

Abstract

Breast Cancer is considered to be the second most cause of cancer among women in the world. Mammography being the standard medical imaging tool for detecting early instances of Breast Cancer can help in curing the disease before it gets severe. However, if the screening of mammogram is done incorrectly and which is very common due to human error and lack of expert radiologists then it can create a unavoidable problem and may harm the patients with unnecessary treatments and operations or lack of treatments in case of false negative and might eventually responsible for the death of a patient. Deep learning methods have recently gained popularity among the medical image research community due to the rapid rise of computational resources and availability of clinical image data. It has boosted the performance of different computer vision tasks, e.g. object detection, segmentation, super-resolution, classification etc. In order to solve the classification problem, deep convolutional neural networks (CNNs) have the ability to automatically extract the hierarchical low-level features and from low level features it can extract the high level features to finally give the prediction for any diseases. Convolutional Neural Networks (CNNs) are used in this project as a part of deep learning which predicts the early instances of breast cancer from the mammogram images. In this report, various deep learning models are used in order to predict breast cancer using mammogram images. With the help of these deep learning models, low level and high level features are automatically extracted in a hierarchical way that may help to detect early prediction of breast cancer. Various deep learning CNN models such as VGG16, VGG19 and ResNet50 are implemented with the help of transfer learning techniques to analyse the performance and efficiency of the classification model and predict the mammogram images in three different classes with associated DL Risk Score. Ultimately, a best accuracy of 88% is achieved on the DDSM (Digital Dataset for Screening Mammography) dataset with Precision of 0.90, 0.87, 0.88 for Benign, Malignant and Normal cases respectively. The inclusion of Risk Factors (such as Patients' age, Breast Density Score, Age at first birth) along with the corresponding mammogram images have the ability to increase the precision of the deep learning based prediction.

Keywords: Deep Learning; Convolution Neural Networks; Transfer Learning; Breast Cancer Prediction; Mammogram; Classification

Table of Contents

Chapters:	Page No.
1. Introduction	01
1.1 Motivation	01
1.2 Problem Description	04
1.3 Thesis Organisation	04
2. Context and Literature Survey	05
2.1 Breast Cancer Prediction	05
2.2 Machine Learning based techniques	06
2.3 Literature Survey	11
2.4 Deep Learning techniques	12
2.5 Summary	19
3. Methodology: Design and Implementation	20
3.1 Description of Datasets	20
3.2 Deep Learning pipeline	22
3.3 Implementation of CNN Training Phase	33
3.4 Summary of CNN Model Design and Training	36
4. Results	44
4.1 Experimental Setup	44
4.2 CNN model trainable parameters	45
4.3 Visualisation of results	45
4.4 Hyperparameter tuning	47
4.5 Performance of CNN Models	48
4.6 Final Prediction and Classification with DL Score	57
5. Conclusion	60
5.1 Achievements	60
5.2 Limitations	60
5.3 Future Works	61
Bibliography	63

List of Figures

Figure Names	Page No.
Figure 1.1: Example of three different mammograms	03
Figure 2.1: Evolution of breast cancer screening systems	06
Figure 2.2: k-NN classifier used for Breast Cancer Prediction	07
Figure 2.3: Decision tree for breast cancer screening	07
Figure 2.4: SVM Classifier for breast cancer prediction	08
Figure 2.5: ANN for Breast Cancer Prediction	09
Figure 2.6: Max- Pool and Avg. Pool applied on mammogram	14
Figure 2.7: Three different scenarios of learning	16
Figure 2.8: Examples of data augmentation techniques	17
Figure 2.9: Dropout in a CNN	18
Figure 3.1: Age distribution in Mini- DDSM Dataset	21
Figure 3.2: Density distribution in Mini- DDSM Dataset	21
Figure 3.3: Metadata which includes risk factor	22
Figure 3.4: Deep Learning based breast cancer prediction	23
Figure 3.5: Artefacts present in Mammograms	24
Figure 3.6: The process of artefacts removal	25
Figure 3.7: Image Normalization for mammograms	27
Figure 3.8: Label encoding for mammogram classes	28
Figure 3.9: VGG16 Architecture	29
Figure 3.10: VGG19 Architecture	29
Figure 3.11: ResNet50 Architecture	30
Figure 3.12: Sigmoid and Softmax activation functions	31

Figure Names	Page No.
Figure 3.13: Metadata of Mini- DDSM	33
Figure 3.14: Cleaned Dataset of Risk Factor	33
Figure 3.15: Image Transformation (Data Augmentation)	35
Figure 3.16: VGG16 CNN Architecture for breast cancer prediction	37
Figure 3.17: VGG19 CNN Architecture for breast cancer prediction	38
Figure 3.18: ResNet50 Architecture for breast cancer prediction	39
Figure 3.19: VGG16 CNN Architecture with Risk Factor.....	42
Figure 4.1: Hyperparameters of different CNN models	47
Figure 4.2: Performance of different CNN model	48
Figure 4.3: Accuracy vs No. of Epochs (VGG16 model)	49
Figure 4.4: Loss vs No. of Epochs (VGG16 model)	50
Figure 4.5: ROC for VGG16 model	51
Figure 4.6: Accuracy vs No. of Epochs (VGG19 model)	51
Figure 4.7: Loss vs No. of Epochs (VGG19 model)	52
Figure 4.8: ROC for VGG16 model	53
Figure 4.9: Accuracy vs No. of Epochs (ResNet50 model)	53
Figure 4.10: Loss vs No. of Epochs (ResNet50 model)	54
Figure 4.11: ROC for ResNet50 model	55
Figure 4.12: Accuracy vs No. of Epochs (VGG16 model with risk factors)	56
Figure 4.13: Loss vs No. of Epochs (VGG16 model with risk factors)	56
Figure 4.14: ROC for VGG16 model with risk factors	57
Figure 4.15: Prediction model for Normal Case	58
Figure 4.16: Prediction model for Malignant Case	58
Figure 4.17: Prediction model for Benign Case	59
Figure 5.1: Population Statistics in DDSM dataset	60

Chapter 1

Introduction

1.1 Motivation

Breast Cancer is one of the most common forms of cancer amongst women in the world. Breast cancer is the second most cause of cancer deaths among women worldwide. Every 4 minutes in India, a woman gets diagnosed with breast cancer, and one woman succumbs to breast cancer every 13 minutes, making it the most lethal cancer among Indian women. Women in India are generally diagnosed at a later, more advanced stage with very poor prognosis. About 1 in every 28 women are expected to develop breast cancer during their lifetime. By 2030, breast cancer will cause most deaths among women in India than any other disease [1]. Cancer survival becomes very difficult in later stages of its growth. More than 50% of Indian women suffer from stage 3 or stage 4 of breast cancer. Post cancer survival for women with breast cancer was reported 60% for Indian women, as compared to 80% in the United States of America [2]. Due to the lack of proper prognosis and health facilities in countries like India, it is utmost important for the survival of patients to detect breast cancer at an early stage.

Test screenings have been used to detect the early instances of breast cancer before a patient develops any kind symptoms like lumps that can be felt to the touch of a hand. The general conventional golden method used for early detection of breast cancer is mammography which is basically x-rays around the breast area. This scan shows the backgrounds in black and dense areas in white pixels, which may correspond to calcifications or masses (e.g. lumps or cysts). If suspicious areas in the breast are detected, mammograms are followed by breast ultrasounds for analysing these masses. Ultrasound images are also used for detailed imagery of the breast, usually when a malignant tumour is detected. [3] If any of the screening methods mentioned earlier detects a potential risk of developing breast cancer, then biopsies can be done to confirm the results of screening tests. Biopsies consist of extracting cells or a small part of the breast's tissue and then sending them to a lab to be analysed by pathologists to finally get the test reports. Due to the invasive nature of biopsies, it is desirable to use medical imagery tools to detect early signs of breast cancer that can be treated efficiently rather than conducting a biopsy. Screening of mammograms has been found to reduce mortality rate of the disease to a great

extent. Mammography is the primary imagery method used for early breast cancer detection. Mammography being the standard medical imaging tool for detecting early instances of Breast Cancer can help in curing the disease before it gets severe. However, if the screening of mammogram is done incorrectly and which is very common due to human error and lack of expert radiologists then it can create an unavoidable problem and may harm the patients with unnecessary treatments and operations or lack of treatments in case of false negative which might eventually responsible for the death of a patient. [4] Hence, despite its benefits, screening mammography is associated with a very high risk of false positives or false negatives due to the lack of expert radiologists. Breast cancer detection using mammograms or any form of cancer detection using medical imagery, relies on the conventional diagnosis done by expert radiologists [5]. Radiologists have to go by what they see on medical images to declare if the patient has breast cancer or not. But many a time there comes a situation where the expert gives false positive or false negative results and this leads the patient's life to danger which can not be accepted at any cost. Traditional medical image analysis is less encouraging currently because of following reasons:

- Lower number of experts and higher number of patients, so time taken by experts will be more and it will affect diagnosis and detection.
- Time required to train radiologists or experts for medical image analysis is long and expenses are also large.
- It is a tedious task to segment, detect and classify a medical image while analysed by a human being.

Early detection of breast cancer through screening tests such as mammograms is a way to maximise patients' survival rate by treating the disease at an early stage. However, no matter the expertise of radiologists examining mammograms, external factors such as fatigue, distractions and human error need to be minimised [6], as the rate of missed breast cancers during initial mammogram screenings are as high as 30% [7]. To convey the complexity of mammogram interpretation, Figure 1.1 illustrates the three different kinds of mammograms containing either normal or abnormal (benign and malignant) cases, and how similar they all look to an untrained eye.

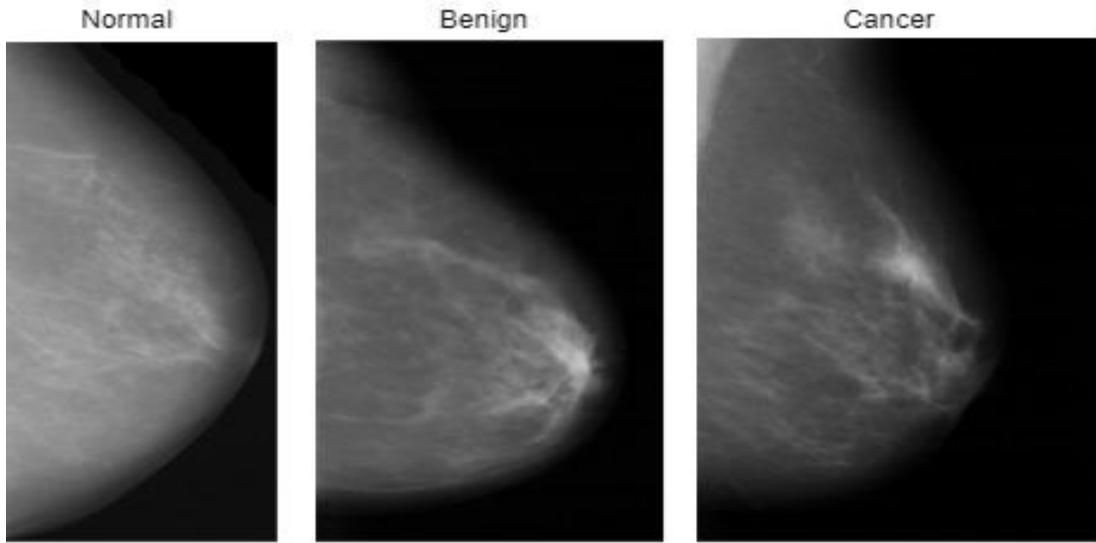


Fig1.1: Example three different mammograms containing either normal or abnormal (benign and malignant) cases.

Computer-Assisted Detection (CAD) software can help minimise the number of false interpretations and increase the accuracy of mammography screening [8]. The motivation behind this project is to apply various deep learning based techniques that can be implemented to a system which can learn the low and high level features from the mammogram images and then detect early instances of breast cancer in order to prevent late treatments due to false negatives as well as preventing unnecessary treatments in cases of false positives. Ultimately, the long-term target of this project is to apply these research on deep learning based methods for early detection of breast cancer in the medical college hospitals across Assam and India to improve the screening process and prevent late treatments due to false negatives as well as preventing unnecessary treatments in cases of false positives. This will allow a general artificial intelligence system which is capable of detecting breast cancer with higher accuracies than radiologist diagnoses.

1.2 Problem Description

With the help of deep learning based techniques, in theory, we could increase the accuracy of the mammogram screenings for predicting early signs of breast cancers. However, these techniques require a very large amount of data to learn the breast cancer's underlying patterns and adapt to the new test cases, and require some powerful computing resources to accelerate the process while training and learning from the mammogram images, making them very hard to optimise.

1.3 Thesis Organisation

This thesis consists of six different chapters. The first chapter is entitled Introduction. It illustrates the initial understanding of the project through motivation, background and problem description of breast cancer screening systems. The second chapter is entitled as Context Survey and Literature Review. It elaborates different conventional computer aided systems in predicting breast cancer and also describes the evolution of these algorithms and methods. Literature Review gives a good idea about current deep learning works in context of breast cancer prediction. The fourth chapter is entitled Methodology: Design and Implementation. This chapter talks about the mammogram dataset used in this project and also illustrates how data have been collected and pre-processed which will eventually be used as an input to the CNN model to give early prediction of breast cancer. This Chapter gives the idea about the complete design part of the deep learning CNN model for breast cancer prediction. In the later part of the chapter it explains about the implementation of the project. The fourth chapter is entitled as Results. Here performance of different CNN models have been discussed. In this Chapter, we also discuss the prediction given by the breast cancer CNN model. Fifth and the final chapter is entitled as Conclusion. Here in this sixth chapter, achievements of this deep learning based early prediction of breast cancer project are described along with its limitations and future scope.

Chapter 2

Context and Literature Survey

2.1 Breast Cancer Prediction

2.1.1 Medical Imagery screening tests

Before any kind of appearance of symptoms, screening tests have been used to predict the early signs of breast cancer. Mammograms which are low concentration X-ray around the breast region are usually used as initial screening tests. In mammogram screening if any suspicious areas are found, it is followed by ultrasound and also breast MRIs (Magnetic Resonance Imaging) for further analysis of the screening test [9]. If any of the tests mentioned earlier is found suspicious or it reveals a potential presence of abnormality in the breast, then biopsies are conducted to confirm the results. In biopsies, cells or small portions of tissue from the breast are collected and sent to the laboratory for further analysis by the pathologists to finally get a sure result on the breast cancer status [10].

But due to the very expensive, complex and invasive nature of biopsies, it is preferred to use medical imagery tools like mammography for the early prediction of breast cancer. Mammograms are the main medical imagery tool used by the radiologists for predicting early stages of breast cancer. However, breast cancer prediction using medical imagery relies on the ability of the expert radiologists for accurate diagnosis. The diagnosis depends on the correct interpretation of mammograms which may be prone to errors due to the difficulty of radiologists to correctly interpret them [11].

2.1.2. Machine Learning based techniques

During the late 1990s, machine learning based techniques started to assist the expert radiologists. With the help of machine learning, hidden patterns in the mammogram that previously could not be perceived by the experts were recognised by these new algorithms [12]. This was a shift from Computer Aided Diagnosis (CAD) systems that were fully dependent on humans to systems which were trained on medical imagery datasets to give prediction of early signs of breast cancer [12]. Following figure shows the development of breast cancer screening systems throughout the years from 1980s.

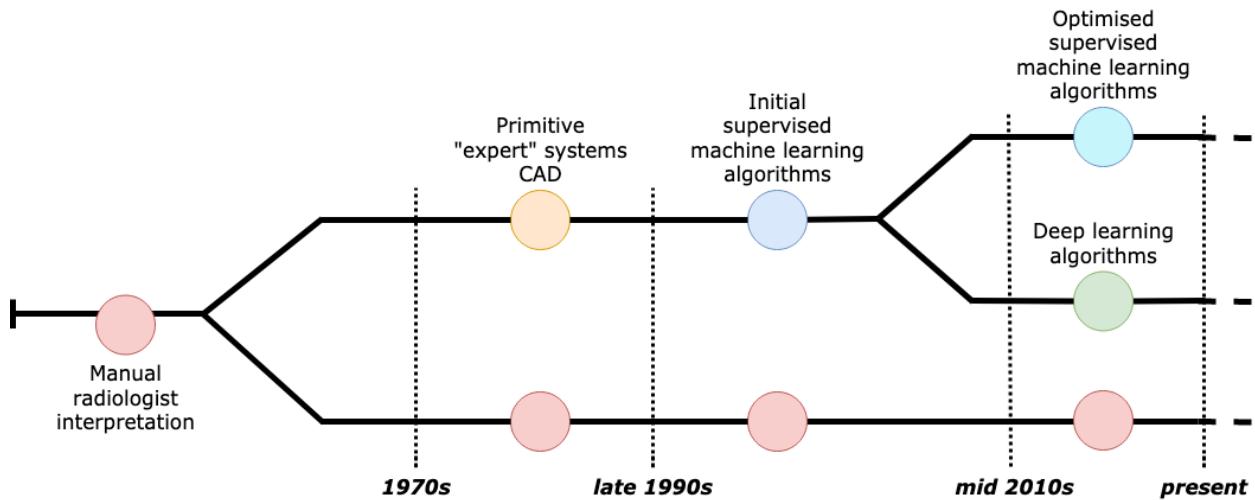


Figure 2.1: evolution of Breast Cancer Screening Systems [12]

2.2 Machine Learning based techniques

2.2.1 Types of machine learning algorithms

K-Nearest Neighbours

K- Nearest Neighbours or commonly known as kNN is a non- parametric model as it does not learn from the data pattern but rather classifies test samples just by looking at the k- Nearest Neighbours [13]. Euclidean distance is used to determine the nearest neighbours of the sample data point. Following is the equation involving the calculation of the distance between two data points “s” and “p” in n- dimensional space

$$d(s, p) = \sqrt{\sum_{i=1}^n (s_i - p_i)^2}$$

k-NN has been used for prediction of breast cancer on the datasets of mammograms such as the “Wisconsin Breast Cancer Detection” (WBCD). This dataset contains ten extracted features such as thickness of the clump, cell sizes etc [14].

In Figure 2.2 a k-NN classifier is shown which is used to distinguish between a benign and a malignant class.

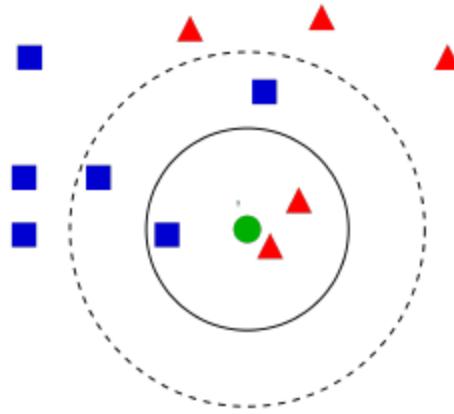


Figure 2.2: k-NN Classifier used for Breast Cancer Prediction

Decision Tree

Unlike k- Nearest Neighbours, the decision trees algorithm is very simple for training and fitting the breast cancer datasets. Decision tree is a tree which is made of many types of nodes. In a decision tree each non- leaf node checks one of the feature vectors and depending upon the value of the feature vectors attribute, branching out of the nodes to a deeper node takes place. When a leaf node is achieved after multiple checks, a classification can be done. In figure 2.4 an example of a decision tree is shown which is applied to breast cancer detection.

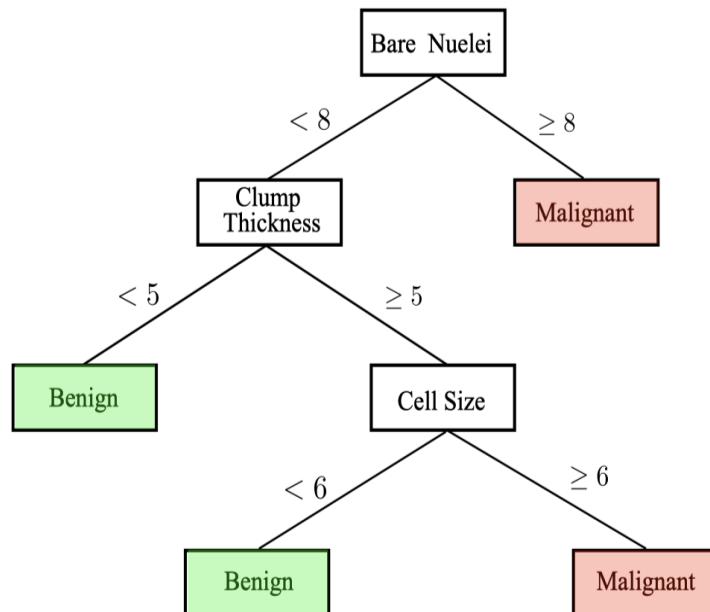


Figure 2.3: Decision tree for breast cancer screening

Support Vector Machine

Support Vector Machine or commonly known as SVM is a linear model which is used for classification and also can be used for regression tasks. SVM can not only solve linear problems but also non- linear problems for many real world practical applications. The SVM uses an algorithm where it creates a hyperplane that can separate the data into different classes that we are trying to classify or predict. Figure: 2.5 shows an example of SVM Classifier which is used to separate benign and malignant cancer cases with the help of features extracted from mammogram datasets [15].

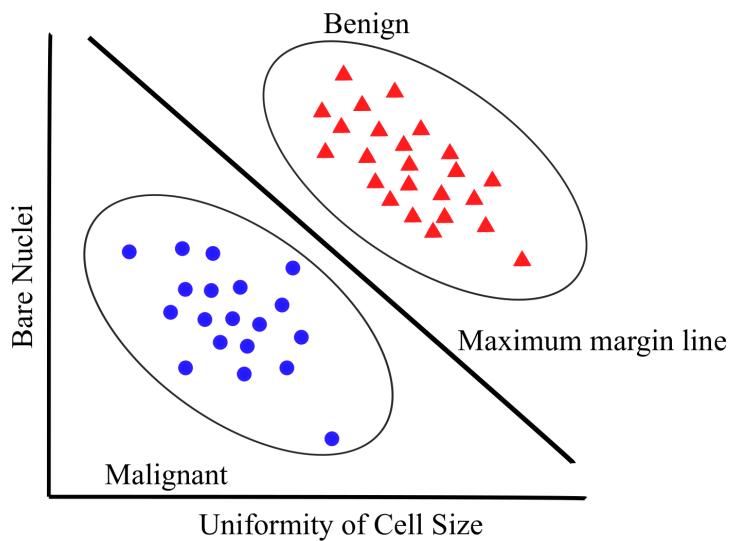


Figure 2.4: SVM Classifier for separating benign cases from malignant cases

Sarosa et al. in 2018 applied Support Vector Machine (SVM) algorithm to a dataset containing raw mammogram images from CBIS- DDSM Dataset where the features are extracted using some image processing techniques like Grey level co-occurrence matrix. This SVM algorithm achieved an accuracy of 63.03% [16].

In machine learning algorithms like SVM, feature extraction and selection, image processing techniques etc hold a very crucial role in the performance of breast cancer prediction.

Artificial Neural Network

Artificial Neural Network or simple ANN is an information processing and learning tool which is loosely inspired by the way our human brain and more precisely neural systems process information and make decisions. ANN is the core building block of deep learning. Artificial Neural Network consists of many neurons which activate an output if the linear sum of its weighted inputs is greater than a fixed threshold value which is usually 0.5 (but not necessarily always). In an ANN neurons are placed in a hierarchical order which are connected with weights. Figure 2.6 shows an example of ANN which is fully connected. This ANN is used to classify benign and malignant cases of mammogram using six input features, eight hidden neurons which are fully connected with the input feature nodes and a single output neuron for the final prediction [15].

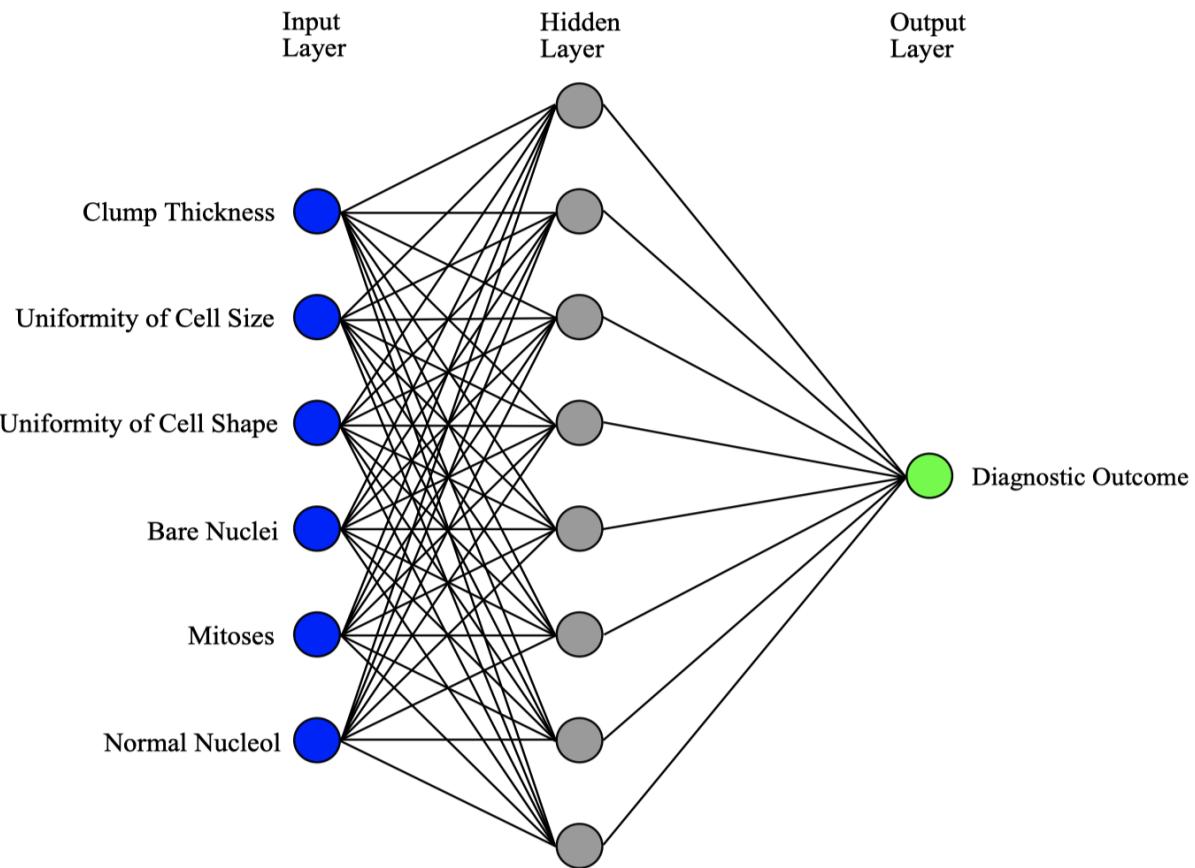


Figure 2.5: Artificial Neural Network for Breast Cancer Prediction

Like any other supervised machine learning algorithms, neural networks learn from the data to give predictions by minimizing the cost function between the predictions and the ground truth labels. Backpropagation algorithm is used to reduce the cost function. During backpropagation the error from the output layer is propagated to hidden layers for each training example which is used to update the values of the weights by using gradient descent algorithm.

When first applied to breast cancer dataset, Artificial Neural Networks (ANNs) showed a very promising result with accuracy of 95.2% in predicting breast cancer. The ANN model was designed with only a simple 3 layer architecture [16].

2.2.1 Comparison among machine learning based algorithms

All the machine learning algorithms that we described above have one thing in common among them and that is the quality of features extracted from the mammogram images. These feature based machine learning algorithms such as k- NN, Decision Trees, SVM and Artificial Neural Networks (ANNs) can not make use of the raw 2D space mammogram images as input directly. These algorithms can only work with the datasets which contain extracted features. For example: Wisconsin Breast Cancer Dataset.

So, there was a need to automate the tasks of extracting features from the raw mammogram images rather than optimising the existing machine learning algorithms and fine tuning the hyperparameters. Deep Learning and Convolutional Neural Networks (CNN) caught major attention when AlexNet won the ImageNet competition back in 2012 and 2014. It was quite a revolution in the field of Computer Vision when a deep Convolutional Neural Network (CNN) model was able to solve such a complicated task of image classification with thousands of classes. CNN could be the most efficient way to automate the tasks of extracting features from the 2D mammogram images in 2D space which can be used to predict breast cancer rather than extracting the features manually from scratch or using flattened image vectors as features where all spatial information is lost.

2.3 Literature Survey

In recent years, deep CNN models have gained so much popularity in the field of breast cancer classification. CNN models are used for detecting and predicting breast cancer efficiently.

Li Shen et al. has designed a deep learning based system for whole image classifiers to detect Breast Cancer. The system was initially trained on lesion annotations of breast mammograms and then the whole image classifier was developed using the trained model. **Nan Wu et al. in 2019** developed a deep convolutional neural network for Breast Cancer Screening which was trained on a very large dataset to achieve an AUC of 0.89. The dataset was initially trained on patches of the breast mammogram and annotations were known. Using the patch classifier, an end to end whole breast mammogram screening model was designed. **Li Shen** developed an end-to-end algorithm for whole-image classification of breast cancer. It requires lesson annotations only at the beginning stage and later whole image classification can be achieved by transferring the same model which was initially trained on lesson annotations. It also described how once the model is trained on a particular dataset can be transferred to classify breast cancer for different dataset. **Karin Dembrower et al.** introduced a deep learning system which can give predictions more accurately on which women are likely to have breast cancer in future looking at the risk score associated with the prediction. They developed the deep CNN model to reduce the False Negative and False Positive cases in breast cancer screening. **Adam Yala et al.** has developed a Deep learning based system where full-field mammogram images were use to train the model which improved risk discrimination compared with the previous conventional Tyrer Cuzick model.

2.4 Deep Learning Techniques

2.4.1 Convolutional Neural Networks

Motivation for CNNs over conventional artificial neural networks

Convolutional Neural Networks are nothing but a deep collection of artificial neural networks which is inspired by the human visual cortex system. The architecture of the CNN makes them very efficient at performing very complex visual problems.

Convolutional Neural Networks are very different from the traditional Artificial Neural Networks. CNNs are not fully connected. CNN can work with large images because of the architecture they have. They are connected partially where neurons from one layer are connected to a few neurons of the previous layer. Because of this kind of nature in CNNs, it can easily work with images having larger dimensions and sizes. This makes CNN very convenient to work with datasets like mammograms. CNN can also process the images faster than any other traditional machine learning methods.

Structure of a CNN Network

A Convolutional Neural Network consists of repeated stacks of convolutional layers and pooling layers which are then followed by few shallow fully connected ANN layers required for the final high level classification task. The main idea behind convolutional and pooling operation is to reduce the input images to a form where it only has the useful information from the images. After extracting the features from the convolution and pooling layers, these features are fed into fully connected ANN layers. It becomes easier for fully connected layers to process information after the convolutional layers pooling layers because of the image dimensions as well as the quality of features it extracted [17]. Figure 2.7 shows an example of Convolutional Neural Network which can classify the early instances of breast cancer.

Convolutional Layers

In the first layer of any convolutional layers neurons are connected only to pixels of the input image in their receptive fields. The neurons are not connected to every pixel. As we go deeper into the convolutional layers, the neurons of the present layers are connected to neurons of the previous layer in a small region. Because of this CNN first focuses on low level features such as

edges, textures etc. As the convolutional layers get deeper they start recognising the high level features like shape, sizes, a particular region etc. In the convolutional layer the more receptive fields are spaced out, the smaller the next layer will be which also reduces the complexity of CNN. In CNN, convolution operation is used which is a mathematical operation where a filter ‘f’ moves over an image ‘I’ in a sliding window manner to calculate a weighted sum. The following equation explain the convolution operation,

$$\hat{I}(x, y) = (I * f)(x, y) = \sum_k \sum_l I(k, l) \cdot f(x - k, y - l)$$

The filters used in convolution operation are learned during the training phase with the help of optimization techniques such as gradient descent. Because of the different filters available in different layers, many complex features can be extracted with the help of convolutional layers.

Pooling Layers

Pooling layer is used in CNN to downsample the image as it traverses from one layer to the next. Pooling operation thus reduces the load on the GPU. The difference between convolutional layer and pooling layer is that the pooling layer has no trainable parameters as the neurons in the pooling layers have no weights involved. There are two types of pooling operations that are widely used. The first one is Max- Pool or Maximum Pooling. As the name suggests it returns the maximum value for a defined portion of the window of pixels. Another pooling operation is Average Pooling where it returns the average of all the pixel values. Max- Pool is preferred to use most of the time in CNN over Average Pooling because in Max- Pool, it can retain the most important features and suppress any noises or unnecessary pixels from the image [18].

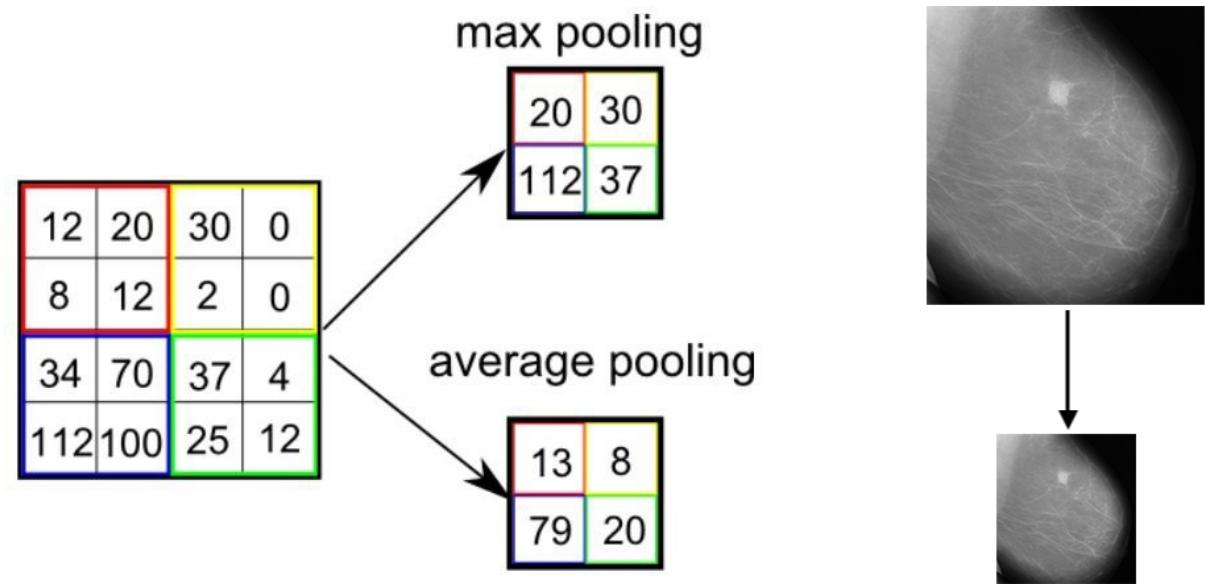


Fig 2.6: Max- Pooling and Average Pooling operation applied on mammogram

Activation Functions in CNN

Similar to activation functions used in Artificial Neural Nets, activation functions in CNN are also used to connect convolutional and pooling layers. Rectified Linear Unit (ReLu) is the most common type of activation function that is widely used in convolutional neural networks (CNNs). At the end of the convolutional layers and pooling layers, the output is flattened and then fully connected to a dense layer. The dense neural network takes the flattened 1 dimensional feature vectors which is the output of the stacked convolutional and pooling layers. The fully connected layers perform the high- level task of classification based on the features learned by the convolutional layers. Depending upon the number of classes to predict, a softmax activation Function can be used in case of multi- class classification whereas sigmoid activation function can be used in binary classification.

2.4.2 Convolutional Neural Network Architecture

VGG (Visual Geometry Group) and AlexNet are two most popular CNN models which have won the ImageNet Challenge back in 2012 and 2014. It still remains a very popular CNN model in domains like medical imagery. AlexNet is a CNN model which has five convolutional layers. Unlike traditional CNN in AlexNet all convolutional layers are not separated by pooling layers.

Following the same concept in VGG architecture multiple convolutional layers with smaller filter sizes are stacked which can pick up more complex features. There are many variants of VGG exists which are based on the number of layers. Two most popular VGG models are VGG16 and VGG19 which have sixteen and nineteen layers respectively. Due to the depth of VGG networks, this takes a long time for training and also it may suffer from a very common problem of vanishing gradient [19]. Because of the depth of networks like VGG, the weights of the convolution filters reduce as we move deeper into the network. This affects the backpropagation and leads to an exponentially smaller gradient while going backward to the initial layers and in turn it affects the network from learning.

To avoid this problem of vanishing gradient, more complex CNN architectures were created such as ResNet (which make use of residual modules), GoogLeNet (which uses inception modules). These architectures improve the accuracy of the CNN model with a cost of little computing resources.

2.4.3 Application of Deep Learning in Breast Cancer Prediction

In practice, deep learning modules require lots and lots of data while implementing to achieve acceptable performance levels. But in reality, it's very hard to get or collect labelled datasets because they are not always abundantly available as it requires time and expensive computing resources to collect and pre- process the large datasets with ground truth labels. One of the challenges of implementing a deep learning based breast cancer prediction system is with datasets of mammograms exceeding thousands, it's very hard to gain access to the computing resources needed for processing the raw mammogram images. Whereas with small amounts of data it suffers from overfitting. Overfitting occurs when the model does not generalise well to the new dataset and can only correctly recognize the cases it has been before while training. In overfitting the model learns the data too well and it kind of memorizes it.

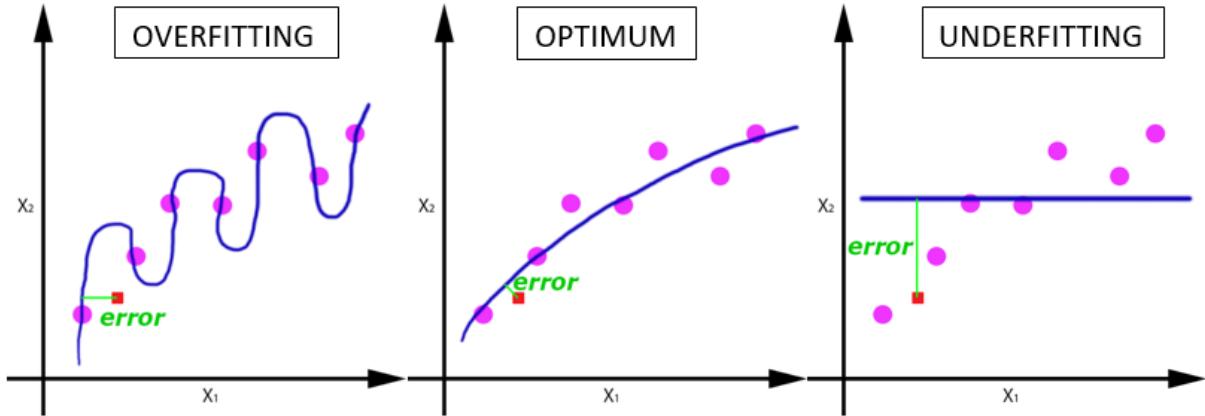


Figure 2.7: Three different scenarios of learning

Transfer Learning

Transfer learning is a commonly used technique in deep learning when there is scarcity of data. Transfer learning makes use of pre-trained CNN models which are already trained on very large datasets like ImageNet dataset. The knowledge gathered by high-performing CNNs in different domains which have larger datasets that can be easily transformed to a related domain such as medical imagery [20].

The transfer learning models have the ability to classify millions of images and thus it can be adapted to any classification task. This can be done by simply replacing the dense fully connected layers which makes the final prediction.

Regularisation techniques

Overfitting is one of the drawbacks of deep learning. To avoid the problem of overfitting many new regularisation techniques have been introduced.

Data Augmentation

To solve the problem of overfitting caused by smaller datasets, data augmentation techniques have been widely used. This data augmentation technique is often used when a deep learning model which has a large number of parameters try to learn with a smaller dataset. The images are augmented and new images are created by using image transformation techniques such as

translation, rotation, scaling, skew, zoom, horizontal flip, vertical flip, addition of noises, brightness and contrast change etc.

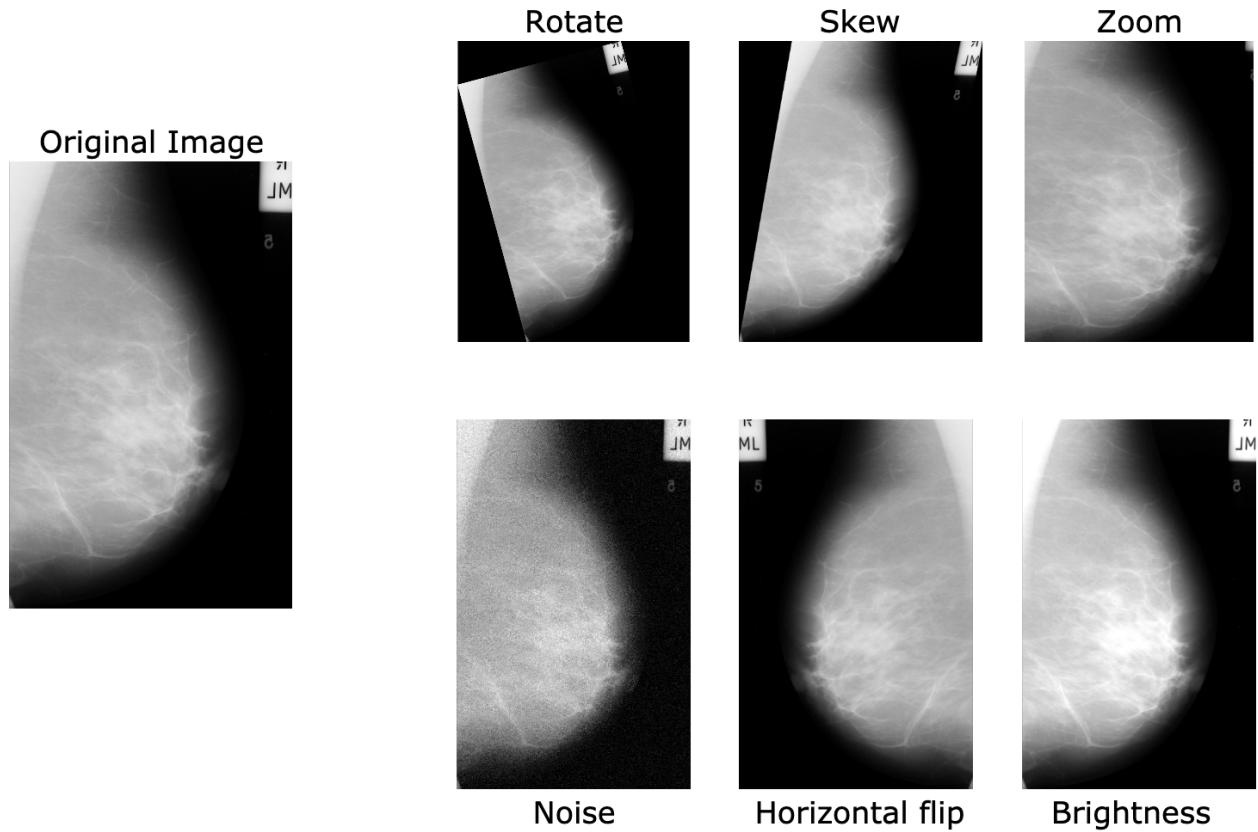


Figure 2.8: Examples of data augmentation techniques

Dropout

This is the most popular yet simple regularisation technique used in deep learning CNN networks. Dropout, as the name suggests is a regularisation algorithm where we randomly drop neurons of any layer (generally in the fully connected layers) except at the output layers during the forward propagation and backpropagation while training the CNN model. We also drop the weights connected between input and output neurons. By doing this, it prevents neurons from adapting too much from their neighbouring neurons and overfitting the CNN model.

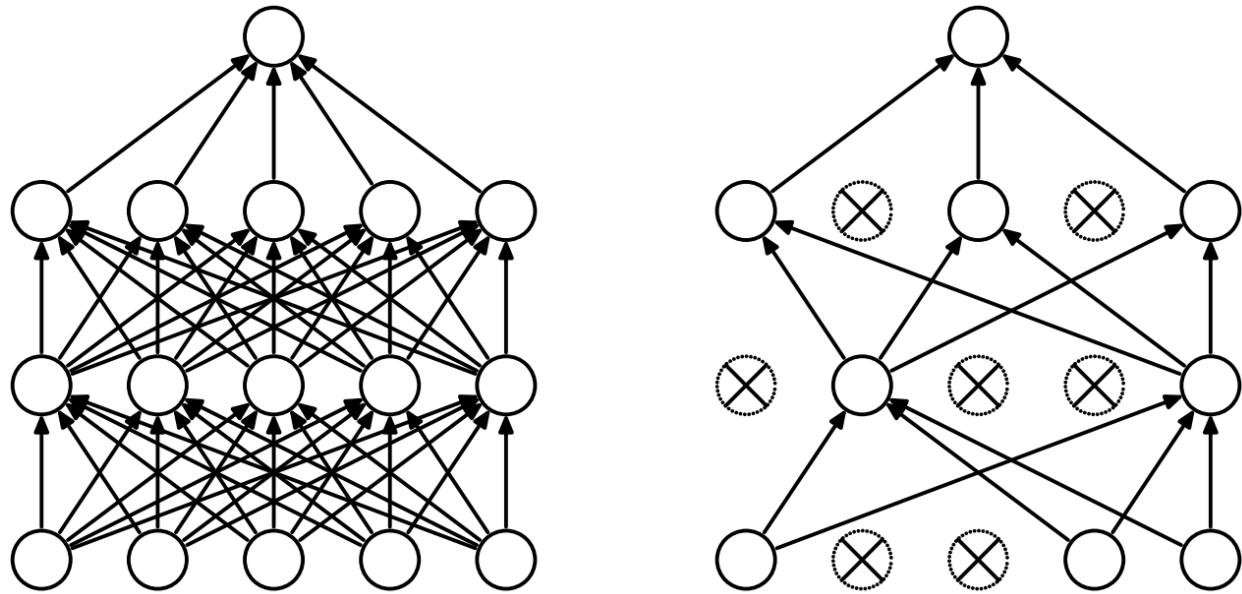


Figure 2.9: Dropout in a CNN

Contribution of technological advances related to deep learning

With the rise of Graphical Processing Units (GPU) in recent years, it has boosted the development of deep learning environments very rapidly. With the help of GPU we can now easily handle large amounts of image data and process multiple images parallelly. GPUs are extremely powerful for large matrix operations such as multiplication which helps in rendering the high resolution images.

GPUs can be very much difficult to use from a low level perspective using low level programming languages. This is why we have a collection of high level APIs and libraries available to make use of GPU's features for deep learning operations. Few very popular libraries are Tensorflow [21], Pytorch [22] etc. Libraries like Tensorflow coupled with Keras API [23] make it very easy and simple to build a CNN model and train as well as implement the complete model to give predictions.

Deep Learning would not have been successful in the field of medical imagery without the rise of computational resources like GPUs and also high level software, libraries and APIs like Tensorflow, Pytorch, Keras etc.

2.5 Summary

From the 1980s to give assistance to the radiologists, Computer Aided Diagnostic systems have been developed and it has been in use since then. Then during the 1990s these expert systems were being replaced by machine learning algorithms. The only problem with the machine learning systems were they require hand- extracted features from the mammogram images. These machine learning algorithms can not directly work with raw mammogram images. Performance of these algorithms are heavily dependent on the quality of features extracted from the images by means of some conventional image processing tools and methods. With the advancement in computational resources, deep learning and more specifically Convolutional Neural Networks (CNNs) has gained popularity recently which has the ability to automatically learn which features from the mammogram images are useful. While extracting features, it also preserves the spatial characters and properties of the mammogram image. However, the only drawback with deep learning is that it requires a large amount of data to avoid overfitting. It's very difficult to collect large amounts of data especially in medical imagery such as mammograms. But once we get a good enough quality of dataset and apply different regularisation techniques such as data augmentation and dropout, deep learning and CNN have the ability to easily outperform all the existing conventional methods, algorithms and systems for the tasks of early screening of breast cancer using mammogram datasets.

Chapter 3

Methodology: Design and Implementation

Design and Implementation of the Deep Learning model for early prediction of Breast Cancer

In this chapter based on the deep learning applications for early prediction of Breast Cancer covered in Chapter 2, we describe the dataset preparation, deep learning design pipeline and implementation for the different CNN models .

Dataset

3.1 Description of the mammogram datasets

The “Digital Database for Screening Mammography” (DDSM) dataset has been used in this project. The dataset is solely used for this Project/ Research purpose only. DDSM is an open-sourced public dataset.

3.1.1 DDSM

The “Digital Database for Screening Mammography” (DDSM) is a mammogram dataset which was released in 2007 by University of Florida for research purposes. It is an open- source database available online publicly. The DDSM dataset contains 2,620 scanned mammograms of benign, malignant and normal classes. All images of the DDSM dataset are stored in Lossless JPEG (LJPEG) image file format.

3.1.2 Data Source

In this project we are using the Mini-DDSM dataset which is a light weighted version of the “Digital Database for Screening Mammography” which is publicly available online in Kaggle website for deep learning research [24]. Just like the DDSM dataset, in the Mini- DDSM dataset, mammogram images are compressed in Lossless JPEG (LJPEG) image file format. The Mini-DDSM dataset that we have collected also comes with a metadata which has the following risk factors with it,

1. Age of the patient
2. Breast Density Score

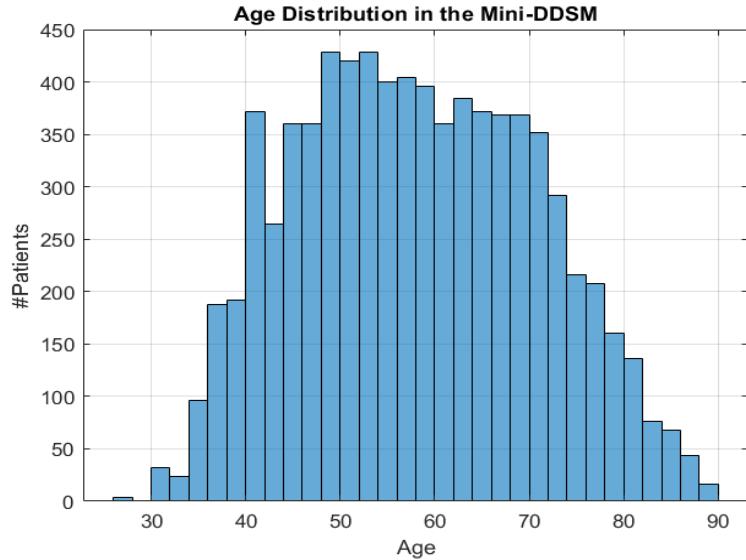


Figure 3.1: Age distribution in Mini- DDSM Dataset

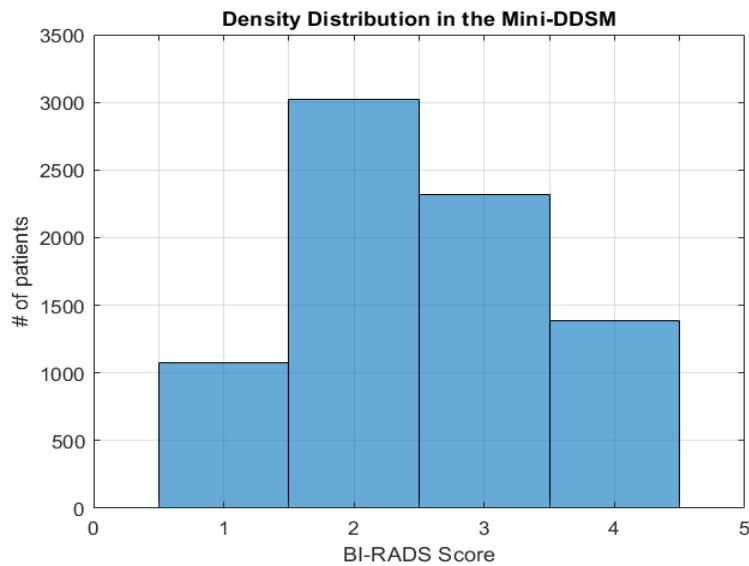


Figure 3.2: Density Distribution in Min- DDSM Dataset

Figure 3.1 illustrates the age distribution in the Mini- DDSM dataset. In Figure 3.2, we have shown the density distribution in the Mini- DDSM Dataset.

3.1.3 Risk factors

Risk factors are the additional information which comes along with mammogram images as a metadata. In the Mini- DDSM dataset we have a metadata which comes with a CSV file as shown in below figure 3.3

	A	B	C	D	E	F	G	H
1	fullPath	fileName	View	Side	Status	Age	Density	Tumour_Contour
2	Benign\0029\C_0029_1.LEFT_CC.png	C_0029_1.LEFT_CC.png	CC	LEFT	Benign	66	3	Benign\0029\C_0029_1.LEFT_CC_Mask.png
3	Benign\0029\C_0029_1.LEFT_MLO.png	C_0029_1.LEFT_MLO.png	MLO	LEFT	Benign	66	3	Benign\0029\C_0029_1.LEFT_MLO_Mask.png
4	Benign\0029\C_0029_1.RIGHT_CC.png	C_0029_1.RIGHT_CC.png	CC	RIGHT	Benign	66	3	-
5	Benign\0029\C_0029_1.RIGHT_MLO.png	C_0029_1.RIGHT_MLO.png	MLO	RIGHT	Benign	66	3	-
6	Benign\0033\C_0033_1.LEFT_CC.png	C_0033_1.LEFT_CC.png	CC	LEFT	Benign	60	3	-
7	Benign\0033\C_0033_1.LEFT_MLO.png	C_0033_1.LEFT_MLO.png	MLO	LEFT	Benign	60	3	-
8	Benign\0033\C_0033_1.RIGHT_CC.png	C_0033_1.RIGHT_CC.png	CC	RIGHT	Benign	60	3	Benign\0033\C_0033_1.RIGHT_CC_Mask.png
9	Benign\0033\C_0033_1.RIGHT_MLO.png	C_0033_1.RIGHT_MLO.png	MLO	RIGHT	Benign	60	3	Benign\0033\C_0033_1.RIGHT_MLO_Mask.png
10	Benign\0217\C_0217_1.LEFT_CC.png	C_0217_1.LEFT_CC.png	CC	LEFT	Benign	56	2	-

Figure 3.3: Metadata which include risk factor

This metadata contains two important risk factors which are Age of the patients and the Density score associated with the mammogram of the patients. However these risk factors are not yet ready to be used with the CNN model. We have to come up with some data cleaning and encoding techniques to convert these risk factor based metadata into feature vectors which we could then use in the fully connected layers right after the extraction of features from CNN layers.

3.2 Deep Learning pipeline

Deep Learning pipeline designed for the task of early prediction of breast cancer can be described in the following phases:

1. Dataset pre-processing: In this phase we do all the required pre-processing on the mammogram dataset by removing noise and artifacts etc. We also load the dataset in memory and process it further to get image labels for the mammogram classifications.
2. CNN Model Design: In this phase we build different CNN models such as VGG16, VGG19, ResNet50 which is followed by fully connected ANN layers to give predictions for three different classes of Benign, Malignant and Normal.

3. Training of the CNN model: In this phase we train the different CNN models with and without the inclusion of risk factors along with the mammogram images by means of training algorithms. We also fine tune the hyperparameters for best performance using a bag of tricks approach.
4. Final Prediction and Classification with DL Score: This phase is used for the final prediction of the mammogram image along with the associated DL Score for the corresponding classes. In the Malignant case the DL Score corresponds to the Risk of having Cancer and thus referred as DL Risk Score.

Following is the complete block diagram for the breast cancer prediction starting from model design, training and prediction.

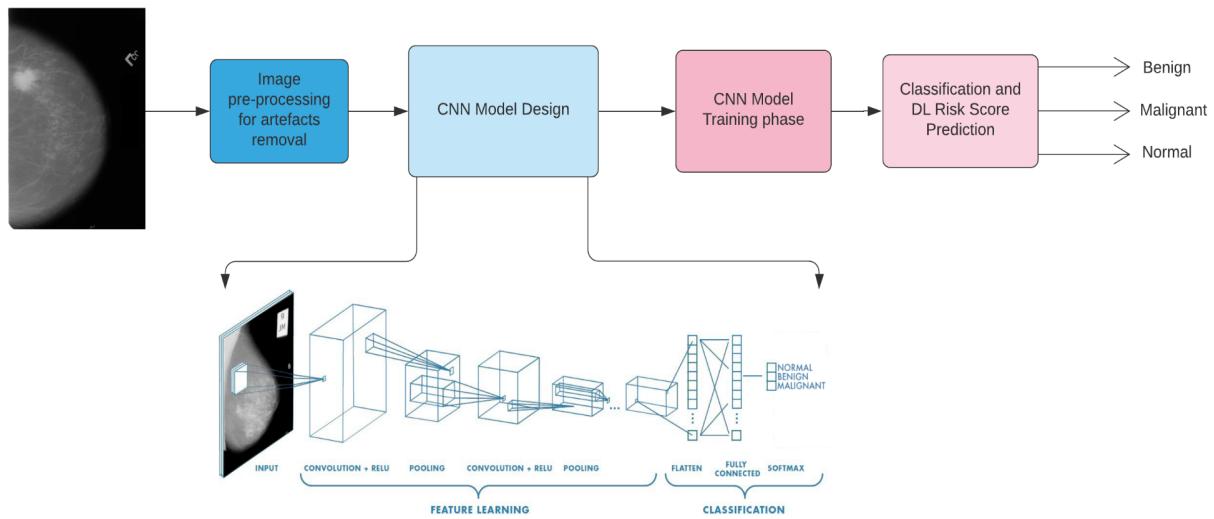


Figure 3.4: A complete block diagram of deep learning based breast cancer prediction system

3.2.1 Data Pre-Processing

Artefacts and noise removal from mammogram

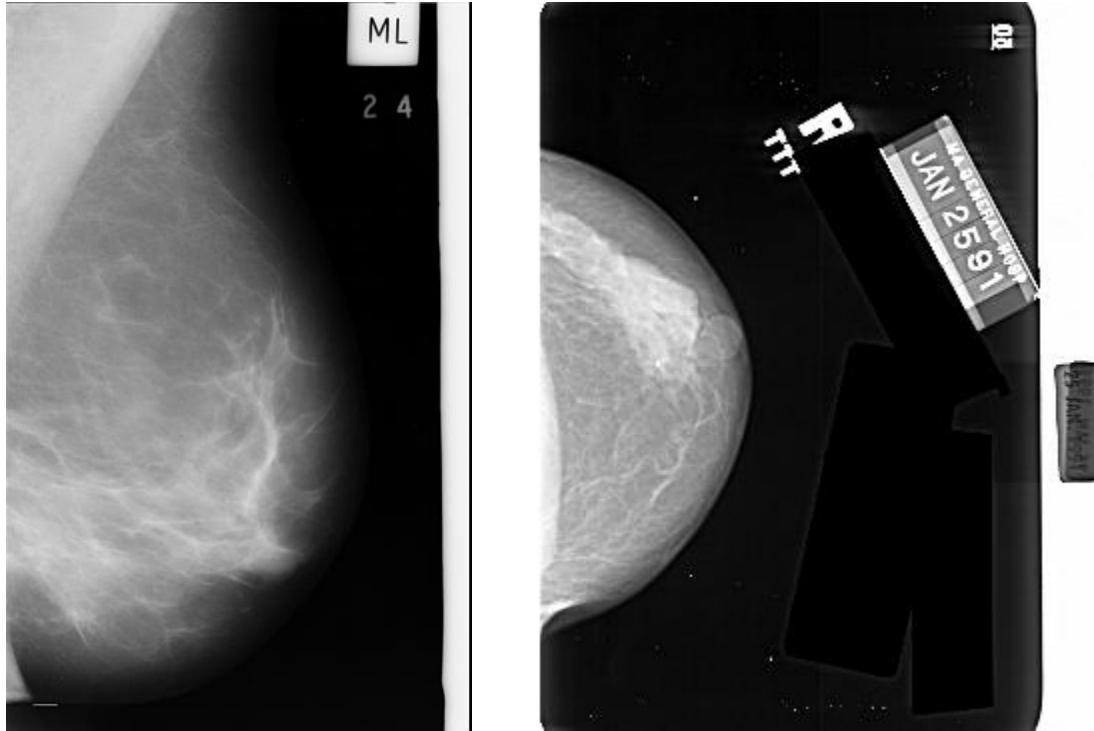


Figure 3.5: Artefacts present in the mammogram

In this step we have used various image processing techniques to remove the unwanted part present in the mammogram images. Various image processing techniques such as Wedge suppression, segmentation of the breast region from the background, area morphology [25] etc have been used for the artefacts removal from the mammogram images.

In Figure 3.6 we illustrate the different processes involved in cleaning the mammogram images using the above algorithms.

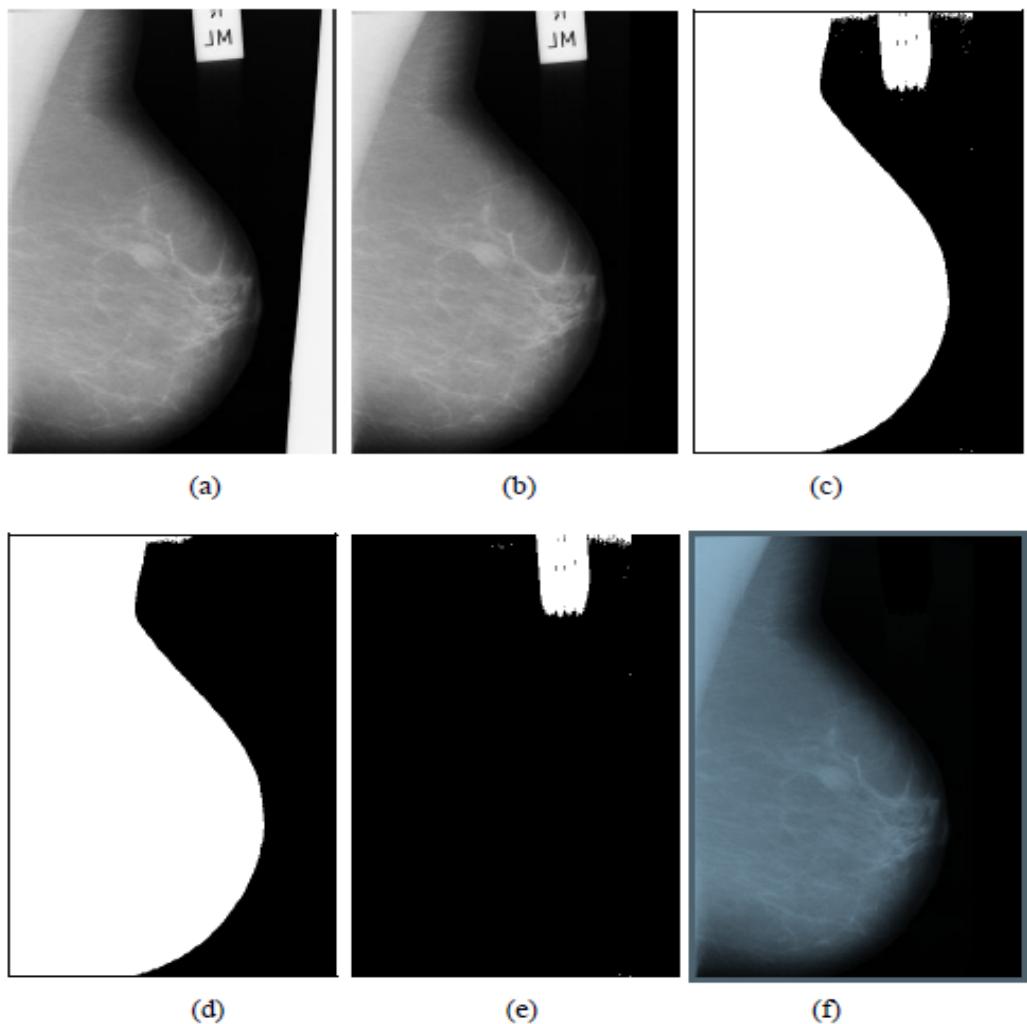


Figure 3.6: The process of artefacts removal: (a) Raw mammogram, (b) Removal of Wedges, (c) Segmentation of the breast region, (d) Binary masks of the breast region, (e) Binary masks of the artefacts and (f) Cleaned Mammogram image

Splitting the datasets

Datasets splitting is a very important part for any machine learning or deep learning model training. The datasets are basically splitted between a training set and validation/ test set to avoid any leakage of data while training.

As in machine learning or deep learning models, 80% training and 20% testing is often used which is seen in breast cancer detection papers [26]. During training only training set mammogram images are allowed to train the CNN model. Whereas, the test/ validation dataset images are kept aside to avoid any form of cheating to the learning model which we use while validation or may be testing the CNN model after it finishes training. Test/ validation set is basically used to evaluate the performance of the CNN model. In many scenarios in deep learning, apart from the training set we have a test and validation set separately. After the end of every epoch, validation dataset is used to validate the model. Validation dataset is helpful for hyperparameter tuning as well.

Data Loading

The Mini- DDSM dataset that we have used contained 2,196 mammogram images that has 1.2 GB of storage. The dataset is loaded in batches before it is fed to the CNN model for training.

Data Normalisation

In this step the mammogram images are resized and rescaled down so that all the images have the same size and to avoid inconsistent input sizes. The pixel intensities of the mammogram images found in the mini- DDSM dataset ranges from 0 to 255. But as the weights in the CNN network have smaller values and input images have larger values, it makes the training process slow and thus lead to decrease in overall accuracy of the CNN model. Therefore the pixel intensities are normalised by dividing them by 255 so that they range from 0 to 1. The following Figure 3.7 shows the pixel values of the mammogram before and after applying normalisation.

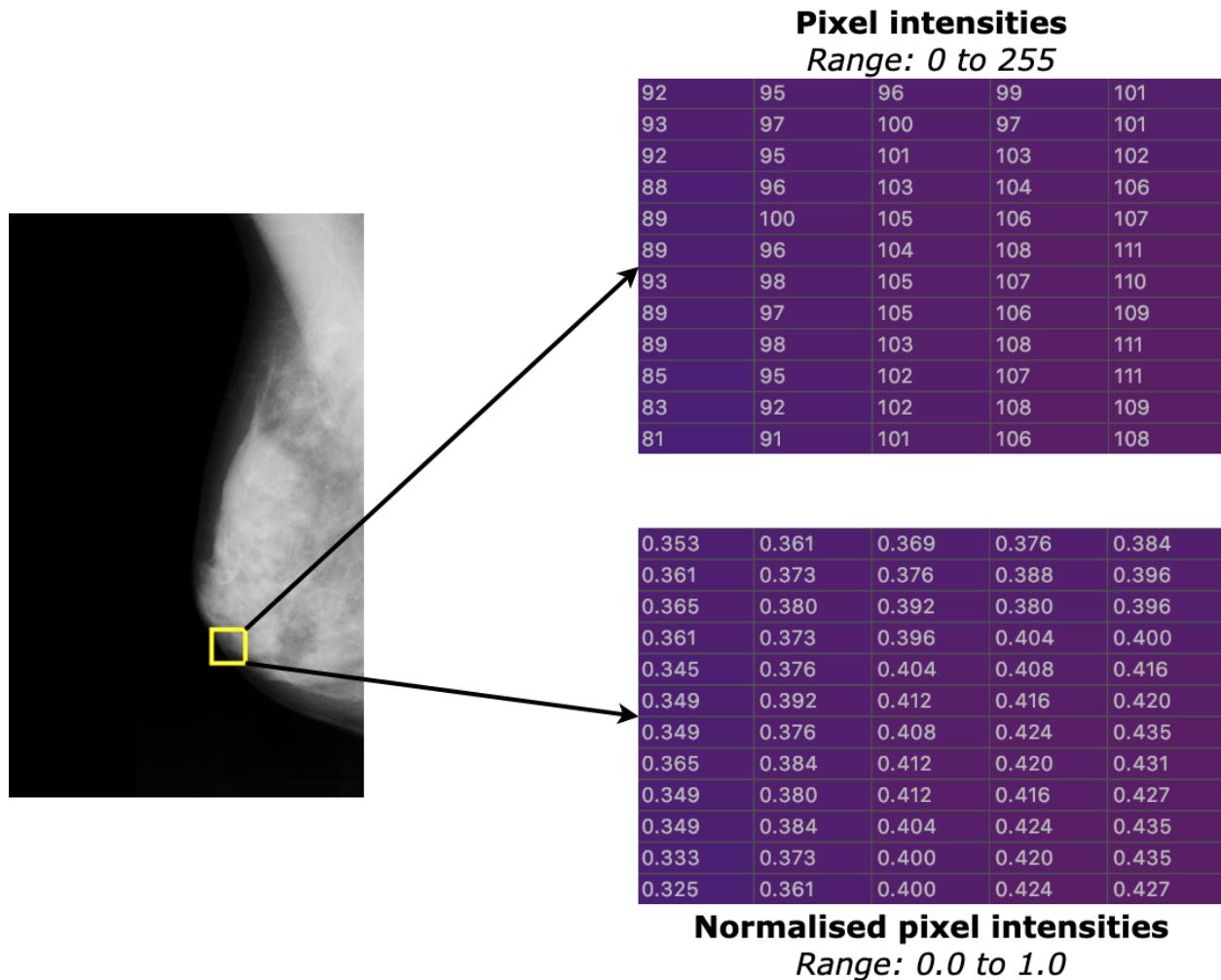


Figure 3.7: Image Normalisation for mammogram

Mammogram label encoding

All the three labels (Benign, malignant and normal) for the mammogram images are in categorical (string) format which we have to encode in numerical format so as to calculate the error and train the classification model. One- hot encoding is applied to the labels of Mini-DDSM mammogram dataset where each class is encoded into three numerical values either 1 or 0. The one- hot encoded labels are shown in the following table

Categorial (string) format	One- hot encoding
Benign	1 0 0
Malignant	0 1 0
Normal	0 0 1

Figure 3.8: Label encoding for mammogram classes

3.2.2 CNN Model Design

In this phase the pre- processed training dataset is fed to the CNN Model for training. The CNN model will learn the features from the pre- processed mammogram images and gives prediction which is then compared with the ground truth labels (Benign, malignant and normal) for calculating the error and update the kernel weights of the CNN layers and to finally predict the cases accurately after the end of the learning with the help of optimisation techniques when training phase is over.

Convolutional Neural Network Model

In this Project to train the Mini- DDSM dataset, state of the art CNN architectures are used rather than creating own custom CNN model from scratch. These pre-trained CNN Models are trained on very large datasets from ImageNet Challenge which has 1000s of classes. This approach or technique used in deep learning is known as transfer learning. Transfer learning technique is already proven to work with breast cancer mammogram image datasets or any medical imagery datasets [27].

There are many popular CNN architectures that can be used as a base CNN model before fully connected layers for the task of breast cancer prediction. We have used Keras library on top of Tensorflow to build these base CNN models using the already well established CNN architectures. These CNN architectures include VGG16, VGG19 and ResNet50. Keras library has already come up with these CNN models. The fully connected layers of these CNN models designed for the ImageNet Challenge had 1000s of different classes at the output neurons. Hence, for our task of breast cancer prediction, we dropped the last layers and replaced our own output layers which have three classes (Benign, Malignant and Normal).

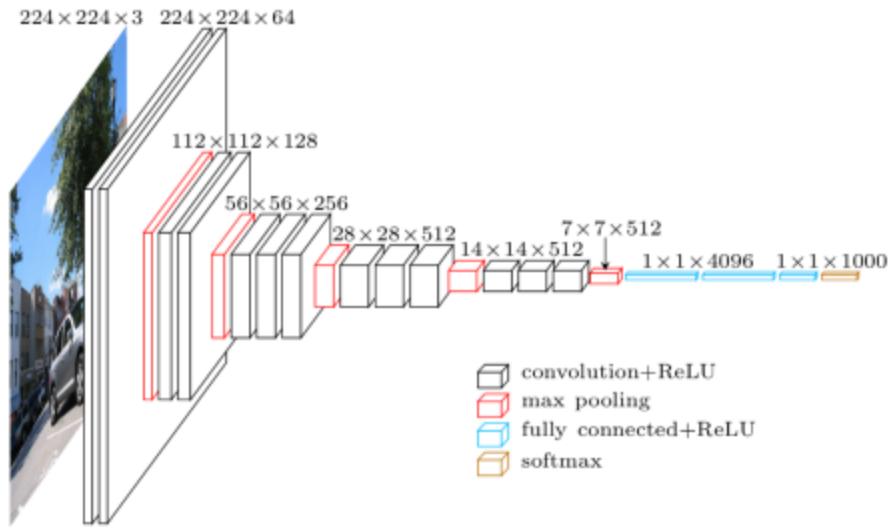


Figure 3.9: VGG16 Architecture

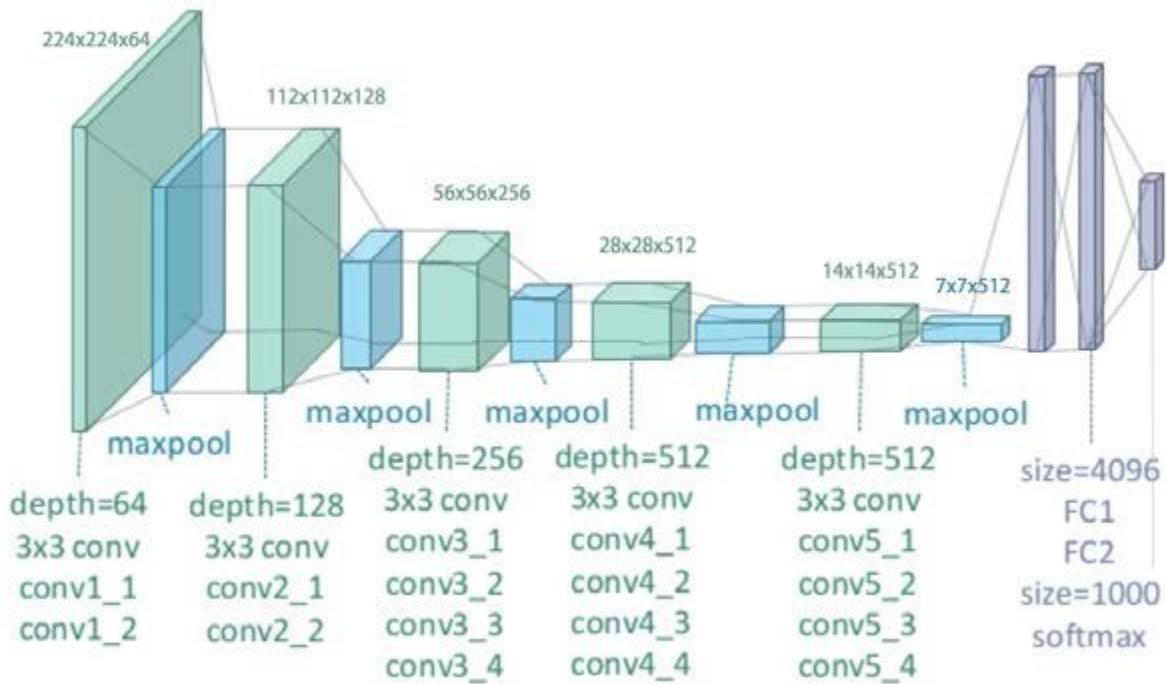


Figure 3.10: VGG19 Architecture

34-layer residual

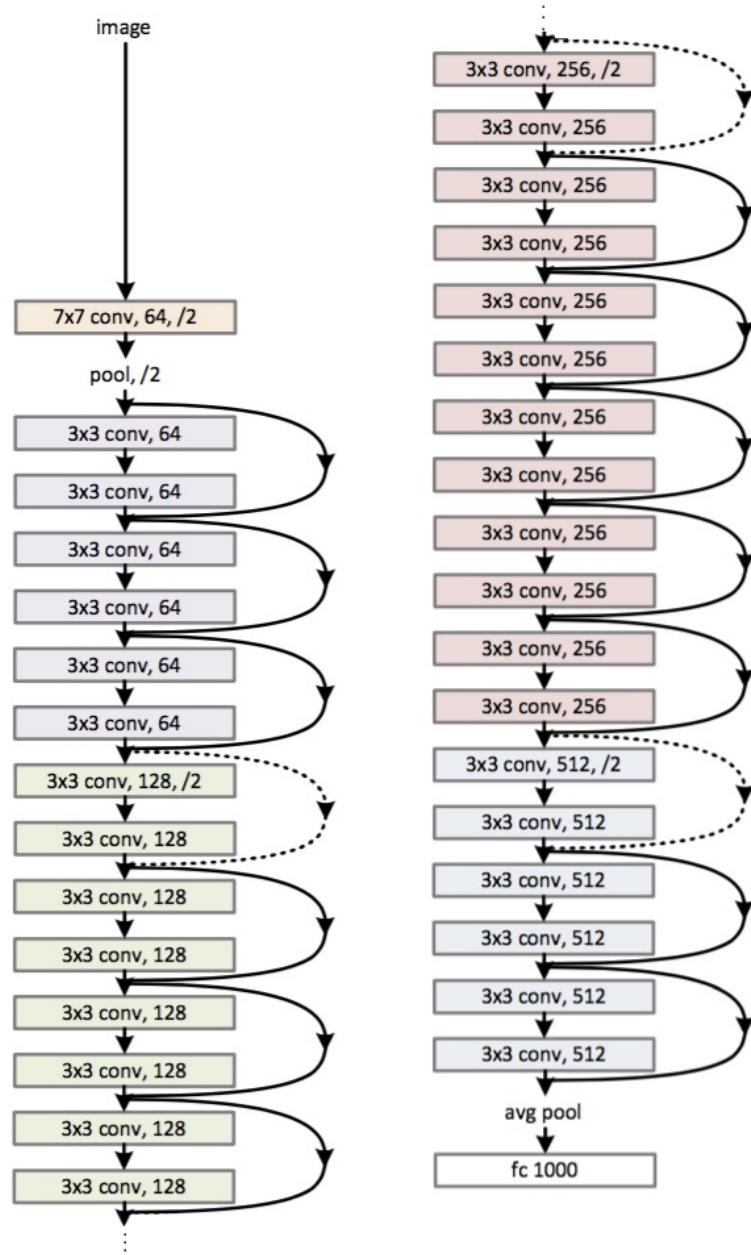


Figure 3.11: ResNet50 Architecture with Skip connections

Activation functions

There are many types of activation functions used in deep learning or CNN. The CNN models such as VGG or ResNet generally use ReLu activation functions in their hidden layers. In our CNN model at the output of the CNN model since there are three classes so we picked softmax

activation function which is often used for multi- class classification tasks. Softmax activation gives probability for each class which sums up to 1. The probability of each class depends on each other as the mammogram must belong to a single class only either Benign, malignant or normal. If there had been only two classes, we would have used sigmoid activation function instead of softmax activation function. Hence, the output nodes of the CNN model for Mini-DDSM dataset will use softmax activation to classify the mammogram and give early prediction for breast cancer.

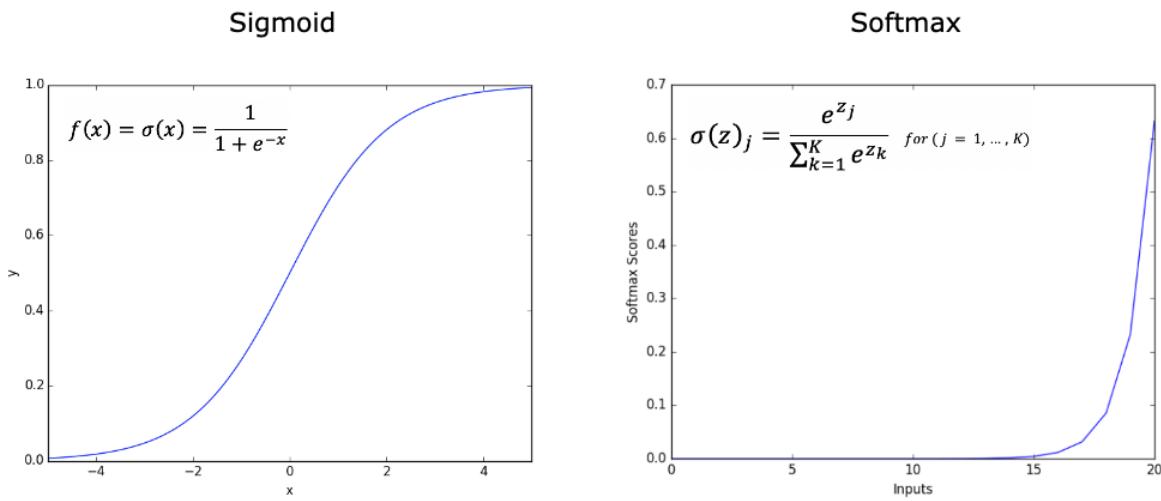
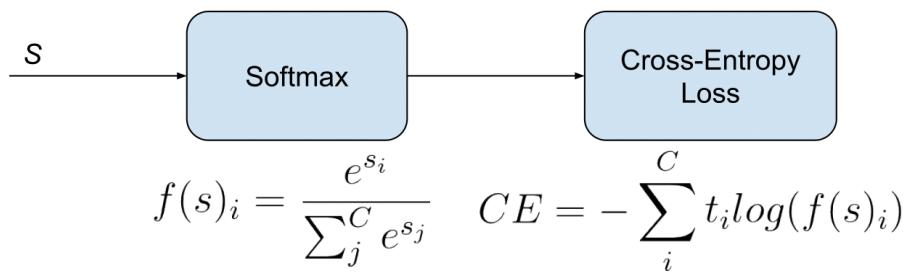


Figure 3.12: Sigmoid and Softmax activation functions

Cost Function

Cross entropy is the most popular cost function that has been used for binary and multi- class classification tasks and hence in breast cancer classification also we can use cross entropy loss function. Since in our case we have multi- class classification problem so we used categorical cross entropy cost function which is given by the following equation



Here, s is the value of the output neuron before applying softmax activation, $f(s)_i$ is the final output that has been predicted, c is the total number of classes which is three in our case, CE stands for Cross Entropy which is the cost function that we have used. t_i is the actual ground truth label corresponding to the i^{th} example.

Cross entropy is an ideal choice for cost function in this case because the predictions are given by the probabilities with the help of softmax activation function.

Optimiser

It is the most important phase of the deep learning model. Optimisers in any deep learning CNN model are used to minimize the cost function and control the model by using hyperparameters. There are many optimisers like SGD (Stochastic Gradient Descent), Adam (Adaptive moment estimation), RMSProp etc. SGD is a traditional optimiser which converges slowly. It also requires more fine tuning. Adam (Adaptive moment estimation) is the most used optimiser. Adam optimiser is the best choice as an optimiser for our CNN model.

Transfer Learning Pre-trained weights

Transfer Learning technique is implemented with the help of weights from ImageNet. All the layers from the base model architecture (VGG, ResNet) are frozen and only the fully connected layers of the CNN are allowed to train. The training phase for the breast cancer prediction model ends when the maximum number of epochs is achieved or the early stopping criteria is fulfilled.

Validation of the CNN model

To make the model generalise well for the unseen data, the training dataset is further divided to form a validation set with a split ratio of 75% and 25% respectively. By calculating the loss and accuracy, the validation dataset is used to make the prediction at the end of every epoch.

Early Stopping on training phase

Before the maximum number of epochs is reached the training of the CNN model is stopped by monitoring the validation accuracy or loss if it does not improve after a certain number of

epochs. Early stopping prevents the model from overfitting and also reduces the computational costs.

3.3 Implementation of CNN model training phase

3.3.1 Initial Dataset Pre-processing

When downloaded from the source website, the mammogram datasets come with two master folders. One folder contains the raw mammogram images with three subfolders given by their classes(Benign, malignant and Normal). We organised the folders to get training and validation folders where each folder consists of three sub- folders with file name “Benign”, “Malignant” and “Normal” respectively. The dataset also come up with a CSV file as metadata along with the mammogram images of following type which contains certain risk

fullPath	fileName	View	Side	Status	Age	Density
Benign\0029\C_0029_1.LEFT_CC.png	C_0029_1.LEFT_CC.png	CC	LEFT	Benign	66	3
Benign\0029\C_0029_1.LEFT_MLO.png	C_0029_1.LEFT_MLO.png	MLO	LEFT	Benign	66	3
Benign\0029\C_0029_1.RIGHT_CC.png	C_0029_1.RIGHT_CC.png	CC	RIGHT	Benign	66	3
Benign\0029\C_0029_1.RIGHT_MLO.png	C_0029_1.RIGHT_MLO.png	MLO	RIGHT	Benign	66	3
Benign\0033\C_0033_1.LEFT_CC.png	C_0033_1.LEFT_CC.png	CC	LEFT	Benign	60	3
Benign\0033\C_0033_1.LEFT_MLO.png	C_0033_1.LEFT_MLO.png	MLO	LEFT	Benign	60	3
Benign\0033\C_0033_1.RIGHT_CC.png	C_0033_1.RIGHT_CC.png	CC	RIGHT	Benign	60	3
Benign\0033\C_0033_1.RIGHT_MLO.png	C_0033_1.RIGHT_MLO.png	MLO	RIGHT	Benign	60	3
Benign\0217\C_0217_1.LEFT_CC.png	C_0217_1.LEFT_CC.png	CC	LEFT	Benign	56	2
Benign\0217\C_0217_1.LEFT_MLO.png	C_0217_1.LEFT_MLO.png	MLO	LEFT	Benign	56	2

Figure 3.13: Metadata of Mini- DDSM

This format is not usable in our deep learning model directly. We had to do some data analysis and do some pre- processing to clean it and apply one-hot encoding to finally convert it to some useful dataset of risk factors which is of the following type

Age < 50	Age > 50	Density = 1	Density = 2	Density = 3	Density = 4
0	1	0	0	1	0
0	1	0	1	0	0

Figure 3.14: Cleaned Dataset of Risk Factor (Only few examples)

Here in this dataset each row denotes risk factors for a training example and each column represents a particular risk factor. These feature vectors are used along with the mammogram images while training the CNN model. Risk factors used here give additional information of the mammogram status and help increase the prediction of the mammogram and help increase the prediction for the early instances of breast cancer.

3.3.2 Mammogram image resizing

The mammogram images from the Mini- DDSM dataset are imported using numpy data loader and converted to numpy array for fast processing. The numpy image array then resized to 224 x 224 so that it can be used as an input to the CNN architecture such as VGG16, VGG19 and ResNet50. 224 x 224 is chosen as the input size because the base CNN architectures were pre-trained on datasets having input size 224 x 224. The mammogram image numpy array after resizing is also normalised which is done by dividing each pixel of the mammogram images by 255.

3.3.3 Label encoding for the mammogram classes

As in our project we are solving a multi-class classification problem so we are using ‘Label Encoder’ from Scikit- Learn library to one- hot encode the labels of the mammogram images

3.3.4 Data Augmentation

Since we are using a Mini- DDSM dataset and we have only 2,196 mammogram images, our model might suffer from overfitting. So, to overcome this we apply different image data augmentation techniques such as vertical flip, horizontal flip, shear, zoom, rotation etc with the help of Image data generator function available on Tensorflow/ Keras.

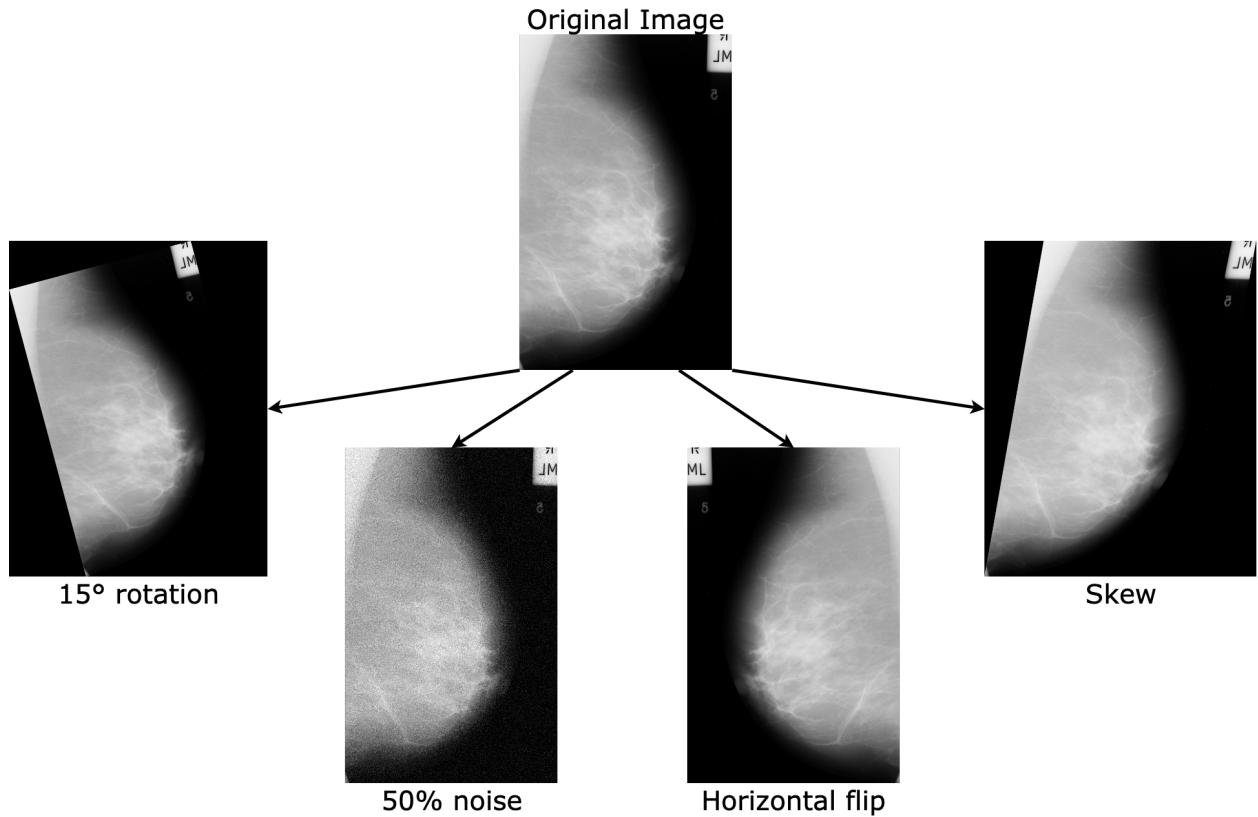


Figure 3.15: Example of different image transformation as a part of data augmentation technique

3.3.4 CNN Model Training

Risk Factor based RF-LR Model: In this model we use the available risk factors that comes with the Mini- DDSM dataset which are "Age at mammography" and "Breast Density" (1,2,3,4). We simply built the Logistic Regression model based on these two risk factors. We trained the RF-LR model based on the risk factors.

Image Only Model

Here we use only the mammogram images alone to train the different CNN models (VGG16, VGG19, ResNet50). We used the sequential model from Keras API in order to build the CNN. As per the transfer learning strategy we also include the weights as “ImageNet”. Following are the sequential model configuration for the base CNN models developed in Tensorflow/ Keras

```
base_model = VGG16(input_shape = (224, 224, 3), weights = 'imagenet', include_top = False)
model = Sequential()
model.add(base_model)
model.add(Dropout(0.5))
model.add(Flatten())
model.add(BatchNormalization())
model.add(Dense(128, kernel_initializer = 'he_uniform'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(Dropout(0.5))
model.add(Dense(3, activation = 'softmax'))
for layer in base_model.layers:
    layer.trainable = False
```

Similarly VGG19 and ResNet50 can be developed with a similar configuration using Tensorflow/ Keras.

```
base_model = VGG19(input_shape = (224, 224, 3), weights = 'imagenet', include_top = False)
model = Sequential()
model.add(base_model)
model.add(Dropout(0.5))
model.add(Flatten())
model.add(BatchNormalization())
model.add(Dense(128, kernel_initializer = 'he_uniform'))
model.add(BatchNormalization())
```

```
model.add(Activation('relu'))  
model.add(Dropout(0.5))  
model.add(Dense(3, activation = 'softmax'))  
for layer in base_model.layers:  
    layer.trainable = False
```

Following Figure 4.14 and Figure 4.15 are the structures of the VGG16 and VGG19 CNN models respectively that we have used in the breast cancer prediction model. Similarly on Figure Figure 4.16 we have shown the ResNet50 architecture as a sequential model in Keras.

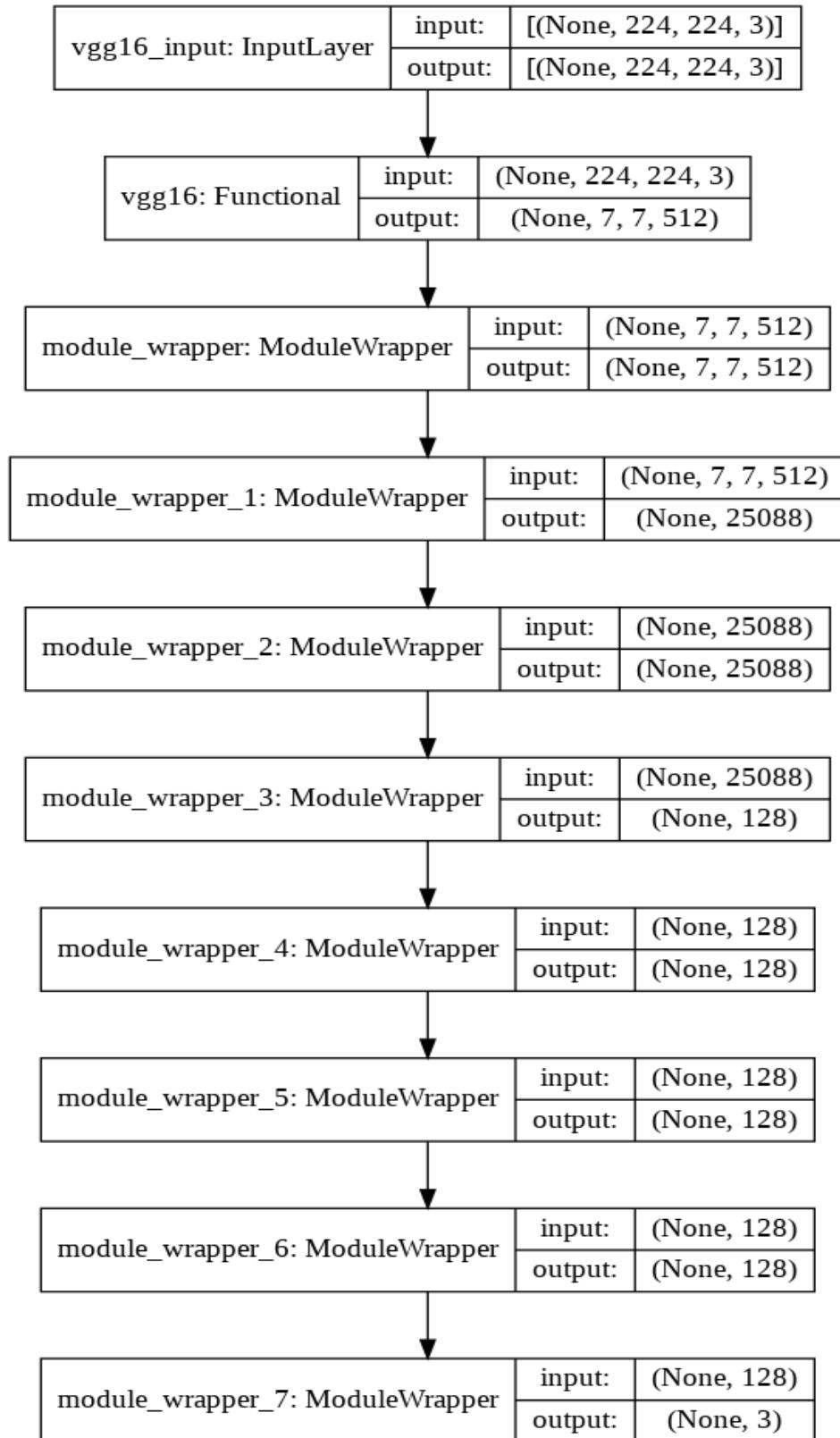


Figure 3.16: VGG16 CNN Architecture for breast cancer prediction

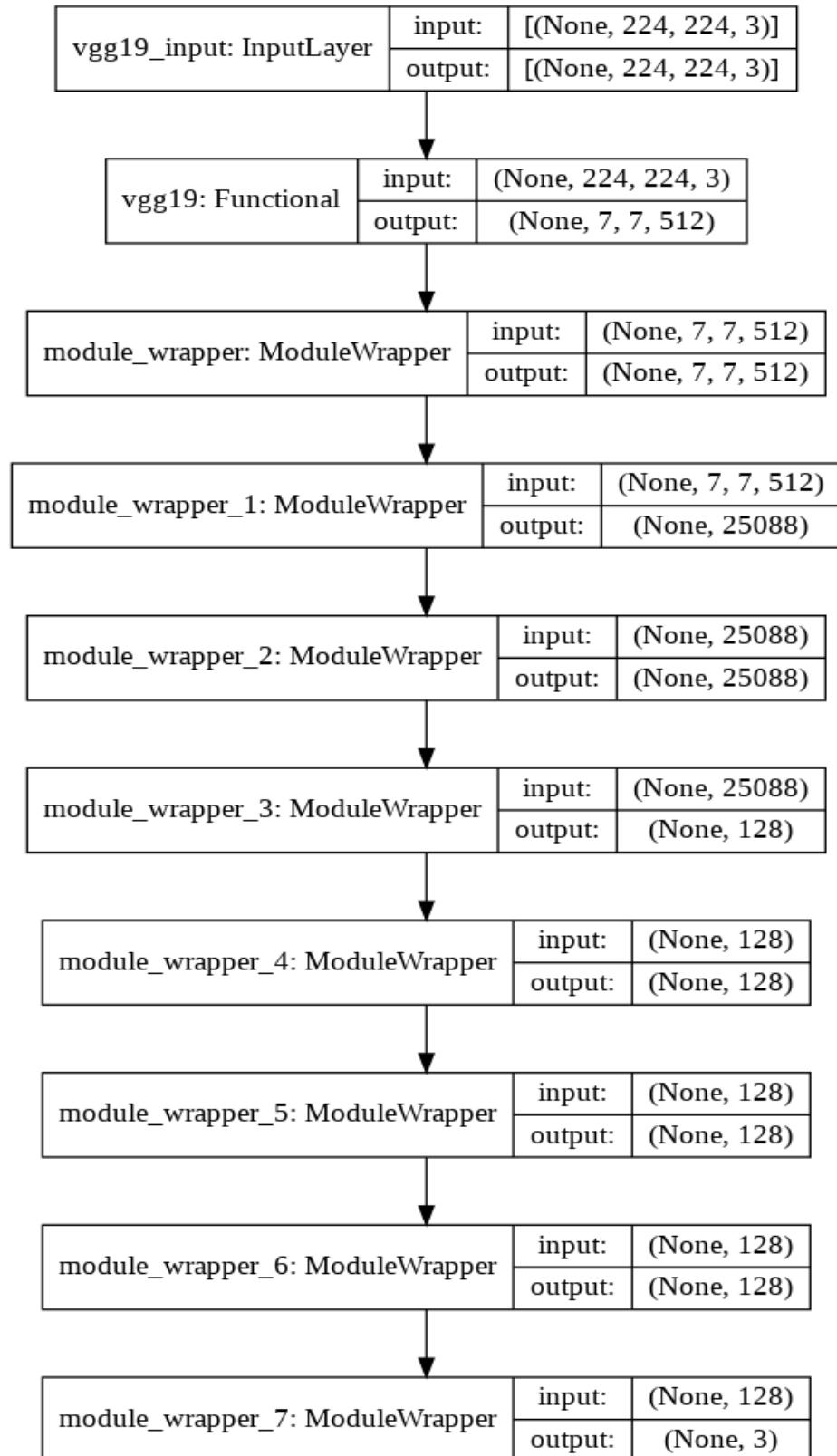


Figure 3.17: VGG19 CNN Architecture for breast cancer prediction

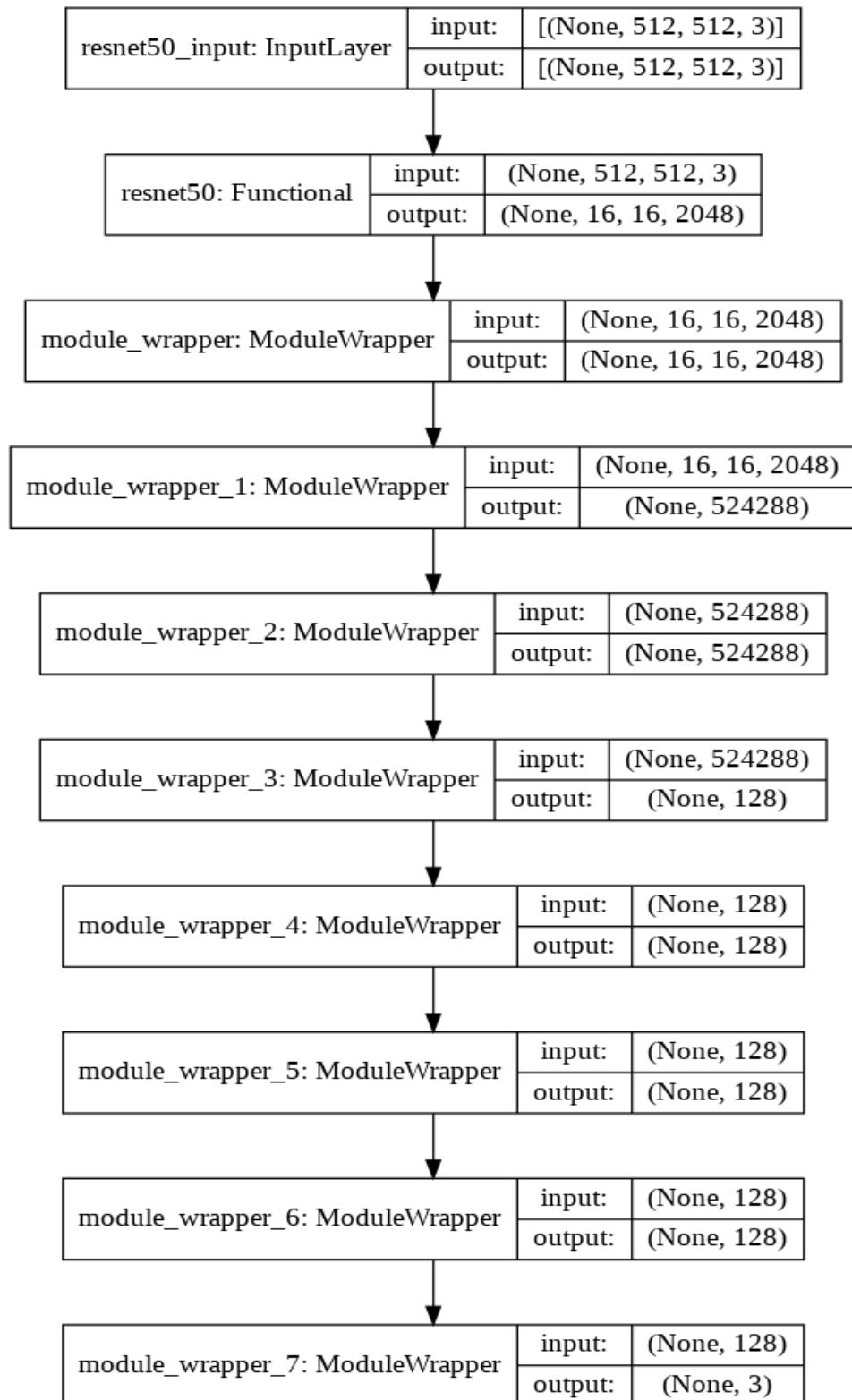


Figure 3.18: ResNet50 Architecture for breast cancer prediction

With the help of Keras API we could include all the three CNN architectures to build our CNN model for predicting early instances of breast cancer. The fully connected layers are dropped which is done by adding `include_top = False` so that after learning and extracting the features from the mammogram with the help of pre-trained weights from ImageNet, we could make use of fully connected layers to train the extracted feature which will give prediction for early instances of breast cancer.

Hybrid Model (Image with Risk Factors)

In this case, we make use of the risk factors which we have as a metadata along with the mammogram images. We have two risk factors available with us as discussed earlier and these are again one-hot encoded to form the following feature vectors.

Age < 50	Age > 50	Density = 1	Density = 2	Density = 3	Density = 4
0	1	0	0	1	0
0	1	0	1	0	0

Figure 3.19: Risk factors as a feature vectors

In feature vectors of risk factors we have six features corresponding to each mammogram. These features are one-hot encoded. The features include the following

1. Age < 50
2. Age > 50
3. Density = 1
4. Density = 2
5. Density = 3
6. Density = 4

We include these feature vectors after the base CNN model. The size of the total Risk factor based feature vectors is 2196 x 6. Where 2196 denotes the number of mammogram images and 6

denotes the number of features. These feature vectors are added right after the base CNN model where the feature maps are flattened.

We made use of Keras functional API as follows in Figure 4.17 to concat the feature vectors for the corresponding mammogram image. These new feature vectors are added with the feature extracted by the base CNN model from mammogram images.

```
base_model = VGG16(include_top = False, input_shape = (224, 224, 3))
model1 = Dropout(0.5)(base_model.layers[- 1].output)
flat = Flatten()(model1)
model2 = concatenate([flat, aux_input], axis = - 1)
model3 = BatchNormalization(axis = 1)(model2)
model4 = Dense(128, activation = 'relu')(model3)
model5 = BatchNormalization(axis = 1)(model4)
model6 = Dropout(0.5)(model5)
out = Dense(3, activation = 'softmax')(model6)
model = Model(inputs = [base_model.inputs, aux_input], outputs = [out])
for layer in base_model.layers:
    layer.trainable = False
```

In the following Figure 4.18 we show the complete CNN architecture of VGG16 with the inclusion of Risk factor based feature vector discussed earlier.

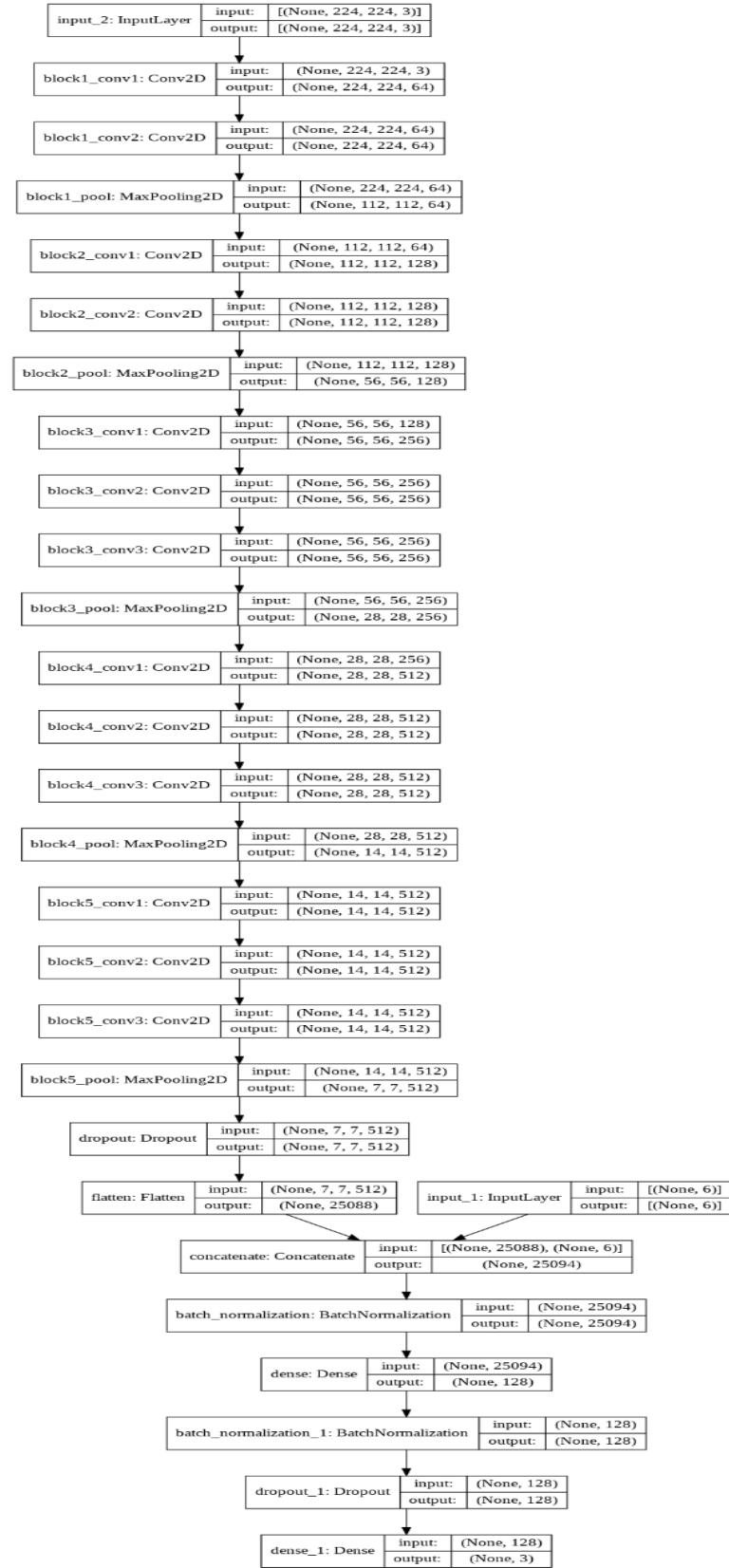


Figure 3.20: VGG16 model architecture with the inclusion of Risk factor vector

3.3.5 Training Steps

In the training phase the maximum number of epochs and learning rate is set to 100 and 0.001 respectively. There are few other parameters of the optimiser which are set as follows

```
optimizer = Adam(lr = 0.001, beta_1 = 0.9, beta_2 = 0.999, epsilon = 0.1, decay = 0.0)
model.compile(loss = 'categorical_crossentropy', optimizer = optimizer, metrics = ['accuracy'])
```

3.3.6 CNN Model and Weights saving

Once the training phase of the CNN model is over we save the configuration of the model, hyperparameters, optimizers, weights etc using HDF5 file format which can be done in Tensorflow Keras with the help of following function

```
from tensorflow import keras
model = keras.models.load_model('path/to/location')
```

These saved models and their weights can be loaded anytime when we need to make predictions on a test or validation dataset. We need not to train the CNN model again and again.

3.4 Summary of CNN Model Design and Training Phase

1. Dataset used: Mini- DDSM (Digital Dataset for Screening Mammography)
2. Dataset Pre- processing: 75% is used for Training and 25% is used for Testing/ Validation
3. CNN model parameters:
 - (a) Base Models: VGG16, VGG19 and ResNet50
 - (b) Activation Function at Output Nodes: Softmax
 - (c) Loss Function/ Cost Function: Categorical Cross Entropy
 - (d) Optimiser: Adam
 - (e) Transfer Learning: Pretrained weights from base CNN models which were trained on ImageNet Challenge Dataset

Chapter 4

Results

The performance of the CNN model that we have developed for the early prediction of breast cancer is evaluated using various statistical evaluation parameters like accuracy, precision, recall and F1 Score.

Here in this chapter we also discuss the hardware and software requirements, programming language and deep learning libraries and frameworks used in the Project.

4.1 Experimental Setup

4.1.1 Deep Learning Framework

Programming Language

Python is used for this project to implement the CNN model as Python is one of the popular programming languages and also availability of libraries and frameworks for deep learning makes it the obvious choice for this Project. To avoid manual implementation of many common machine learning functionalities and image processing, visualization techniques etc Python is chosen. Python has got some really good deep learning libraries such as SciKit- Learn [29], Tensorflow, Keras, Pandas [30], Matplotlib [31], Numpy [32].

Choice of the Deep Learning Framework

Due to the complexity of implementing the deep learning model and looking at the nature of the mammogram dataset, it is preferred to use a powerful computing resource which is in the form of Graphical Processing Units (GPU). An Nvidia Tesla P100 32 GB GPU enabled workstation is used for designing, implementing and testing the CNN models.

To make use of the power of GPU, deep learning frameworks are used which work with CUDA (to provide parallel computing and GPU optimisation) and provide CNN support and pre-trained models. Tensorflow coupled with Keras and Pytorch are two most popular deep learning frameworks used nowadays. Tensorflow/ Keras is relatively older than Pytorch and also has more online support. Hence for this Project of early prediction of breast cancer, we decided to go

with Tensorflow coupled with Keras to build the breast cancer classifier based on Convolutional Neural Network.

4.2 CNN Model trainable parameters

VGG16

VGG16 CNN model that we are using for breast cancer prediction has total 18,027,331out of which 3,262,211parameters are trainable and rest 14,765,120are non-trainable as we are fixing those weights using a pre-trained VGG16 model.

VGG19

VGG19 CNN model that we are using for breast cancer prediction has total 23,337,027out of which 3,262,211parameters are trainable and rest 20,074,816are non-trainable as we are fixing those weights using a pre-trained VGG19 model.

ResNet50

VGG16 CNN model that we are using for breast cancer prediction has total 36,835,203out of which 13,046,531parameters are trainable and rest 23,788,672are non-trainable as we are fixing those weights using a pre-trained ResNet50 model.

4.3 Visualisation of results

Accuracy

Accuracy is only a good evaluating parameter for the breast cancer prediction model when the dataset is balanced. When the number of mammogram images in each class of Benign, Malignant and Normal are not balanced, that is all three classes have different numbers of images then the accuracy could mislead the evaluation of the CNN model. The accuracy of a model is given by

$$\text{Accuracy} = \frac{TP + TN}{P + N}$$

Where, TP stands for True Positive

TN stands for True Negative

P stands for positive cases

N stands for negative cases

If let say, our model is predicting the mammogram image as normal always then it would still achieve as high as 64% of accuracy with Mini- MIAS dataset despite never predicting any abnormal classes [28]. So, there is a need for other additional evaluation parameters then just accuracy.

Precision

Precision is the probability of the model predicting the positive cases correctly given that the model predicted positive cases. It is given by the following equation

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

Where, TP stands for True Positive

FP stands for False Positive

Recall

Recall is the probability of the model predicting the positive cases correctly given all the actual positive cases. It is given by the following equation

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

Where, TP stands for True Positive

FN stands for False Negative

F1 Score

To combine both the features of Precision and Recall, F1 Score is used. To get a high F1 Score both the Precision and Recall should be high. F1 Score is given by the following equation

$$F1\ Score = \frac{TP}{TP + \frac{FN + FP}{2}}$$

Confusion Matrix

Confusion Matrix is a very popular visual metric plot made for each predicted class. In a confusion matrix each row denotes the actual ground truth labels and each column corresponds to the model's prediction. Confusion matrix is beneficial to inspect which classes are detected correctly the most number of times and which predicted classes being misclassified.

DL Score

DL Score is the probability given by the trained CNN model for each class. It denotes the confidence score of the model to classify the cases on mammograms. The softmax activation which is present at the output of the CNN model results in probability for each of the classes and gives the DL score. For cancer cases we refer to this DL score as DL Risk Score.

ROC

The Receiver Operator Characteristic or commonly known as ROC is an evaluation metric for classification problems. ROC is a probability curve that plots the True Positive Rate against False Positive Rate at different threshold values. The Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and it is calculated by measuring the area under the ROC curve. In our project with the help of Sklearn library we draw the ROC and calculate the AUC from the ROC curve by importing the following libraries

```
from sklearn.metrics import roc_curve, auc
```

4.4 Hyperparameters tuning

In this phase a set of approaches are used and various experiments are done to fine tune the hyperparameters used in the CNN model. After using and experimenting with different parameters using bag of tricks approach, we fixed these parameters as follows for the best performance of the CNN model for prediction

CNN Model	input image size	Transfer learning	Optimiser	Learning rate	Cost function	No. of epochs
VGG16	224x224	yes	Adam	0.001	Cross Entropy	100
VGG19	224x224	yes	Adam	0.001	Cross Entropy	100
ResNet50	224x224	yes	Adam	0.001	Cross Entropy	100
VGG16 with Risk factors	224x224	yes	Adam	0.001	Cross Entropy	100

Figure 4.1: Hyperparameters of different CNN models

4.5 Performance of different models

For the LR-RF model we used two risk factors which are Age at mammography and Density of Breast. 7,796 total cases are considered with the two risk factors of which 2680 are Benign cases, 2,708 are Malignant cases and the rest 2408 are Normal cases. The dataset is fairly balanced for any machine learning algorithm to work on. Using the LR-RF model which was developed using the Risk factors (Age at mammography and Breast Density) with the help of Logistic Regression we achieved an overall accuracy of 41\%. with Precision of 0.41, 0.41, 0.59 for Benign, Malignant and Normal cases respectively, Recall of 0.59, 0.61, 0.05 for Benign, Malignant and Normal cases respectively and similarly F1 Score of 0.48, 0.49, 0.09 for the given three cases.

Now, we discuss all the experimental results which are performed on the Mini- DDSM dataset for early prediction of Breast Cancer. We used a validation dataset for the evaluation of the CNN model. The Mini- DDSM dataset that we have use has the following number of mammogram images for validation as well as testing of the CNN model

Benign = 219

Malignant = 222

Normal = 219

In Image based CNN models (Image Only and Hybrid), only the base CNN architecture, risk factors dependency is varied. All other parameters given below remain constant across all the experiments

1. Fully Connected layers with 512 and 32 hidden neurons and 3 neurons at the output
2. Dropout factor = 0.5
3. Input dimension of the model = 224X224
4. Data Augmentation factors
5. Cost Function = Categorical Cross Entropy
6. Learning Rate = 0.001
7. Optimizer = Adam

4.5.1 Performance on different models for early prediction of breast cancer

Models	Overall Accuracy	Precision	Recall	F1 Score
LR-RF	0.44(95% CI: 0.343,0.537)	0.44, 0.41, 0.62	0.62, 0.57, 0.08	0.52, 0.48, 0.14
VGG16	0.88(95% CI: 0.816,0.944)	0.87, 0.88, 0.89	0.86, 0.86, 0.92	0.86, 0.87, 0.91
VGG19	0.85(95% CI: 0.780,0.920)	0.83, 0.85, 0.85	0.84, 0.82, 0.89	0.84, 0.83, 0.87
ResNet50	0.61(95% CI: 0.514,0.706)	0.61, 0.56, 0.65	0.44, 0.57, 0.82	0.51, 0.57, 0.72
VGG16 with Risk Factors	0.88(95% CI: 0.816,0.944)	0.86, 0.89, 0.90	0.88, 0.88, 0.89	0.87, 0.88, 0.89

Figure 4.2: Performance of different CNN models for breast cancer early prediction

Image Only Model (VGG16)

Using VGG16 we achieved an overall accuracy of 88%. with Precision of 0.87, 0.88, 0.89 for Benign, Malignant and Normal cases respectively, Recall of 0.86, 0.86, 0.92 for Benign,

Malignant and Normal cases respectively and similarly F1 Score of 0.86, 0.87, 0.91 for the given three cases. The training phase progression of this CNN Model can be visualized by the following plots of Accuracy vs Number of epochs and Loss vs Number of Epochs for Training and Validation set.

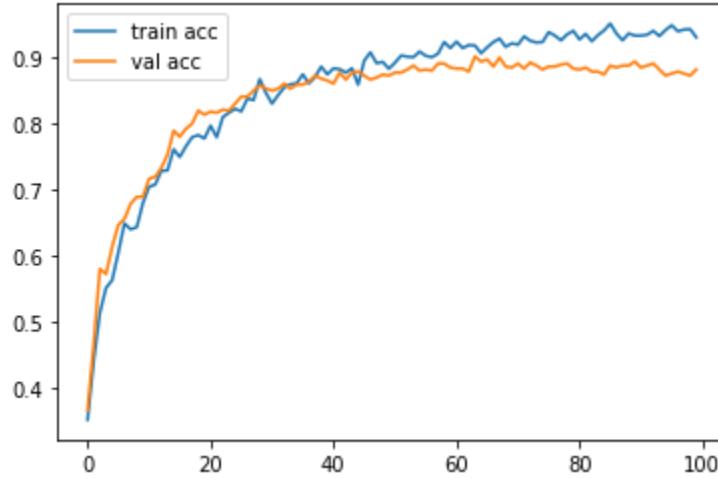


Figure 4.3: Accuracy vs No. of Epochs (VGG16 model)

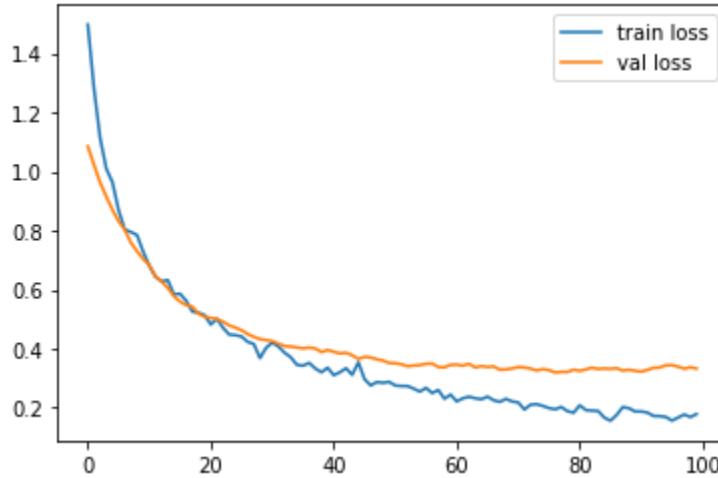


Figure 4.4: Loss vs No. of Epochs (VGG16)

Following in the Figure, we have shown the confusion matrix for the VGG16 model

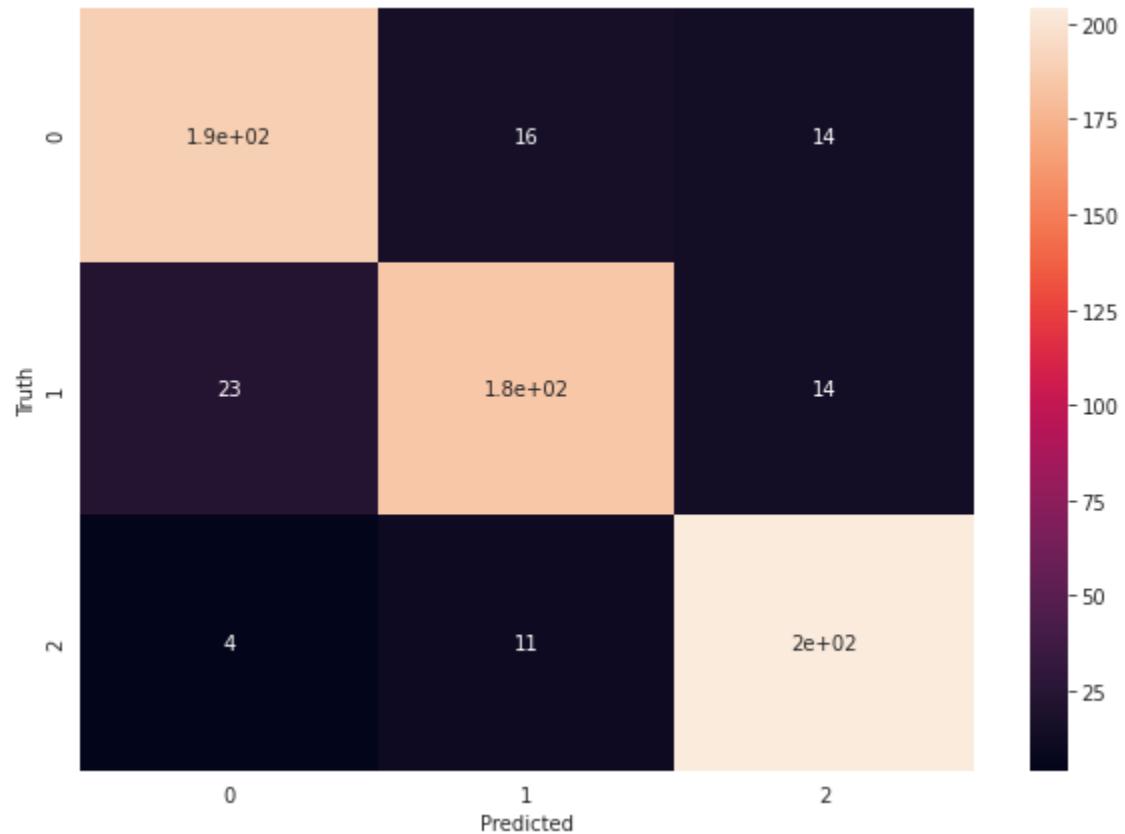


Figure 4.5: Confusion Matrix for VGG16 Image Only Model

Statistical Analysis on DL Scores (VGG16)

We also perform some statistical analysis on the DL Scores given by the prediction model to calculate AUC Score. In Figure 4.5 we have shown the ROC (Receiver Operating Curve) of VGG16 model which is used to calculate the AUC scores given in the following table for each three classes of Benign, Malignant and Normal. In Figure 4.5 and Figure 4.6 Class 0 denotes Benign, Class 1 denotes malignant and Class 2 denotes Normal.

	Micro Average ROC	Macro Average ROC	Benign (Class 0)	Malignant (Class 1)	Normal (Class 2)
AUC Scores	0.98	0.98	0.97	0.97	0.99

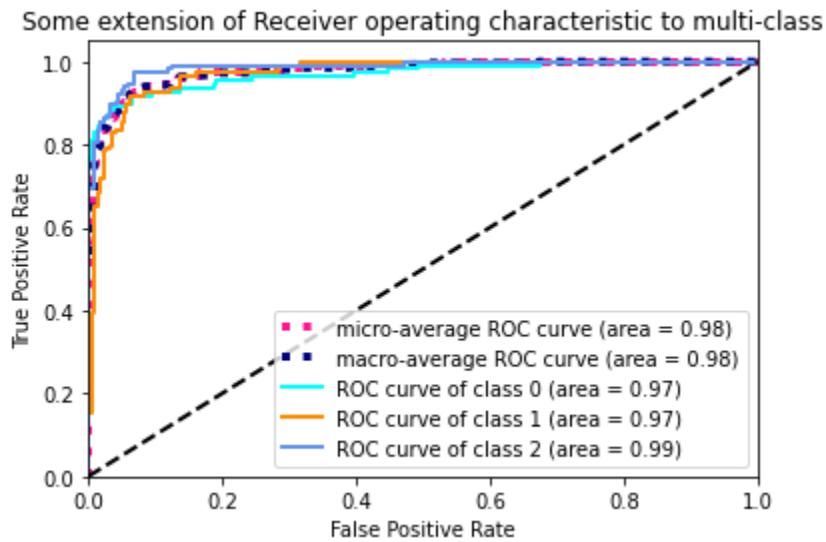


Figure 4.6: ROC for VGG16 model

Image Only Model (VGG19)

Using VGG19 we achieved an overall accuracy of 85%. with Precision of 0.83, 0.85, 0.85 for Benign, Malignant and Normal cases respectively, Recall of 0.84, 0.82, 0.89 for Benign, Malignant and Normal cases respectively and similarly F1 Score of 0.84, 0.83, 0.87 for the given three cases. The training phase progression of this CNN Model can be visualized by the following plots of Accuracy vs Number of epochs and Loss vs Number of Epochs for Training and Validation set.

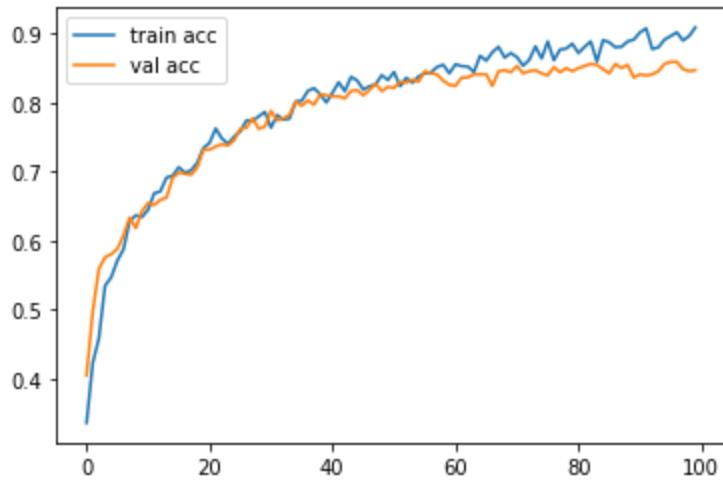


Figure 4.7: Accuracy vs No. of Epochs (VGG19)

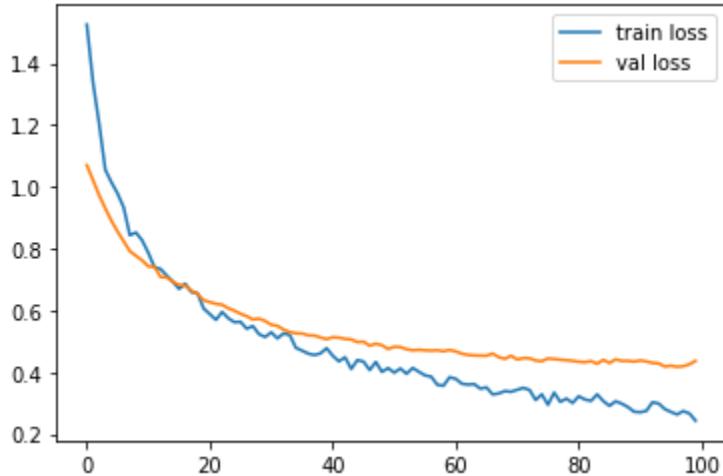


Figure 4.8: Loss vs No. of Epochs (VGG19)

Statistical Analysis on DL Scores (VGG19)

We also perform some statistical analysis on the DL Scores given by the prediction model to calculate AUC Score. In Figure 4.8 We have shown the ROC (Receiver Operating Curve) of VGG19 model which is used to calculate the AUC scores given in the following table for each three classes of Benign, Malignant and Normal. In Figure 4.8 Claas 0 denotes Benign, Class 1 denotes malignant and Class 2 denotes Normal.

	Micro Average ROC	Macro Average ROC	Benign (Class 0)	Malignant (Class 1)	Normal (Class 2)
AUC Scores	0.95	0.95	0.94	0.93	0.98

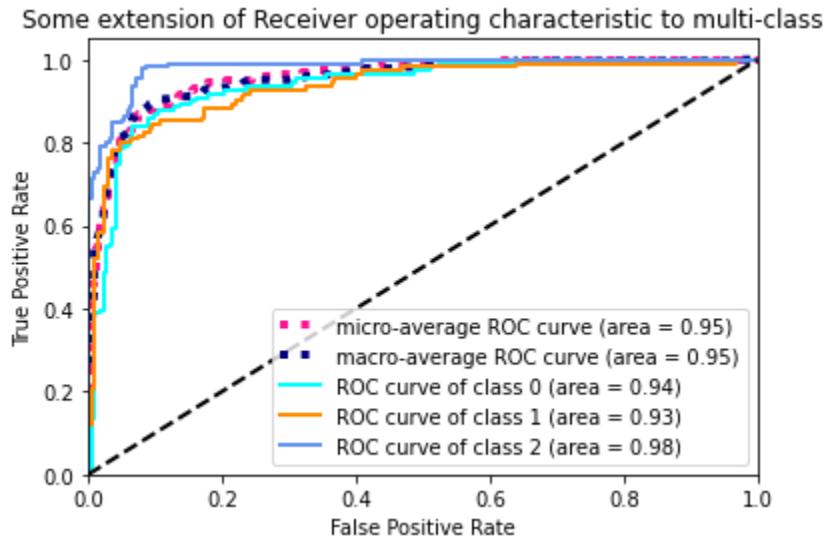


Figure 4.9: ROC for VGG19 model

Image Only Model (ResNet50)

Using ResNet50 we achieved an overall accuracy of 61%. with Precision of 0.61, 0.56, 0.65 for Benign, Malignant and Normal cases respectively, Recall of 0.44, 0.57, 0.82 for Benign, Malignant and Normal cases respectively and similarly F1 Score of 0.51, 0.57, 0.72 for the given three cases. The training phase progression of this CNN Model can be visualized by the following plots of Accuracy vs Number of epochs and Loss vs Number of Epochs for Training and Validation set.

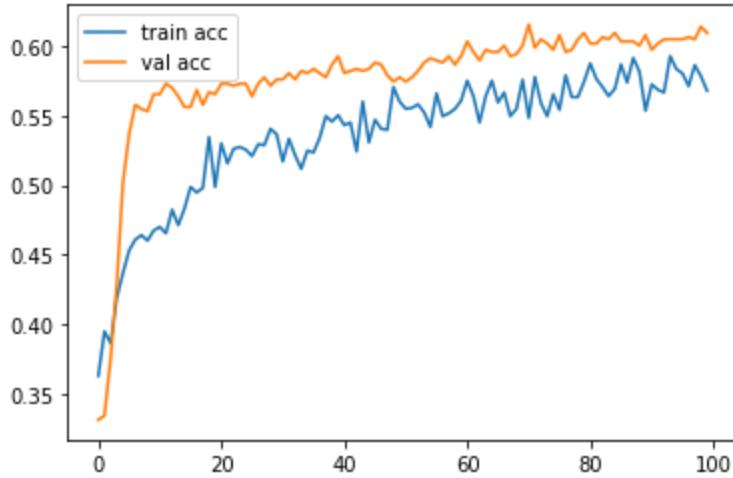


Figure 4.10: Accuracy vs No. of Epochs (ResNet50)

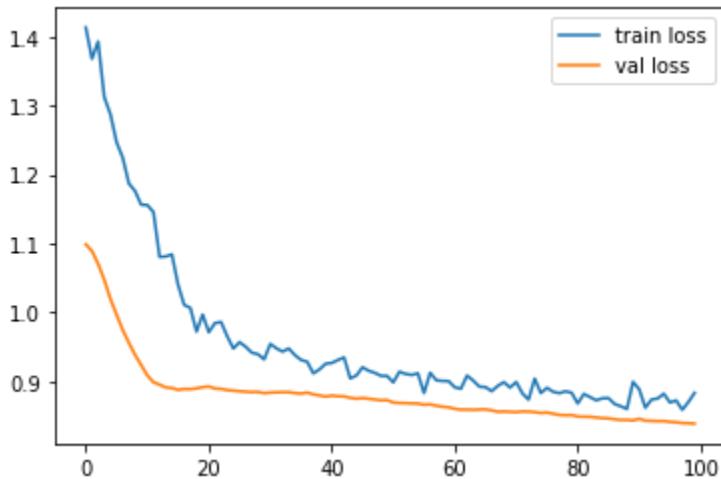


Figure 4.11: Loss vs No. of Epochs (ResNet50)

Statistical Analysis on DL Scores (ResNet50)

We also perform some statistical analysis on the DL Scores given by the prediction model to calculate AUC Score. In Figure 4.11 We have shown the ROC (Receiver Operating Curve) of ResNet50 model which is used to calculate the AUC scores given in the following table for each of the three classes of Benign, Malignant and Normal. In Figure 4.11 Claas 0 denotes Benign, Class 1 denotes malignant and Class 2 denotes Normal.

	Micro Average ROC	Macro Average ROC	Benign (Class 0)	Malignant (Class 1)	Normal (Class 2)
AUC Scores	0.80	0.79	0.73	0.77	0.88

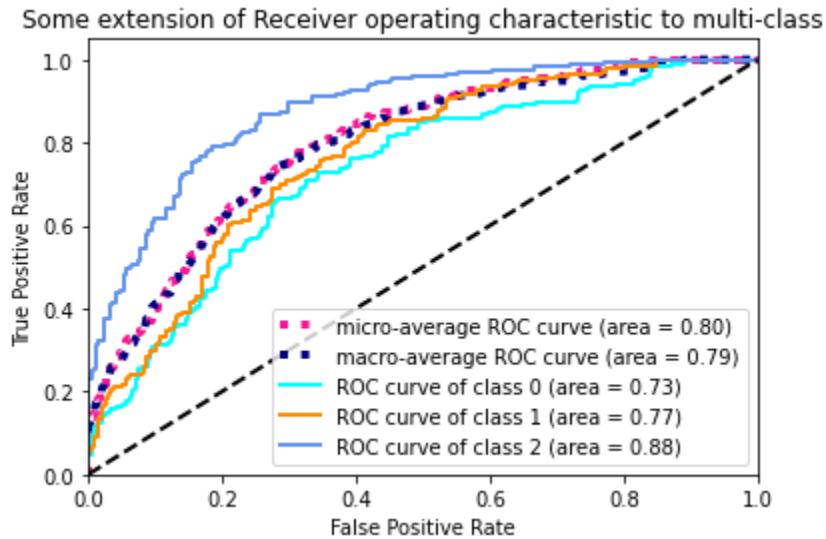


Figure 4.12: ROC for ResNet50 model

From the above experiments it is seen that VGG16 is performing better than the rest of the CNN models in terms of Accuracy, Precision, Recall and F1 Score. Also from the statistical analysis performed based on DL Scores, it is clear that the AUC score for the VGG16 model is better than the rest of the model. Hence the best model for this task of Early Prediction of Breast Cancer is VGG16 with overall accuracy of 88% for the validation/ test set examples and Precision of 0.87, 0.88, 0.89 for Benign, Malignant and Normal cases respectively, Recall of 0.86, 0.86, 0.92 for Benign, Malignant and Normal cases respectively and similarly F1 Score of 0.86, 0.87, 0.91 for the given three cases.

4.5.2 Performance of Hybrid Model (VGG16 CNN model with Risk Factors)

For this experiment we make use of the risk factors along with the CNN model (VGG16) as discussed in earlier chapters. We achieved an overall accuracy of 88% and Precision of 0.90, 0.87, 0.88 for Benign, Malignant and Normal cases respectively, Recall of 0.86, 0.88, 0.91 for Benign, Malignant and Normal cases respectively and similarly F1 Score of 0.88, 0.88, 0.89 for

the given three cases. Hence the inclusion of additional information as risk factors improves the CNN model prediction by increasing the Precision. This result might not be so significant for this project but it has the ability to boost the performance if we consider more risk factors and more quality mammogram datasets.

The accuracy vs epochs and loss vs epochs for both training and test/validation datasets are shown in following figures.

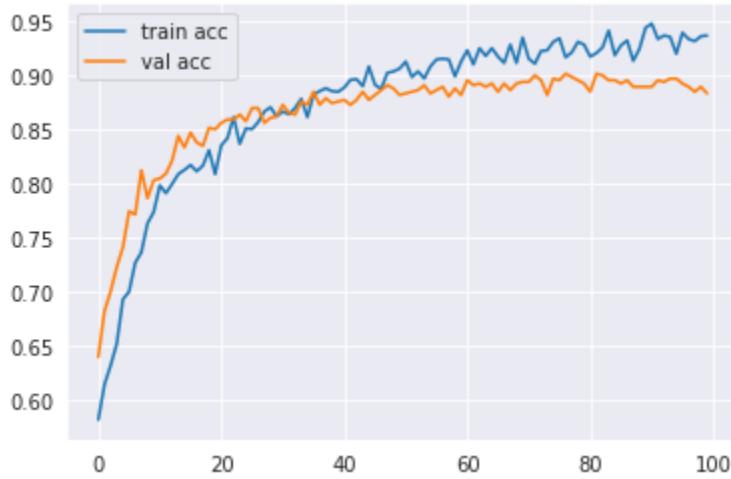


Figure 4.13: Accuracy vs No. Epochs (VGG16 with Risk Factors)



Figure 4.14: Loss vs No. of Epochs (VGG16 with Risk Factors)

In the following figure we have shown the confusion matrix for the Hybrid model (Image with risk factors) based on VGG16 architecture

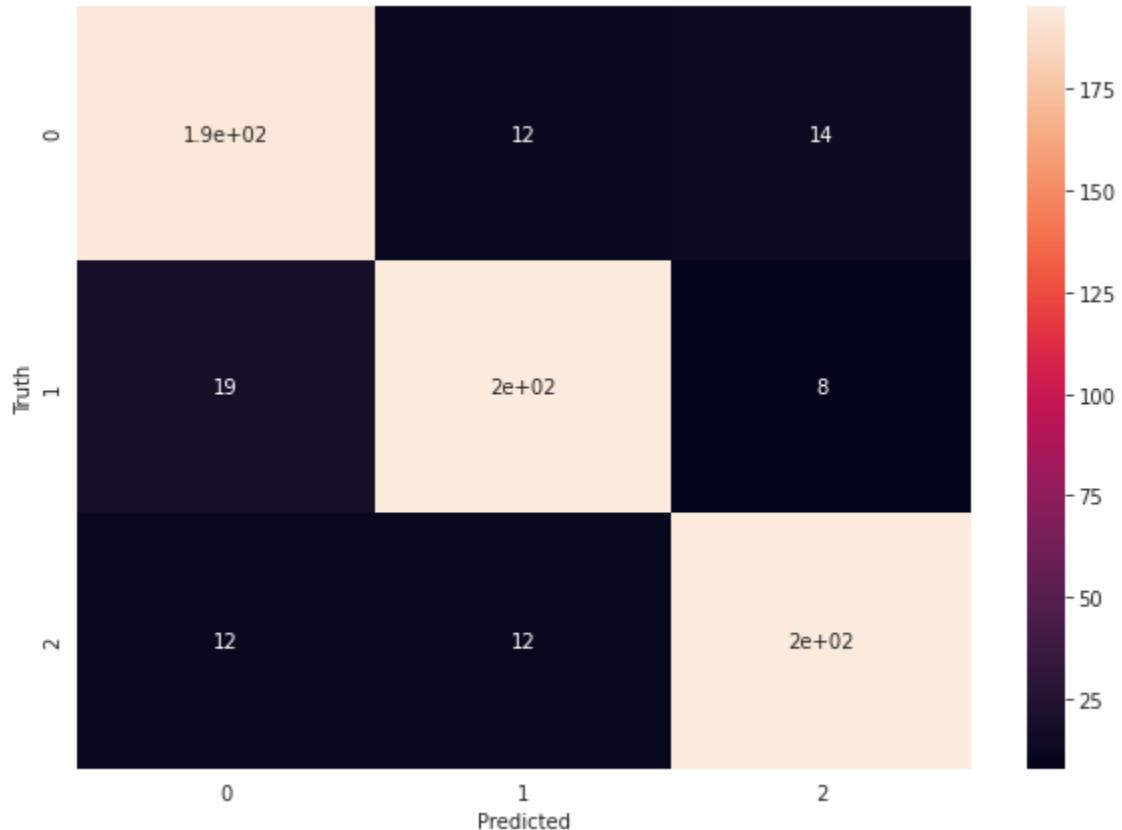


Figure 4.15. : Confusion Matrix for Hybrid Model

Statistical Analysis on DL Scores (VGG16 with Risk factors)

We also perform some statistical analysis on the DL Scores given by the prediction model to calculate AUC Score. In Figure 4.11 We have shown the ROC (Receiver Operating Curve) of ResNet50 model which is used to calculate the AUC scores given in the following table for each of the three classes of Benign, Malignant and Normal. In Figure 4.14 Class 0 denotes Benign, Class 1 denotes malignant and Class 2 denotes Normal.

	Micro Average ROC	Macro Average ROC	Benign (Class 0)	Malignant (Class 1)	Normal (Class 2)
AUC Scores	0.96	0.96	0.95	0.96	0.98

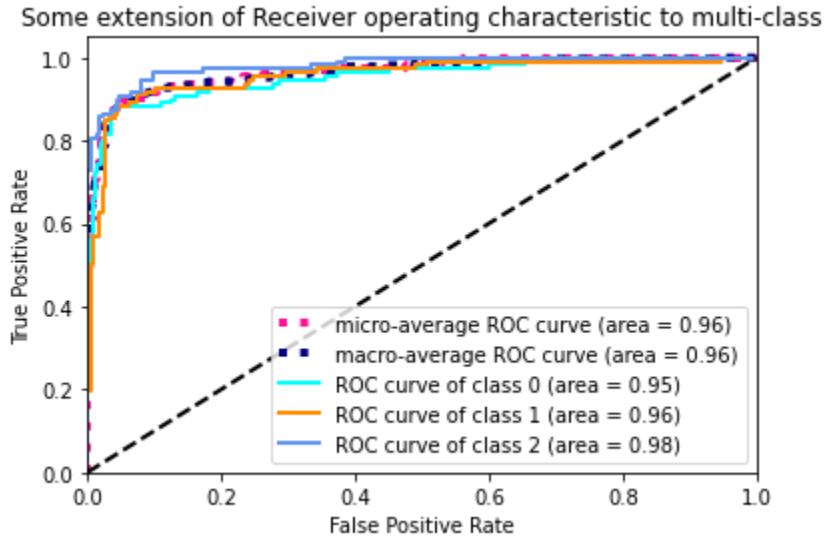


Figure 4.16: ROC for VGG16 (with Risk Factors)

4.6 Final Prediction and Classification with DL Score

Here we discuss about the final prediction given by the CNN model which has been trained with different deep learning techniques. The prediction is given in terms of classification and associated DL Score for Benign and Normal cases. Whereas for Malignant (Cancer) cases the DL Score is referred to as DL Risk Score as it gives the probability score of a patient having breast cancer in the future.

Let us consider a Normal case given in the following Figure 4.15 which we already know but the model has never seen before or the mammogram was not used during the training phase. Using the trained CNN model, we can predict the mammogram whether it is a case of Benign, Malignant or Normal. We see that the CNN Prediction model shown in Figure 4.15 results in "Normal" case with DL score of 0.83

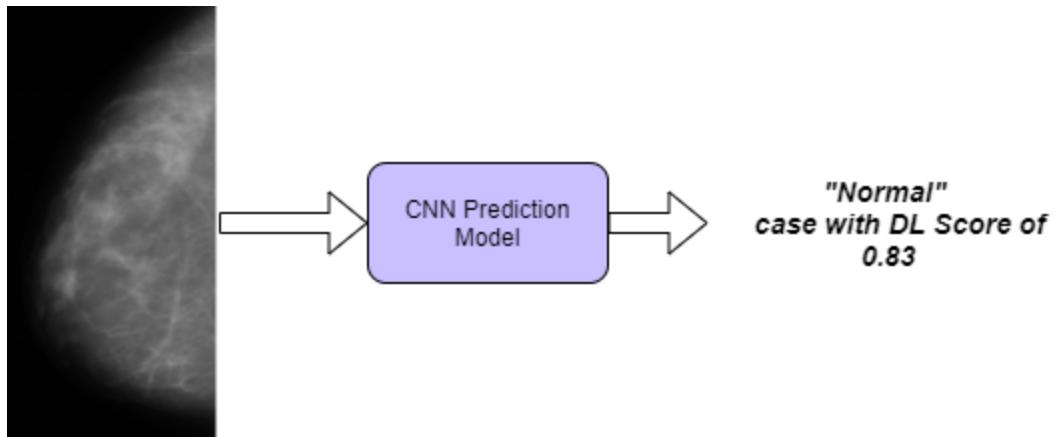


Figure 4.17: Prediction model for Normal Case

Now, let us consider a Malignant case. For validation of the CNN Prediction Model we already know that the patient would have breast cancer but the CNN model does not. Using the trained CNN model, we can predict the mammogram whether it is a case of Benign, Malignant or Normal. We see that the CNN Prediction model shown in Figure 4.16 results in “Cancer” case with DL Risk Score of 0.843

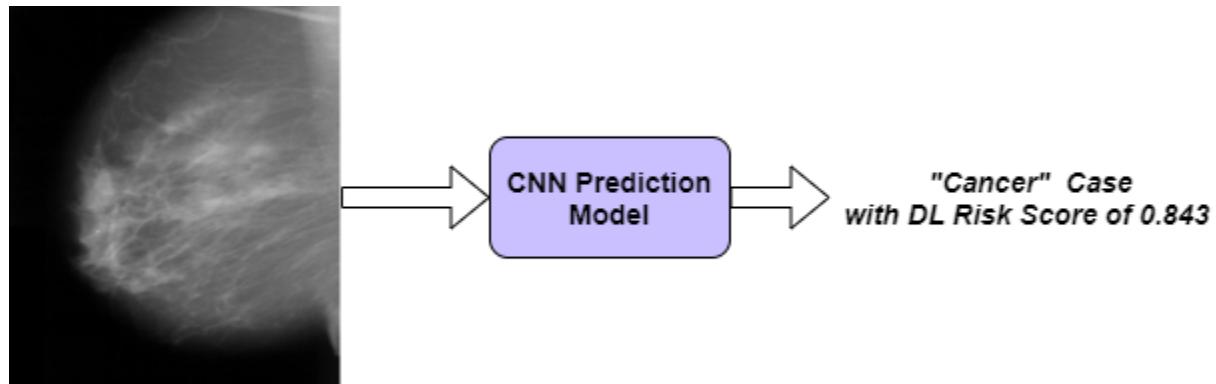


Figure 4.18: Prediction model for Malignant Case

Now, similarly to validate the prediction of Benign cases, let us consider a mammogram image from our test/ validation set. We already know that this mammogram image belongs to the “Benign” class but the CNN model has never seen it before. Using the trained CNN model, we can predict the mammogram whether it is a case of Benign, Malignant or Normal. We see that

the CNN Prediction model shown in Figure 4.17 results in the “Benign” case with DL Score of 0.912.

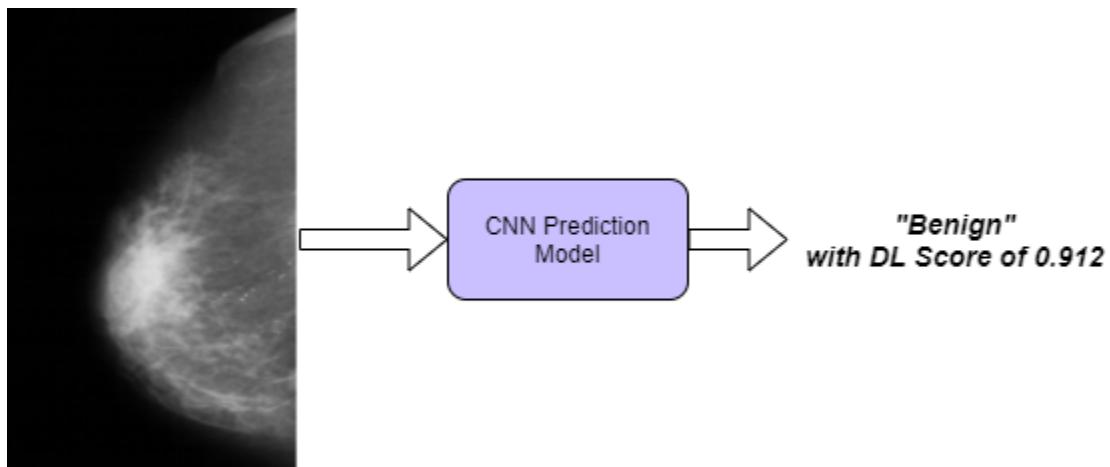


Figure 4.19: Prediction model for Benign Case

Chapter 5

Conclusion

5.1 Achievements

The main objective of the project was to develop and implement a deep learning based system which is capable of predicting early instances of breast cancer from mammogram images. After exploring and trying wide range of deep learning techniques like data pre-processing, transfer learning, hyper-parameter tuning, choice of base CNN model etc we are able to build an end to end system for classifying the mammogram in three cases namely Benign, Malignant and Normal classes with an accuracy of 92% which has the ability to outperform human level precision.

5.2 Limitations

Bias is the most concerning limitation of any deep learning technique and breast cancer prediction is also no exception. Choice of data is the key. Biases arise if we fail to collect data from all the possible samples. Data is the well known limitation in all breast cancer prediction and detection systems. Most widely used datasets of mammograms, which is DDSM contains mammograms from white females of North America only. This naturally gives bias to the CNN model while learning the mammogram data. Figure 6.1 shows the population statistics in the DDSM dataset [29].

Race	Data source	
	MGH	WFUSM
Asian	2.06%	0.20%
Black	4.12%	20.40%
Spanish Surname	6.55%	1.80%
American Indian	0.00%	0.10%
Other	0.75%	0.10%
Unknown	30.34%	0.30%
White	56.18%	77.00%

Figure 5.1: Population Statistics in DDSM dataset

Different body types linked to the geographic location of the patients used to create these databases can have a direct impact on the mammograms prediction and the CNN model can not generalise to females from other cultures. In a recent study with 53,000 North American females showed how diets that include dairy milk consumption might increase the risk of breast cancer by a maximum of 80% based on the consumption [30]. This means that if these Deep Learning algorithms were implemented in clinics outside western countries, they might not generalise well to other body morphologies (e.g. due to different diets based on the geolocation's culture). This limitation could be resolved by collecting more varied data from multiple locations around the world, not just a single region, which would also help deep learning algorithms as it always strives for more and more data. Another limitation in terms of the prediction system's usability is the confidence of the predictions. Indeed, when given new test samples, the model predicts a class label, e.g. benign, malignant or normal. However, these do not indicate the prediction's confidence, as it can be anywhere between the decision boundary's limit (not confident) and far from the decision boundary (confident). Therefore, from a clinical point of view, it is hard to make a decision based on the predictions made by a system similar to this one. Ideally, a probability-based confident metric would be coupled with the predictions to motivate the next step after the diagnosis. For example, if the confidence of a malignant tumour is high (e.g. 90-99%), then breast-conserving surgery or chemotherapy can be recommended, whereas if the confidence is low (e.g. 55-69%), then further screening tests can be recommended instead.

5.3 Future Work

The main area of work that requires improvements is the mammogram preprocessing as it is often an area where significant performance gains can be found [31] by using techniques such as global contrast normalisation (GCN), local contrast normalisation, and Otsu's threshold segmentation etc. Artefacts present in the mammogram images such as tags on the x-rays and black backgrounds should all be removed using computer vision techniques to avoid having the CNN learn irrelevant features.

Another area where improvements can be made is the fine-tuning to achieve better performance on the datasets and avoid overfitting. The choice of dataset which is to be used in the breast cancer prediction system should be free from any kind of biases be it skin colour of the patients, races, cultures, ages, diet habits, food habits, regions and continents the patients belong

to etc. For that we have to plan well defined strategies considering all the ethics and laws to collect the mammogram datasets considering all the factors which are creating biases for the CNN model for breast cancer prediction.

Bibliography

1. Sofi, Javaid. "Breast Cancer In India". Times Of India Blog, 2021, <https://timesofindia.indiatimes.com/blogs/poverty-of-ambition/breast-cancer-in-india/>.
2. "Statistics Of Breast Cancer In India | Cytecare Hospitals". Cytecare Hospital In Bangalore, 2021, <https://cytecare.com/blog/statistics-of-breast-cancer>.
3. Smith, R. A., Saslow, D., Sawyer, K. A., Burke, W., Costanza, M. E., Evans III, W. P., ... & Sener, S. (2003). American Cancer Society guidelines for breast cancer screening: update 2003. CA: a cancer journal for clinicians, 53(3), 141-169.
4. Ramos-Pollan, R., Guevara-Lopez, M. A., Suarez-Ortega, C., Daz-Herrero, G., Franco-Valiente, J. M., Rubio-Del-Solar, M., González-De-Posada, N., Vaz, M. A. P., Loureiro, J. and Ramos, I. (2012), 'Discovering mammography-based machine learning classifiers for breast cancer diagnosis', Journal of Medical Systems 36(4), 2259-2269.
5. Osareh, A. and Shadgar, B. (2010), Machine learning techniques to diagnose breast cancer, in '2010 5th International Symposium on Health Informatics and Bioinformatics, HIBIT 2010', pp. 114-120.
6. Polat, K., & Güneş, S. (2007). Breast cancer diagnosis using least square support vector machine. Digital signal processing, 17(4), 694-701.
7. Elter, M. and Horsch, A. (2009), 'CADx of mammographic masses and clustered microcalcifications: A review', Medical Physics 36(6), 2052-2068.
8. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R. B. and Sieh, W. (2017), 'Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography', Scientific Reports 9(1).
9. Oeffinger, K. C., Fontham, E. T., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y. C. T., ... & Wender, R. (2015). Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. Jama, 314(15), 1599-1614.
10. Martin, Laura J. (2019), 'WebMD: Breast Biopsy for Breast Cancer Diagnosis',

11. Elter, M., & Horsch, A. (2009). CADx of mammographic masses and clustered microcalcifications: a review. *Medical physics*, 36(6Part1), 2052-2068.
12. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.
13. Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. *Designs*, 2(2), 13.
14. Wolberg, W. H., Street, W. N. and Mangasarian, O. L. (1995), 'UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set'.
15. Yue, W., Wang, Z., Chen, H., Payne, A. and Liu, X. (2018), 'Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis', *Designs* 2(2), 13.
16. Wu, Y., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., & Metz, C. E. (1993). Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology*, 187(1), 81-87.
17. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R. B. and Sieh, W. (2017), 'Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography', *Scientific Reports* 9(1).
18. Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012), *ImageNet Classification with Deep Convolutional Neural Networks*, Technical report, Google.
19. Simonyan, K. and Zisserman, A. (2014), 'Very deep convolutional networks for large-scale image recognition'
20. Falconi, L. G., Perez, M. and Aguilar, W. G. (2019), 'Transfer Learning in Breast Mammogram Abnormalities Classification with Mobilenet and Nasnet', *International Conference on Systems, Signals, and Image Processing* 2019-June, 109{114.

21. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (pp. 265-283).
22. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703.
23. Charitha Dissanayake Lekamlage, Fabia Afzal, Erik Westerberg and Abbas Cheddad, "Mini-DDSM: Mammography-based Automatic Age Estimation," in the 3rd International Conference on Digital Medicine and Image Processing (DMIP 2020), ACM, Kyoto, Japan, November 06-09, 2020.
24. Wirth, M. A., Lyon, J., Nikitenko, D., & Stapiński, A. (2004, May). Removing radiopaque artifacts from mammograms using area morphology. In Medical Imaging 2004: Image Processing (Vol. 5370, pp. 1042-1053). International Society for Optics and Photonics.
25. Yue, W., Wang, Z., Chen, H., Payne, A. and Liu, X. (2018), 'Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis', Designs 2(2), 13. URL: <http://www.mdpi.com/2411-9660/2/2/13>
26. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R. B. and Sieh, W. (2017), 'Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography', Scientific Reports 9(1). URL: <http://arxiv.org/abs/1708.09427> <http://dx.doi.org/10.1038/s41598-019-48995-4>
27. Falconi, L. G., Perez, M. and Aguilar, W. G. (2019), 'Transfer Learning in Breast Mammogram Abnormalities Classification with Mobilenet and Nasnet', International Conference on Systems, Signals, and Image Processing 2019-June, 109{114.
28. Yala, A., Lehman, C., Schuster, T., Portnoi, T. and Barzilay, R. (2019), 'A Deep Learning Mammography-based Model for Improved Breast Cancer BIBLIOGRAPHY 73 Risk Prediction', Radiology 292(1), 60{66.

29. Fraser, G. E., Jaceldo-Siegl, K., Orlich, M., Mashchak, A. Sirirat, R. and Knutsen, S. (2020), 'Dairy, soy, and risk of breast cancer: those confounded milks', International Journal of Epidemiology.
30. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B. and Sanchez, C. I. (2017), 'A survey on deep learning in medical image analysis'.