is the major building block of a neural network. It is inspired in traditional MLPs, but having some major differences. The main ones are that each layer has a single set of weights for all neurons shared weights, and that each neuron only processes a small part of the input space. It uses all the parameters introduced in previous section.

3.6.2 Pooling Layer

It is common to include these layers between successive stacks of convolutional ones, in order to progressively reduce the size of the image representation.It works by taking each channel of its receptive field, and resizes it by keeping only the maximum of its values. It is usually used with 2 x 2 kernels, and a stride of 2, which halves each side size. This reduces the overall size in 75 percent by picking the largest of 2 x 2 patches. This kind of layers does not have weights that need training, and it only uses the stride and kernel size parameters.

Its utility consists in reducing the amount of weights to learn, which reduces computational time as well as probability of over-fitting.

3.6.3 Fully Connected Layer

These layers are basically neural layers connected to all neurons from previous layer. They do not use any of the introduced parameters, instead they use the number of neurons. The output they produce could be understood as a compact feature vector representing the input image. They are also used as output layers, with one neuron per output, as usual.

3.7 Classification Networks

3.7.1 AlexNet

AlexNet is a convolutional neural network, designed by Alex Krizhevsky in 2012. This network was entered in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second best entry [18]. This is a large, deep convolutional neural network consisting of eight layers as shown in Figure 3.4. The key feature of this model is that it can be trained using multiple GPUs for improved speed in training as well as usage of bigger datasets.

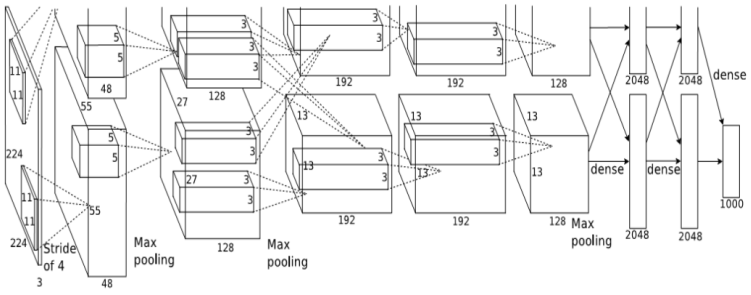


Figure 3.4: AlexNet Architecture

ImageNet is a dataset consisting of 15 million labeled high-resolution images, over 22,000 categories. As it consists of variable-resolution images, the authors down-sampled the images to a fixed resolution of 256x256. The primary concern on this dataset is to prevent overfitting, which was observed using ReLU non-linearity which also influenced the performance of the training model by faster learning. Overlapping pooling reduces computation and controls overfitting.

AlexNet consists of 8 deep layers in total which includes 5 convolutional layers and 3 fully connected (dense) layers as shown in Figure 3.4. Layer one is a convolutional layer with input image size of 224x224x3. Number of filters used in this layer are 96 with filter size of 11x11x3 and stride 4. The output dimension

of layer one is 55x55x96 which is split across two GPUs, so 55x55x48 for each GPU. Layer two is a max pooling layer followed by a convolution with input image of size 55x55x96. The max pooling layer results to 27x27x96. Number of filters used are 256 with filter size of 5x5x48. The output dimension changes to 27x27x256 which is again split across two GPUs, so 27x27x128 each GPU. Layer three, four and five follow on similar lines as shown in Figure 3.4. Layer six is fully connected layer with an input of 13x13x128 which is transformed into a vector. The output then results to 1x2048. Layer seven and eight follow on similar lines. Lastly the output is passed through softmax activation in order to normalize classification vector.

3.7.2 VGG-16

VGG-16 was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry group lab of Oxford University in 2014. This particular network was entered into the 2014 ImageNet challenge [19]. It had 16 weight layers as shown in Figure 3.5. The design of this is very similar to Alexnet except that they have made this a little systematic. They have also stuck to only one filter size, 3, which is the smallest filter size throughout the layers. There are around 138 million parameters.

Figure 3.5: VGG-16 Architecture

Input: VGG takes in a 224x224 pixel RGB image. For the ImageNet competition, the authors cropped out the center 224x224 patch in each image to keep the input image size consistent.

Convolutional layers: The convolutional layers in VGG use a very small receptive field ie. 3x3, smallest possible size that still captures left/right and up/down. The convolution stride is fixed to 1 pixel so that the spatial resolution is preserved after convolution.

Fully Connected Layers: VGG has three fully connected layers. 4096 channels each and the last one has 1000 channels softmax layer.

Hidden layers: All of VGG's hidden layers use RelU (a huge innovation from Alexnet that cut training time.) The input to the network is an image of dimension (224x224x3).

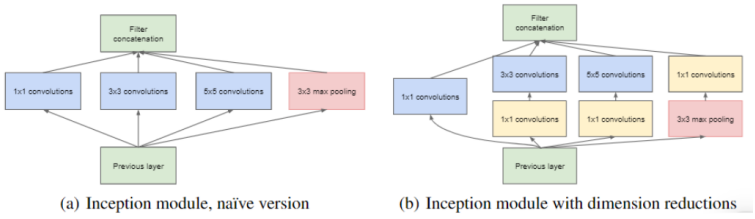
The first two layers have 64 channels of 3x3 kernel with padding 1 and stride 1 followed by a maxpool layer of stride 2, size 2x2. The output dimension results in 112x112x65. The third and fourth convolutional layers have 128 channels of

3x3 kernel, followed by a maxpool layer same as before. The output dimension changes to 56x56x128. The fifth, sixth and seventh layers have 256 channels of 3x3 kernel, followed by a maxpool layer. The output dimension then changes to 28x28x256. The eighth, ninth and tenth convolutional layers have 512 channels of 3x3 kernel, followed by a maxpool layer. The output dimension results in 14x14x512. The eleventh, twelfth and thirteenth layers also have 512 channels of 3x3 kernel, followed by a maxpool layer. The output dimension then changes to 7x7x512.

After the stack of convolution and max pooling layers, we flatten the output to make it a feature vector as shown in Figure 3.5. After this there are three fully connected layers, which takes in input from the last feature vector and outputs a 1x1x4096 vector. The output of the three fully connected layers is passed to the softmax layer in order to normalize classification vectors. All the hidden layers use ReLU as it's activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradients.

3.7.3 GoogleNet

GoogleNet is a deep convolutional network architecture proposed by Google (with collaboration of many universities) in 2014. It is the winner of ILSVRC 2014. It has provided a significant decrease in error rate as compared to previous winners AlexNet (Winner of ILSVRC 2012) and ZF-Net (Winner of ILSVRC 2013) and significantly less error rate than VGG (2014 runner up). It has 22 layers of deep architecture [38]. The architecture was designed to keep computational efficiency in mind. The idea behind that the architecture can be run on individual devices even with low computational resources. The architecture also contains two auxiliary classifier layers connected to the output of Inception (4a) and Inception (4d) layers.



(a) Inception module, naïve version (b) Inception module with dimension reductions

Figure 3.6: Inception module of GoogleNet Architecture

GoogleNet uses many different methods such as 1x1 convolution and global average pooling that enables it to create a deeper architecture. The inception architecture uses 1×1 convolution in its architecture. These convolutions used to decrease the number of parameters (weights and biases) of the architecture.

1x1 convolution is used to reduce the number of parameters of the architecture further which increases the depth.Global average pooling is another method used in the architecture of GoogleNet. It is used at the end of the network. This layer takes a feature map of 7x7 and averages it to 1x1. This also decreases the number of trainable parameters to 0 and increases accuracy of 0.6%. GoogleNet has an inception module which is different from previous architectures such as AlexNet, ZF-Net. In this architecture, there is a fixed convolution size for each layer as shown in Figure 3.6. In the Inception module 1×1, 3×3, 5×5 convolution and 3×3 max pooling performed in a parallel way at the input and the output of these are stacked together to generate the final output. The idea behind that convolution filters of different sizes will handle objects at multiple scales better.

3.7.4 ResNet50

ResNet is known as the Residual Neural Network. This model won the imagenet classification competition in 2015. The evolution of this network helped in training models with more than 150 layers. The input image has to preferably be of the size 224x224.

In traditional deep neural networks, there is a significant degradation in the training accuracy and ResNet addresses this problem by introducing the concept of residual network. The basic framework for the network is based on the VGG networks though the complexity is much less than that of VGG. The network architecture is divided into two networks namely, Plain Network and the Residual Network as shown in figure 3.7. The Plain network consists of convolutional layers mostly containing 3×3 filters and follow two simple design rules: (i) for the same output feature map size, the layers have the same number of filters; and (ii) if the feature map size is halved, the number of filters is doubled so as to preserve the time complexity per layer. We perform downsampling directly by convolutional layers that have a stride of 2. The network ends with a global average pooling layer and a 1000-way fully-connected layer with softmax. The total number of weighted layers is 34.

Based on the Plain network, the Residual network comprises of shortcut connections as shown on the right hand side of the figure 3.7, which turn the network into its counterpart residual version. In the figure, the solid line shortcuts can be directly used if the dimensions of the input and output are same. For the dotted line shortcut, there are two options, (i) The shortcut performs identity mapping with padding for increasing dimensions and (ii) Match the dimensions with 1x1 convolution. [39] Resnet50 has a total of 50 layers. The 2-layer blocks in 3.7 are replaced by 3-layer blocks. There are a total number of are 11 million parameters in ResNet50. The convolutional layers are finally connected to a fully connected layer with 1000 neurons and then the output is achieved.

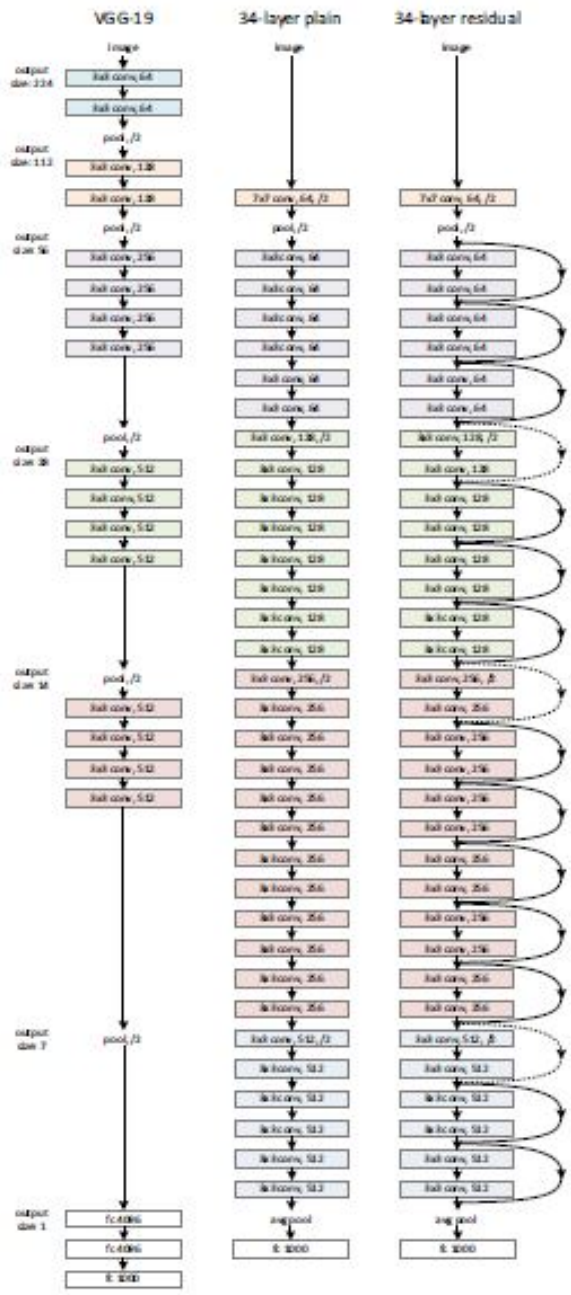


Figure 3.7: Plain and Residual network of Resnet Architecture

3.7.5 EfficientNet

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. Unlike conventional practice that arbitrary scales these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients [40]. In particular, EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet, while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet [40].

The compound scaling method is justified by the intuition that if the input image is bigger, then the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image. This method uses a compound co-efficient ø to scale width, depth, and resolution together. The formula for scaled attributes is as shown below:

$$depth : d = \alpha^{\phi}$$

$$width : w = \beta^{\phi}$$

$$resolution : r = \gamma^{\phi}$$

$$s.t : \alpha \times \beta^2 \times \gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

Here, alpha, beta, and gamma are scaling multiplier for depth, width and resolution respectively and be obtained using grid search. ø is a user-specific co-efficient which takes real numbers like and controls resources which is 2ø. So if we have double resources available than what a model is currently using, we can take find ø using $2ø = 2$ and hence ø is 1 for such cases.

The base model of EfficientNet family, EfficientNet-B0 architecture wasn't developed by engineers but by the neural network itself. They developed this model using a multi-objective neural architecture search that optimizes both accuracy and floating-point operations. Taking B0 as a baseline model, the authors developed a full family of EfficientNets from B1 to B7. The architecture of EfficientNet-B0 is shown in Figure 3.8.

One can see that architecture uses 7 inverted residual blocks but each is having different settings. These blocks also use squeeze excitation block along with swish activation.

Swish Activation: Swish is a multiplication of a linear and a sigmoid activation.

$$Swish(x) = x \times S(x)$$

Inverted Residual Block: In an original residual block (introduced in ResNet), skip connections are used to connect wide layers (aka layers with a large number of channels) and there are fewer numbers of channels inside a block (aka narrow layers). The inverted residual block does the opposite, skip connections connects narrow layers while wider layers are between skip connections.

Squeeze and Excitation Block: Squeeze and Excitation block gives the output of shape (1 x 1 x channels) which specifies the weightage for each channel

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Figure 3.8: EfficientNet-B0 Architechture

and the great thing is that neural network can learn this weightage by itself like other parameters.

MBConv Block: MBConv block takes two inputs, first is data and the other is block arguments. The data is output from the last layer. A block argument is a collection of attributes to be used inside an MBConv block like input filters, output filters, expansion ratio, squeeze ratio etc.

3.8 Segmentation Networks

3.8.1 U-Net

With AlexNet, there was a breakthrough in the field of CNNs and based on that, many other classification networks were introduced. The majority of classification networks were based on supervised learning where pre-defined labels are provided to the algorithm for training.The classification tasks are only based on a single label output but in the biomedical field where a visual aspect is included,the expected output is the exact location of the tumor or lesion.

To help with this, a fully convolutional network was proposed specifically for medical imaging community and named as U-Net. The layers are designed such that it take the shape of the alphabet 'U', as shown in 3.9, and hence gets its name from there .

U-Net helped in the training low number of images yet producing a precise segmented output. The architecture comprises of two paths, the expansive and the contracting path. The network does not have any fully connected layers. excessive data augmentation is an important step while using U-Net since the training dataset is very small specially while using it in the clinical field.

The left side of 3.9 comprises of the contracting path and the right side is the expansive path. The contracting path has successive 3x3 convolutions, followed by a rectified linear unit (ReLU) and finally a max-pooling with stride rate 2 for downsampling. The number of feature mas are doubled at every downsampling.

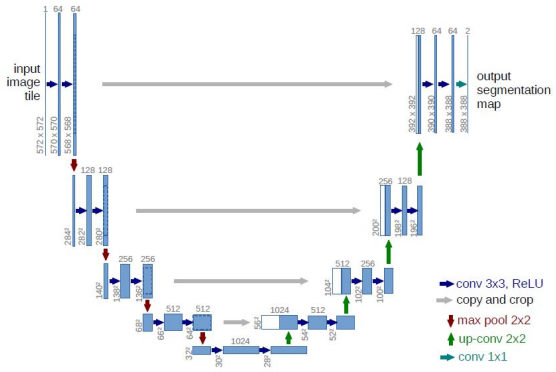


Figure 3.9: U-NET Architecture

For the expansive path, upsampling of feature maps occur which is also known a "up-convolution". This is followd by 2x2 convolution which halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each folowed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution.

The final layer is a 1x1 convolution used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers [28].

3.8.2 SegNet

In the year 2017, after observing some limitations with other segmentation networks, a fully deep convolutional neural network called SegNet was proposed for semantic segmentation. Semantic segmentation is the process in which each pixel in assigned a label.

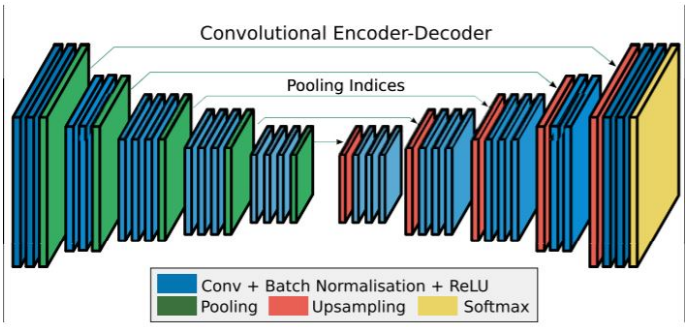


Figure 3.10: SegNet Architecture

The architecture as show in 3.10 consists of an encoder network, followed by a decoder network and then finally a pixel-wise classifier layer. The encoder has 13 covolution layers which are the first 13 layers in the VGG16 architecture as seen in Figure 3.5 which is widely used for classification tasks and thus helps us to use SegNet on large datasets. Since the fully connected layers from the VGG16 network are removed here, the number of parameters decrease significantly in number from 134 to 14.7 million. Each encoder has a corresponding decoder and hence the decoder consists of 13 layers. The final decoder output is fed to a multi-class soft-max classifier to produce class probabilities for each pixel independently.

The encoder performs convolution to produce a set of feature maps which are then batch normalized and given an element wise rectified non-linearity (ReLU). Following this, a max-pooling with stride rate two is performed and the resulting output is sub-sampled by a factor of two.

The decoder network upsamples its input feature maps by reusing the max-pooling indices given by the encoder and further produces sparse feature maps as output. These feature maps are then are convolved with decoder filter to produce dense feature maps. A batch normalization is then applied to each of these feature maps. The high dimensional feature representation at the output of the final decoder is fed to a trainable softmax classifier which classifies each pixel separately. The output of the soft-max classifier is a K channel probability image, where K is the number of classes. The predicted segmentation corresponds to the category with the greatest probability in each pixel.

The major difference between SegNet and UNet is that UNet does not reuse the indices of max-pooling produced by its encoder but rather sends over the entire feature channel produced by the encoder to corresponding and concatenates it to the decoder feature maps [41] .

Chapter 4

Results and Discussion

4.1 Dataset

The input dataset was divided into two categories i.e training dataset and test dataset. Training dataset includes the images which were trained with four models while test dataset was the one which provided unbiased evaluation of the model on the training dataset.

MIAS Dataset: The MIAS database contains the original 322 images (161 pairs) at 50 micron resolution in "Portable Gray Map" (PGM) format and associated truth data which depict CC and MLO views of the scanned images [42]. It is a supervised dataset which consists of the class, radius, severity and X and Y coordinates of each scan. Figure 4.1 shows a mammogram in MIAS dataset showing both, CC and MLO views.

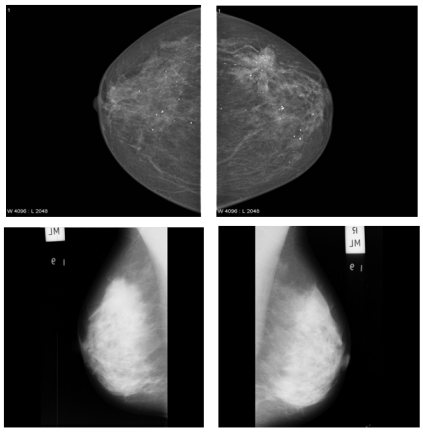


Figure 4.1: Sample image of CC and MLO views in MIAS dataset

DDSM Dataset: Another database used is DDSM which is a database of digitized film-screen mammograms with associated ground truth and other information. The purpose of this resource is to provide a large set of mammograms in a digital format that may be used by researchers to evaluate and compare the performance of computer-aided detection (CAD) algorithms. The database was completed in the fall of 1999. It contains 2620, four view, mammography screening exams [43]. This dataset contains images that have been pre-processed and converted to 299x299 images by extracting the ROIs. The data is stored as tfrecords files for TensorFlow. The dataset contains 55,890 training examples, of which 14% are positive and the remaining 86% negative, divided into 5 tfrecords files. Following the entire plan, the algorithm will detect normal or abnormal scan and from the abnormal scans, it will further detect a benign or malignant tumor. Figure 4.2 shows a mammogram in DDSM dataset showing both, CC and MLO views.

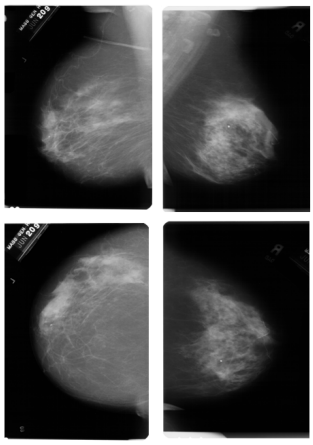


Figure 4.2: Sample image of CC and MLO views in DDSM dataset

4.2 Breast Cancer Detection Results

We implemented our model on 5 models and the training and implementation results including number of layers, loss, validation loss, accuracy, validation accuracy, precision, recall and F1 score have been shown in table 4.1.

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP}$$

Table 4.1: Breast Cancer Detection on MIAS dataset

Name of Model	Number Of Layers	Accuracy (%)	Loss	Validation Loss	Parameters (Million)
AlexNet	8	69.64	1.84	1.94	49.0
EfficientNet	17	72.29	0.49	1.53	5.3
GoogleNet	22	71.67	0.31	0.63	22.2
VGG-16	16	75.46	0.31	0.44	138.0

Table 4.2: Breast Cancer Detection Results on DDSM dataset

Name of Model	Number of Layers	Loss(%)	Validation Loss(%)	Accuracy (%)	Validation Accuracy (%)	Precision	Recall	F1 Score
Alex Net	8	62.46	47.21	49.93	49.69	0.65	0.71	0.67
VGG-16	16	58.18	39.04	68.39	86.42	0.78	0.81	0.79
Efficient Net	17	59.38	37.30	73.00	82.72	0.81	0.77	0.78
Google Net	22	67.03	21.27	72.94	91.36	0.85	0.81	0.82
Res Net	50	65.28	50.16	65.91	73.77	0.79	0.80	0.79

Recall is the ratio of correctly predicted positive observations to the all observations.

$$Recall = \frac{TP}{TP + FN}$$

F1 Score is the weighted average of Precision and Recall.

$$F1score = \frac{Precision + Recall}{2}$$

Here,

$$TP = TruePositives$$

$$FP = FalsePositives$$

$$FN = FalseNegatives$$

Training the model on GoogleNet and EfficientNet has been shown in Fig 4.3 and Fig 4.4.

```
Epoch 15/50
10/10 [==============================] - 37s 4s/step - loss: 0.7386 - accuracy: 0.6407 - val_loss: 0.6191 - val_accuracy: 0.7160
Epoch 16/50
10/10 [==============================] - 37s 4s/step - loss: 0.8732 - accuracy: 0.6369 - val_loss: 0.3280 - val_accuracy: 0.8519
Epoch 17/50
10/10 [==============================] - 36s 4s/step - loss: 0.8823 - accuracy: 0.7132 - val_loss: 0.3686 - val_accuracy: 0.8272
Epoch 18/50
10/10 [==============================] - 37s 4s/step - loss: 0.7695 - accuracy: 0.6459 - val_loss: 0.3661 - val_accuracy: 0.8519
Epoch 19/50
10/10 [==============================] - 37s 4s/step - loss: 0.6555 - accuracy: 0.7417 - val_loss: 0.4883 - val_accuracy: 0.7531
Epoch 20/50
10/10 [==============================] - 37s 4s/step - loss: 0.4975 - accuracy: 0.7395 - val_loss: 0.4553 - val_accuracy: 0.7901
Epoch 21/50
10/10 [==============================] - 38s 4s/step - loss: 0.6470 - accuracy: 0.7244 - val_loss: 0.9063 - val_accuracy: 0.6173
Epoch 22/50
10/10 [==============================] - 36s 4s/step - loss: 0.7336 - accuracy: 0.6609 - val_loss: 0.3751 - val_accuracy: 0.8272
Epoch 23/50
10/10 [==============================] - 37s 4s/step - loss: 0.7639 - accuracy: 0.6664 - val_loss: 0.6080 - val_accuracy: 0.7160
Epoch 24/50
10/10 [==============================] - 38s 4s/step - loss: 0.7570 - accuracy: 0.6702 - val_loss: 1.3385 - val_accuracy: 0.5309
Epoch 25/50
10/10 [==============================] - 37s 4s/step - loss: 0.8580 - accuracy: 0.6220 - val_loss: 0.4578 - val_accuracy: 0.7531
Epoch 26/50
10/10 [==============================] - 36s 4s/step - loss: 0.6703 - accuracy: 0.7294 - val_loss: 0.2127 - val_accuracy: 0.9136
```

Figure 4.3: Training of model on GoogleNet

```
Epoch 40/50
10/10 [==============================] - 43s 5s/step - loss: 0.6319 - accuracy: 0.6992 - val_loss: 0.6499 - val_accuracy: 0.7407
Epoch 41/50
10/10 [==============================] - 44s 5s/step - loss: 0.6058 - accuracy: 0.6622 - val_loss: 0.5525 - val_accuracy: 0.7531
Epoch 42/50
10/10 [==============================] - 44s 4s/step - loss: 0.6748 - accuracy: 0.6897 - val_loss: 0.6315 - val_accuracy: 0.6790
Epoch 43/50
10/10 [==============================] - 43s 4s/step - loss: 0.6171 - accuracy: 0.6818 - val_loss: 0.4726 - val_accuracy: 0.7654
Epoch 44/50
10/10 [==============================] - 44s 5s/step - loss: 0.6730 - accuracy: 0.7156 - val_loss: 0.7650 - val_accuracy: 0.6667
Epoch 45/50
10/10 [==============================] - 43s 5s/step - loss: 0.5163 - accuracy: 0.7846 - val_loss: 0.4175 - val_accuracy: 0.8148
Epoch 46/50
10/10 [==============================] - 44s 4s/step - loss: 0.5069 - accuracy: 0.7695 - val_loss: 0.6308 - val_accuracy: 0.7160
Epoch 47/50
10/10 [==============================] - 44s 5s/step - loss: 0.6137 - accuracy: 0.6748 - val_loss: 0.4929 - val_accuracy: 0.7654
Epoch 48/50
10/10 [==============================] - 45s 5s/step - loss: 0.7335 - accuracy: 0.6694 - val_loss: 0.5459 - val_accuracy: 0.7654
Epoch 49/50
10/10 [==============================] - 45s 5s/step - loss: 0.5219 - accuracy: 0.7377 - val_loss: 0.7747 - val_accuracy: 0.6790
Epoch 50/50
10/10 [==============================] - 45s 5s/step - loss: 0.5938 - accuracy: 0.7300 - val_loss: 0.3730 - val_accuracy: 0.8272
```

Figure 4.4: Training of model on EfficientNet

After training our detection model, we made a binary classifier that classifies the mammograms as benign or malignant. The output obtained from the binary classifier has been shown in Fig 4.5 and Fig 4.6. We trained the model using the desired CNN network and then loaded an image from the DDSM and MIAS dataset for the prediction.

In all the above models we see that when we increase the number of layers, we achieve more accuracy. Here we are scaling up the models. The top accuracy on DDSM is achieved using GoogleNet and for MIAS, VGG16 works the best. So the goal of GoogleNet is to systematically scale up depth, width and resolution of the CNN's. So by randomly increasing layers, there are 3 types of scaling possible namely width scaling, depth scaling and resolution scaling. Depth scaling accounts to adding more feature maps and hence making the network deep. Compound scaling refers to a combination of increasing the width and adding more layers. Saturation happens only if one type of scaling is done. But

Figure 4.5: Binary classifier predicting the mammogram as Benign

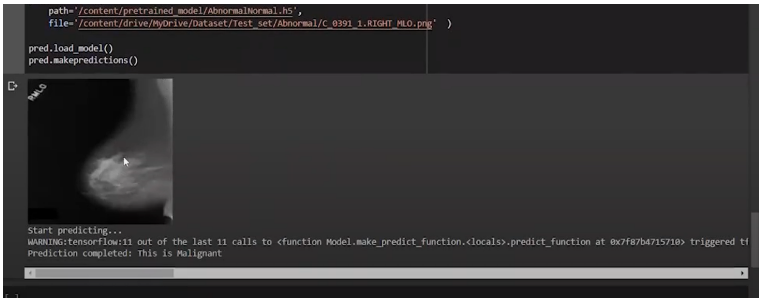


Figure 4.6: Binary classifier predicting the mammogram as Malignant

if we overall increase the parameters, we get more accurate results. The earlier models like AlexNet follow the conventional approach of scaling the dimensions arbitrarily and by adding up more and more layers.

The inconsistency in results and accuracy maybe due to the resolution of images in both the datasets. Due to different size of each image in both datasets, we resize the images so as to obtain a common parameter for training. This results in loss of information and thus decrease the performance of network.

However, our project proposes that if we scale the dimensions by a fixed amount at the same time and do so uniformly, we achieve much better performance. The scaling coefficients can be in fact decided by the user. Though this scaling technique can be used for any CNN-based model, the authors started off with their own baseline model called EfficientNetB0.

While implementing the models, we found out that AlexNet was very slow as compared to the other models with low efficiency. VGG16 improved the efficiency but the training time was too long. Efficient Net further increased the efficiency but the losses were more. The best efficiency with low number of losses was provided by GoogleNet as seen in 4.1.

4.3 Breast Cancer Segmentation Results

Segmentation in mammography images is defined as a stage where it is possible to locate the exact place of the lump in breast [44]. We used the U-net model for segmentation and it predicts the tumour from the mammogram by segmenting it.

Table 4.3 shows the accuracy, validation accuracy, dice similarity coefficient and jaccard index that have been calculated after segmentation of tumor from mammogram.

Dice similarity coefficient is a coefficient used to gauge the similarity of two samples. It has been a popular metric for evaluating the accuracy of automated or semi-automated segmentation methods by comparing their results to the ground truth. The Jaccard index, also known as the Jaccard similarity coefficient is also used to compare two samples. IOU is intersection over union. It is a number from 0 to 1 that specifies the amount of overlap between the predicted and ground truth bounding box. An IoU of 0 means that there is no overlap between the two images.

Table 4.3: Breast Cancer Segmentation Results

Accuracy (%)	99.72
Validation Accuracy (%)	97.42
DSC(Dice Similarity Coefficient)	0.82
IOU/Jaccard Index	0.98

Figures 4.7 and 4.8 present the original image loaded from the DDSM dataset, trained image given by the U-NET model, the ground-truth and the predicted output from the model for both the MLO and CC view. As we can see, in Figure 4.7 the original mammogram shows the pectoral muscle in MLO view and while training this is removed. The idea behind removing the pectoral muscle is that it can hinder the training and provide us false positive results.

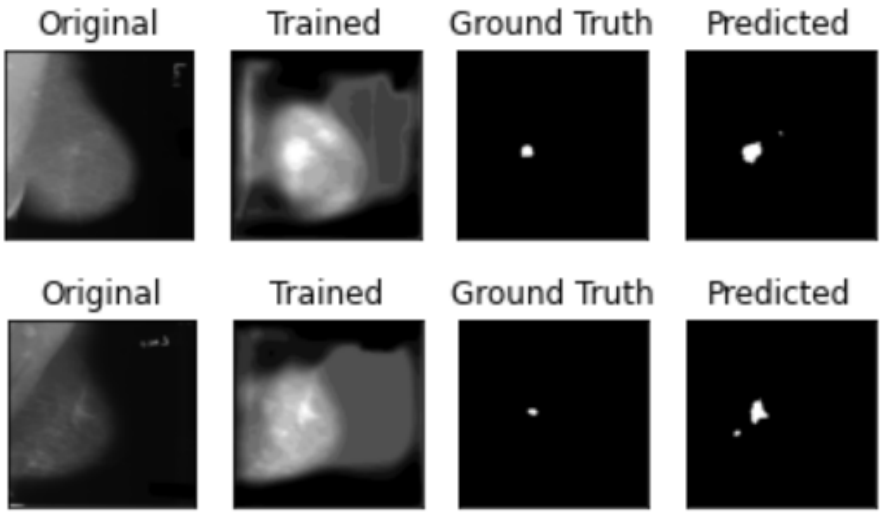


Figure 4.7: Segmentation Result on MLO view mammogram

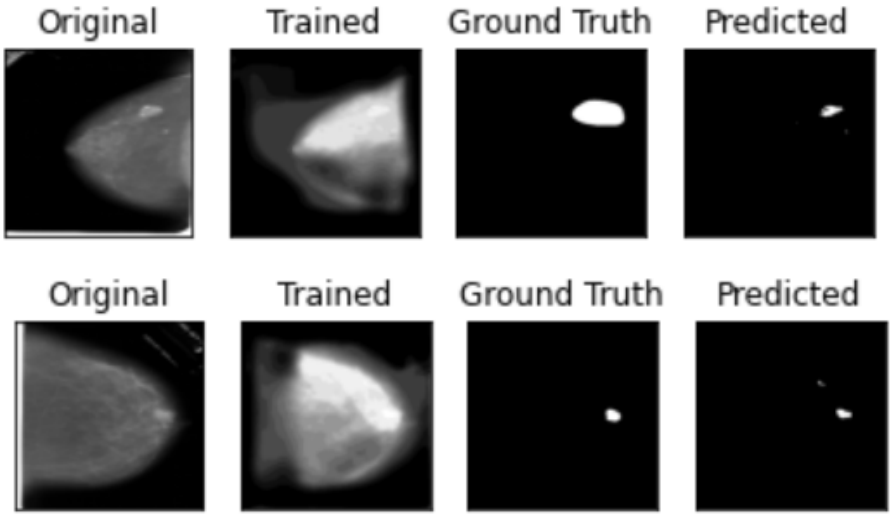


Figure 4.8: Segmentation Result on CC view mammogram

In Figure 4.9 we depict results which have not given us expected output. This maybe due to the density of breast. Some breasts are fatty or grandular in nature which limits the performance of mammography as well. The trained network considers the fatty tissues as abnormality and hence it predicts the whole region as a tumor.

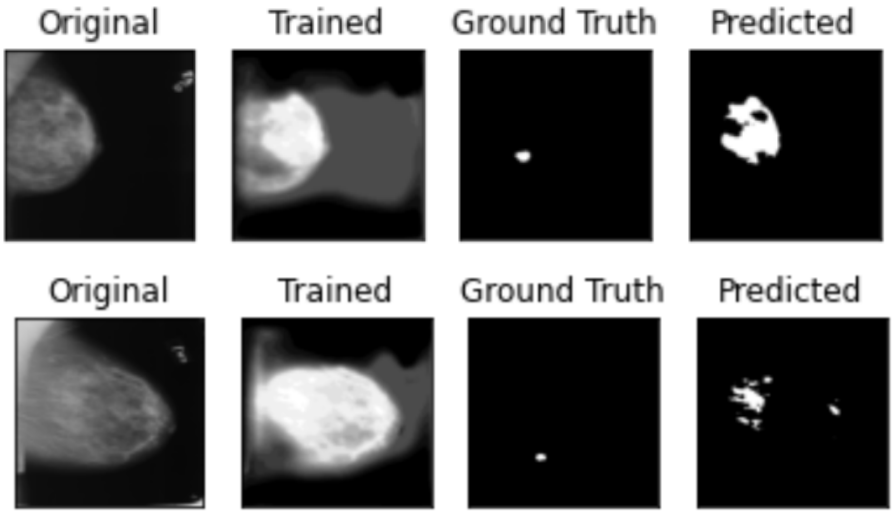


Figure 4.9: Worst Case Scenario

Chapter 5

Conclusion & Future plans

Early detection helps with early diagnosis of breast cancer. In the proposed work, the detection model we developed detects mass abnormality and classifies the mammograms into benign and malignant. We also compare the recent deep learning models. The proposed methodology for the work includes methods for image pre-processing, mass detection, data augmentation, and finally predicting the outcome of new data given to the model. These techniques have resulted in a much more normalized dataset and increased the number of training as well as testing images.

The segmentation model helps us in comparing the original ground truth of the mammogram with the ground truth produced by the trained model. It helps with the localization of the tumor and automates the usual process of contouring the tumor which is manually done by radiologists.

We worked on mammographic images available publicly such as MIAS and DDSM datasets. GoogleNet, despite being a deep and bigger CNN model, achieves better results and produces much better efficiency for the mass detection as compared with AlexNet, VGG16, ResNet and EfficientNet. This is because of high accuracy obtained due to more trainable parameters and faster network. Using UNet model, we successfully locate the calcifications and masses in mammographic images. This helped us achieve a higher DSC coefficient with a better jaccard index.

The classification and segmentation of breast cancer using the tool we have developed will assist the radiologists to prioritize the mammograms. It will help patients to get early treatment which may result in reducing mortality rate in breast cancer cases. It will also help physician to decide on the course of treatment for better patient management.

For both classification and segmentation, it is expected that similar quality of result can be achieved by using much more non-inception type of network of similar depth and width. Also, a well built interface that provides a simple way for the radiologists to classify and segment the mammographic images can be implemented. All these suggest future work towards moving to a feasible and a useful idea for creating a clinical tool which contributes towards healthcare.

Bibliography

[1] D. Selvathi and A. A. Poornila, "Deep learning techniques for breast cancer detection using medical image analysis," in *Biologically rationalized computing techniques for image processing applications*, pp. 159–186, Springer, 2018.

[2] S. Hadush, Y. Girmay, A. Sinamo, and G. Hagos, "Breast cancer detection using convolutional neural networks," *arXiv preprint arXiv:2003.07911*, 2020.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[4] E. Hadgu, D. Seifu, W. Tigneh, Y. Bokretsion, A. Bekele, M. Abebe, T. Sollie, S. D. Merajver, C. Karlsson, and M. G. Karlsson, "Breast cancer in ethiopia: evidence for geographic difference in the distribution of molecular subtypes in africa," *BMC women's health*, vol. 18, no. 1, pp. 1–8, 2018.

[5] S. V. Sree, E. Y.-K. Ng, R. U. Acharya, and O. Faust, "Breast imaging: a survey," *World journal of clinical oncology*, vol. 2, no. 4, p. 171, 2011.

[6] G. Hamed, M. Marey, S. Amin, and M. Tolba, *Deep Learning in Breast Cancer Detection and Classification*, pp. 322–333. 03 2020.

[7] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *International Journal of Cancer*, vol. 127, pp. 2893–2917, jun 2010.

[8] P. I. Ali, W. Wani, K. Saleem, and Correspondence, "Cancer scenario in india with future perspectives," *Cancer Therapy*, vol. 8, pp. 56–70, 11 2011.

[9] M. S. Islam, N. Kaabouch, and W. C. Hu, "A survey of medical imaging techniques used for breast cancer detection," in *IEEE International Conference on Electro-Information Technology, EIT 2013*, pp. 1–5, IEEE, 2013.

[10] D. Singh and A. K. Singh, "Role of image thermography in early breast cancer detection- past, present and future," *Computer Methods and Programs in Biomedicine*, vol. 183, p. 105074, 2020.

[11] Z. Yang, M. Dong, Y. Guo, X. Gao, K. Wang, B. Shi, and Y. Ma, "A new method of micro-calcifications detection in digitized mammograms based on improved simplified pcnn," *Neurocomput.*, vol. 218, p. 79–90, Dec. 2016.

[12] F. Harirchi, P. Radparvar, H. A. Moghaddam, F. Dehghan, and M. Giti, "Two-level algorithm for MCs detection in mammograms using diverse-adaboost-SVM," in *2010 20th International Conference on Pattern Recognition*, IEEE, aug 2010.

[13] *ACR BI-RADS®-Atlas der Mammadiagnostik*. Springer Berlin Heidelberg, 2016.

[14] Y. Li and H. Chen, "A survey of computer-aided detection of breast cancer with mammography," *Journal of Health & Medical Informatics*, vol. 7, no. 4, 2016.

[15] A. Petrosian, H.-P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: classification of mass and normal tissue by texture analysis," *Physics in Medicine and Biology*, vol. 39, pp. 2273–2288, dec 1994.

[16] N. Petrick, H.-P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Transactions on Medical Imaging*, vol. 15, no. 1, pp. 59–67, 1996.

[17] D. Cascio, F. Fauci, R. Magro, G. Raso, R. Bellotti, F. D. Carlo, S. Tangaro, G. D. Nunzio, M. Quarta, G. Forni, A. Lauria, M. Fantacci, A. Retico, G. Masala, P. Oliva, S. Bagnasco, S. Cheran, and E. Torres, "Mammogram segmentation by contour searching and mass lesions classification with neural network," *IEEE Transactions on Nuclear Science*, vol. 53, pp. 2827–2833, oct 2006.

[18] A. Krizhevsky and I. Sutskever, "H. geoffrey e.,"alex net,"," *Adv. Neural Inf. Process. Syst*, vol. 25, pp. 1–9, 2012.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[20] A. Rojas Domínguez and A. K. Nandi, "Toward breast cancer diagnosis based on automated segmentation of masses in mammograms," *Pattern Recognition*, vol. 42, no. 6, pp. 1138–1148, 2009. Digital Image Processing and Pattern Recognition Techniques for the Detection of Cancer.

[21] A. Malek, W. Rahman, A. Ibrahim, R. Mahmud, S. Yasiran, and A. Jumaat, "Region and boundary segmentation of microcalcifications using seed-based region growing and mathematical morphology," *Procedia - Social and Behavioral Sciences*, vol. 8, pp. 634–639, 12 2010.

[22] M. Kavitha and M. Rejusha, "Segmentation of pectoral muscle and detection of masses in mammographic images," *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, pp. 1201–1204, 2015.

[23] A. Jumaat, W. Rahman, A. Ibrahim, and R. Mahmud, "Segmentation of masses from breast ultrasound images using parametric active contour algorithm," *Procedia - Social and Behavioral Sciences*, vol. 8, pp. 640–647, 12 2010.

[24] R. S. C. Boss, K. Thangavel, and D. A. P. Daniel, "Mammogram image segmentation using fuzzy clustering," in *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, IEEE, mar 2012.

[25] S. M., A. A., H. E., and M. T., "Breast cancer detection with mammogram segmentation: A qualitative study," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[26] X. Liu, X. Xu, J. Liu, and J. Tang, "Mass classification with level set segmentation and shape analysis for breast cancer diagnosis using mammography," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, pp. 630–637, Springer Berlin Heidelberg, 2012.

[27] A. Ibrahim, S. Mohammed, H. A. Ali, and S. E. Hussein, "Breast cancer segmentation from thermal images based on chaotic salp swarm algorithm," *IEEE Access*, vol. 8, pp. 122121–122134, 2020.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[29] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, *et al.*, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.

[30] A. K. Singh and B. Gupta, "A novel approach for breast cancer detection and segmentation in a mammogram," *Procedia Computer Science*, vol. 54, pp. 676–682, 2015.

[31] S. Punitha, A. Amuthan, and K. S. Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 348–358, 2018.

[32] R. Sathya and A. Abraham, "Comparison of supervised and unsupervised learning algorithms for pattern classification," *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, no. 2, pp. 34–38, 2013.

[33] J. Wang, L. Perez, *et al.*, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Networks Vis. Recognit*, vol. 11, 2017.

[34] J. Carbonell, R. Michalski, and T. Mitchell, "An overview of machine learning in "machine learning," *An Artificial Intelligence Approach", Morgan Kaufman Publishers (Ed.)*, 1983.

[35] S. C. Dharmadhikari, M. Ingle, and P. Kulkarni, "Empirical studies on machine learning based text classification algorithms," *Advanced Computing*, vol. 2, no. 6, p. 161, 2011.

[36] C. Nebauer, "Evaluation of convolutional neural networks for visual recognition," *IEEE transactions on neural networks*, vol. 9, no. 4, pp. 685–696, 1998.

[37] J. Fieres, J. Schemmel, and K. Meier, "Training convolutional networks of threshold neurons suited for low-power hardware implementation," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 21–28, IEEE, 2006.

[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[40] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105–6114, PMLR, 2019.

[41] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[42] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, *et al.*, "Mammographic image analysis society (mias) database v," vol. 1, no. 21, 2015.

[43] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography, iwdm-2000," in *International Workshop on Digital Mammography. Medical Physics Publishing, Madison WI*, pp. 212–218, 2001.

[44] P. Yousefikamal, "Breast tumor classification and segmentation using convolutional neural networks," 2019.