Breast Cancer Classification Using Convolutional Neural Network

Eman Ahmed F18CSC41

Abdul Basit F18CSC23

Abdul Subhan F18CSC32

A project report submitted in partial fulfilment of the Requirements for the award of the degree of Bachelor of Computer Science

Computer Science Department

Salim Habib University, Karachi

2022

Declaration

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Salim Habib University or other institutions.

Name : ______________________

Reg No. : ______________________

Signature : ______________________

Name : ______________________

Reg No. : ______________________

Signature : ______________________

Name : ______________________

Reg No. : ______________________

Signature : ______________________

Date : ______________________

Approval for Submission

We certify that this project report entitled **"BREAST CANCER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS"** was prepared by **EMAN AHMED, ABDUL BASIT and ABDUL SUBHAN** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science at Salim Habib University.

Approved by,

Signature : ______________________	Signature : ______________________
Supervisor : Dr Sheeraz Arif	Co supervisor: Dr Rizwan Ahmed Khan
Date : ______________________	Date : ______________________

The copyright of this report belongs to Salim Habib University according to the Intellectual Property Policy of Salim Habib University amended on July 21st 2022. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2022 Salim Habib University. All right reserved.

Acknowledgements

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express our gratitude to our research supervisor, Dr Sheeraz Arif and co-supervisor, Dr Rizwan Ahmed Khan for their invaluable advice, guidance and their enormous patience throughout the development of the research.

In addition, we would also like to express our gratitude to our loving parents and friends who had helped and given us the encouragement.

BREAST CANCER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

ABSTRACT

One of the top causes of death for women worldwide is breast cancer. If discovered in its later stages, treatment is challenging; nevertheless, early discovery can greatly improve survival chances and the lives of millions of women. It is crucial for the scientific community to develop a paradigm for early detection, classification, and diagnosis given the broad prevalence of breast cancer. Such frameworks are being developed by the artificial intelligence research community in collaboration with medical professionals to automate the detection process. It is anticipated that the outcomes of the AI framework will assist even more physicians in generating accurate predictions given the increase in research efforts, availability of massive datasets, and improved computational capabilities. In this article, we have done a comparison of state-of-the-art Convolutional Neural Network architectures and our De Novo model on the classification of the mammograms. The mammograms used are from the publicly available CBIS-DDSM dataset.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

Breast cancer is one of the leading causes of death worldwide. It is the second prevalent type of cancer among women [4]. To help better control and treat the cancer before it spreads to other organs, frequent screening programs can detect early breast cancer before any appearance of symptoms. In order to reduce the mortality of breast cancer, it is important to find signs of the cancer. Among the imaging modalities, mammography is the effective method for early detection and screening of breast cancer [5]. Mammography is a specialized medical imaging that uses low-dose X-ray system to see inside the breast. A mammography exam, called a mammogram, aids in the early detection and diagnoses of breast cancer in women. The basic structure of a mammogram is shown in Figure 1.

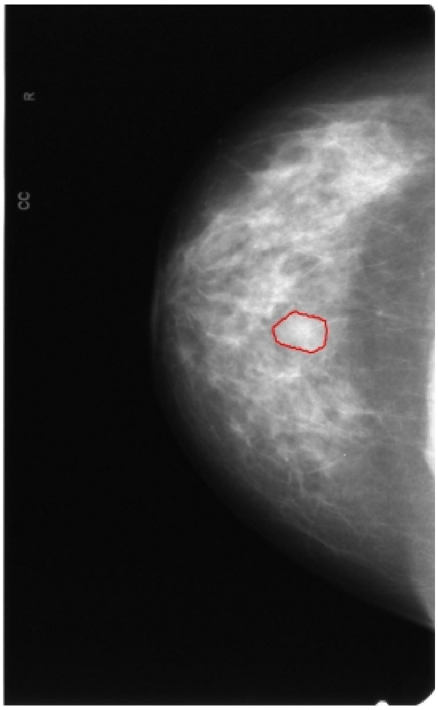

Figure 1: Mammogram of a Human Breast

According to various studies, mammography is used as the ideal tool for early detection of breast cancer to improve the survival chances of women with breast cancer. However, 30%

of the results of mammograms return false-negative results which also caries with the breast density, the age of the patient as well as lesion type. It is harder to diagnose breast cancer in denser breasts as they have low visibility of contrast between the lesions and the background. Double readings of mammograms have been tried to improve the sensitivity and specificity of mammograms, however, the double did not create a statistical difference in detection of lesions as compared to a single reading.

To improve the inconsistency of the results, Computer-Aided Diagnosis (CAD) systems have recently been developed to aid doctors in reading and classifying medical images be helping in the detection of the lesion as well as in measuring the malignancy of the lesion Two groups of the CAD system include CADe and CADx. CADe systems are used by radiologists to identify possible abnormalities while the CADx system is used by the radiologist as a decision making aid in classifying the findings from the imaging. CADx systems work by extracting the various characteristics of an image and consequently using a classifier to measure the total malignancy.

For breast cancer diagnosis, Computer Aided Diagnosis (CAD) method has been developed to increase the efficiency and effectiveness of breast screening [6]. Conventional CAD systems are limited to manually extracting features to describe suspect structures in the breast [7]. Currently researchers aim to develop fully automated end-to-end CADS systems [7].

It is possible to get fully automated CAD system with the rise of deep learning techniques, in which the features are automatically discovered through supervised learning [8]. Particularly, Convolutional Neural Networks (CNNs) are widely being used for developing CAD systems for detecting and classifying breast cancer. A. Shrestha et al. developed a new algorithm to describe deep learning [9]. The models included InceptionV3, DenseNet121, ResNet50, and VGG16 models, for the classification process [10], [11], [12].

The sudden progress and wide scope of deep learning, and the resulting surge of attention and multi-billion dollar investment, has led to a virtuous cycle of improvements and investments in the entire field of machine learning. It is now one of the popoular areas of study world-wide [13]. Healthcare providers generate and record massive volumes of data containing extremely significant signals and information, at a rate which exceeds what the "conventional" methods of analysis can process [14]. Machine learning therefore quickly enters the picture, as it is one of the best ways to integrate, analyze and make predictions based on large, heterogeneous data sets (cf. health informatics [15].

Traditional Machine Learning algorithms are as complex as they may seem. They need a great deal of domain expertise, human intervention, and are only capable of what they're designed for; nothing more, nothing less. In traditional machine learning algorithms, most of the applied features must be identified by a domain expert to reduce the complexity of the data [16] and make the patterns visible more clearly for the learning algorithms to work. The biggest advantage of deep learning algorithms, is that they try to gradually learn high-level features from the data. This eliminates the need for domain expertise and core feature extraction.

The article is structured as follows: Section 2 describes how breast cancer classifier, using Artificial Intelligence, advanced over the years. This sections discusses progress made by conventional ML algorithm (refer Section 2.1) and Deep Learning algorithms (refer Section 2.2) in the detection of breast cancer. In the last of this section, details on Convolutional Neural Network (CNN) (refer Section 2.3), a special architecture of Deep Learning used to learn data representation in images, is presented. Importance on CNN is given as this architecture is able to achieve state-of-the-art results for breast cancer detection. Section 3 presents the detail on mammogram which is used as the imaging modality to detect breast cancer detection in our research. Section 4 describes how we have used CNN architectures to detect cancer in mammograms and finally, Section 5 presents the results of our methodology. Lastly, Section 6 concludes the article with summary and future research directions.

Chapter 2

2 Literature Review

2.1 Conventional practices in medical image analysis

Breast cancer is among the leading causes of death worldwide and early detection of breast cancer is essential to reduce the mortality rate of breast cancer. Imaging techniques such as mammograms are an effective means for early detection. To improve the accuracy and efficiency of the classification of lesions of breast cancer, recently developing methodology such as CAD systems are developed to extract the features and distinguish the lesions.

The major purpose of CAD is to improve the detection rate of abnormal regions while lowering the false-negative rate, which might occur because of human error or stress [17]. The standard CAD workflow is: 1) Algorithms find regions of interest initially using image processing techniques that rely on an extensive range of handcrafted characteristics (texture, shape, grey-level level intensity) that professionals spend hours extracting. 2) The selected regions are then represented by a statistical set of characteristics. 3) A statistical classifier outputs a probability of, or predicts, a disease condition using the engineered characteristics [17].

Machine Learning, as described by Arthur Samuel in 1959 [18], is a technique that recognizes patterns from given inputs, learn those patterns without being explicitly programmed and solve the problems based on inputs [19], [20], [21]. Machine learning is one of the most widely used methods for effectively training machines and developing predictive models for better decision-making [22]. One of the problems with machine learning is that users have to select the features like shapes, areas, histogram pixel of interest region (i.e. cancer regions), which define the class of the image it belongs to [23].

Manual features are extracted using the extraction feature algorithms such as Histogram of Gradient (HOG) and Local Binary (LBP) pattern. These algorithms focus mainly on a single aspect of the image, such as the texture or edge of the image. Thus, these algorithms may fail with different types of images. For example, a feature algorithm that extracts texture-based features may not be able to extract features that contain shape information [24].

For breast cancer detection and analysis, various handmade features-based Machine Learning algorithms have been applied. For the past breast cancer diagnosis, Yassin et al. [25] have outlined different conventional Machine Learning algorithms in recent years. The algorithms are but also not limited to Decision Trees, Logistic Regression Random Forest, Support Vector Machines, Linear Discriminant Analysis, Naive Bayes, and K-Nearest Neighbor [26] [27].

Since there are such limitations of extracting adequate information, there is always a need of the classifier for making classification decisions to handle the acquired feature space. However, it is a complicated task to select an appropriate classifier [28].

2.2 Artificial Intelligence algorithms based on Deep Learning

Deep Learning algorithms do not require manual and explicit feature extractions from input data. They process raw data and then produce predication by adapting to all sorts of features from the input data (audio, text,image etc). Hence, they provide much stronger results in classification problems compared to conventional Machine learning methods [29], [30].

Deep Learning Algorithm are originated from Artificial Neural Network (ANN) which is the network of artificial neurons that mimic the simple function of brain/biological neurons (see Figure 2 for reference) [31].

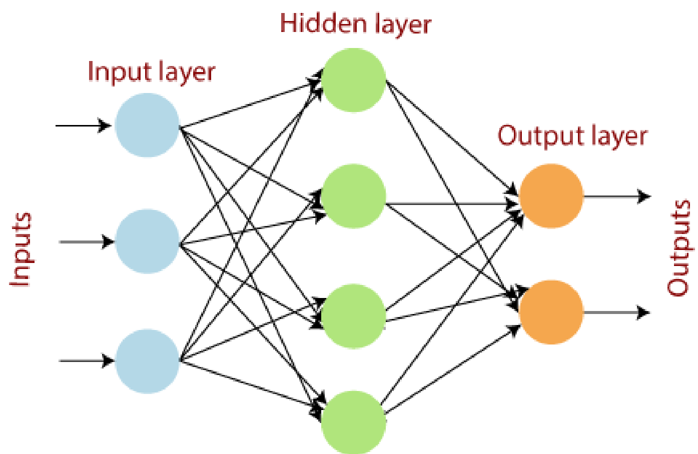

Figure 2: Feed-forward Neural Network

Nodes, also known as neurons, of an ANN are linked together in single or multiple hierarchical layers and can send and receive signals, as shown in Figure 2. The output of the nonlinear activation function determines the reaction (rejection or acceptance) of these sent and received signals. Each neuron or node in a network has a weight associated with its input that might impact the given input and is important for transferring data to the output layer. The input layer sends data and inputs to the hidden layer in the form of a feature vector with a weighted value. The activation units in the hidden layer carry the features vector from the first layer with weighted values and execute calculations as output. The output layer is made up of activation units, each of which corresponds to a label / class in the dataset, and it carries the hidden layer's weighted output and predicts the appropriate class [32]. ANN reduces the error function by employing back-propagation [33] during the training phase. By adjusting the weight values in each layer, the error is minimized.

In classifying breast cancer, mainly two types of ANN have been used namely Shallow Neural Network (SNN) [34] and Deep Neural Network (DNN) [35]. Between these two neural networks, the Deep Neural Network is the most popular form of ANN that is widely used in medical image analysis and diagnosis.s [36].

Without the requirement for expert knowledge or human intervention, a Deep Neural Network can extract relevant features from raw input data. Deep Learning approaches provide substantial improvements in diagnostic, analysis and clinical decision-making processes using medical imaging data [24]. Among the Deep Learning architecture, Convolutional Neural Network is the most powerful, effective and extensively used for medical imaging analysis particularly breast imaging analysis [37].

2.3 Convolutional Neural Network for breast cancer classification

The subtype of feed deep, forward neural networks are Convolutional Neural Networks (CNN). The effectiveness of CNNs have been displayed for various applications that involve visual inputs. [38]. The architecture of CNN uses the input of the entire data in the form of an image to allow it to train the model. Two main techniques are implemented, the first one is Transfer Learning and the second being de novo. The de novo technique works by training the CNN architecture from scratch while the optimal model is obtained by combining various trained models. Comparatively, in the transfer learning method, models that are pre-trained are adopted for analysis. [38]. While the conventional method is transfer learning, it mostly only deals with small or very limited datasets. [39].

The architecture of CNN is based on hierarchical layers which include trainable filters known as the convolution operation and also decreases the size of an image creating pooling layers. The discriminating features are hence extracted from the raw input data which allows the classification to be enhanced. To conclude, the extracted features are used to calculate the probability of the learned classes by the classification. [28]. Figure 3 illustrates the breast cancer detection using deep CNN with Pooling layers as well as convolution. The overfitting of the data is reduced by using the Dropout layer. Following numerous pooling layers and convolutions the fully connected layers are used. The neuron of the next layer is connected to all the neurons of these fully connected layers. Large patterns are hence easily extracted from input data as the low-level information / features are combined together. The last layer includes the softmax function which calculates the occurrence probability of any event which occurs in the data provided for training which concludes to either normal breast image or a malignant or a benign tumor.

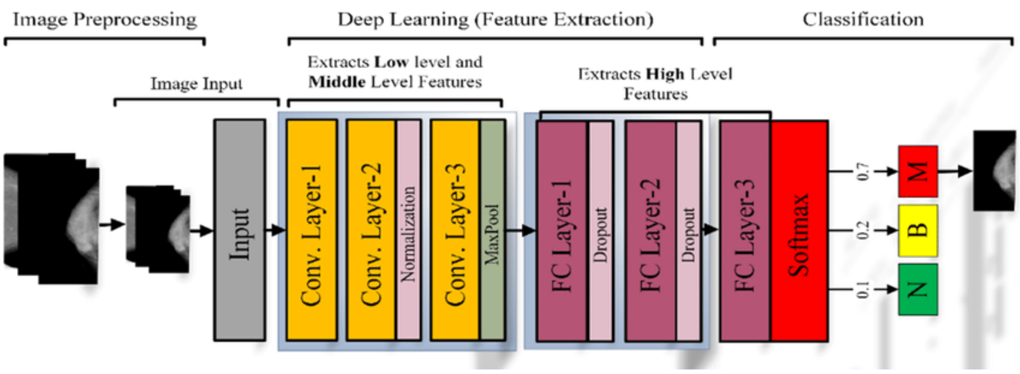

Figure 3: The architecture of deep CNN applied to breast cancer detection and classification [2]. Acronyms used in the figure: Conv. Layer= Convolutional layer, MaxPool= Maximum value pooling and FC Layer= Fully connected layer.

For the breast image analysis, the input layer in the architecture of CNN accepts an input image, as displayed in Figure 3. Convolution filters are known as kernels and are a part of the most vital layers, convolution layers. The convolution operation is implemented in these layers to extract high-level features which include colors, edges, blobs and shapes. All the convolution filters and layers are organized hierarchically which allows the extraction level from low-level to high-level features to increase as the depth of the layers increase [24].

Increasing the scale of deep neural networks or using transfer learning are the two most basic ways to improve CNN performance. This entails expanding both the depth (the number of network levels) and the depth (the number of units at each level) of the network. This is a simple and secure method of building higher-quality models, especially given the vast amount of labelled training data available. However, there are two main downsides to this easy method.

The bigger the network, the more parameters it has, which makes it more prone to overfitting, especially if the number of labelled samples in the training set is insufficient. This is a significant bottleneck since highly labelled datasets are difficult and expensive to collect, and typically need professional human raters to discern between different imaging modalities (mammograms in our case).

Secondly, the consumption of computing resources is greatly enhanced when network size is uniformly raised. When two convolutional layers are coupled in a deep vision network, each uniform increase in the number of their filters results in a quadratic increase in computation. Much of the work is lost if the extra capacity is used inefficiently (for example, if most weights wind up being near to zero). Because the computational budget is always finite, an effective allocation of computing resources is preferred over a haphazard growth in size, even if the primary goal is to improve performance quality [40].

2.3.1 Prominent CNN Architectures for Breast Imaging Analysis

When applied for breast cancer detection using imaging data, studies have shown that Convolutioanl Neural Network-based models improved diagnostic accuracy and reduced the false detection rate [41]. Various Transfer Learning based CNN architectures, such as GoogLeNet [42], VGG [43], LeNet [44], CiFarNet [45], AlexNet [46], Inception [47], ResNet [48], Inception v4 [49] have been used for breast imaging analysis using various combinations of parameters and hyper-parameters. When applied to various breast imaging datasets, all of these CNN designs have shown to be reliable. These CNN architectures might be the only option to train a model without overfitting, because having fewer parameters to train also reduces the risk of overfitting [50].

To accurately train a deep learning model, a large amount of data is normally necessary. It requires time, effort, and resources to develop labelled training mammograms by an expert radiologist. Transfer learning reduces the amount of training data needed for new deep learning models because the majority of the model has already been learned. It is challenging to obtain the required training data and develop models in medical applications, such as breast ultrasound imaging. As a result, it's best to keep the training data acquisition time and effort to a minimum. In such cases, it would be advantageous to use pretrained CNN architectures to transfer learning from one task to the target task. This allows you to employ a model that has already been trained on another domain as the learning target. As a result, the requirement for the work involved in collecting new training data for learning is reduced.

Inception-v3 is one of the few designs concept for scaling up convolutional networks, in the context of the Inception architecture. This architecture has the potential to produce excellent results. Compared to simpler, more monolithic vision networks, this computer vision network have a comparatively low processing cost. Inception-v3 is the highest quality version which has reached 21.2% top-1 error and 5.6% top-5 error in the ILSVR 2012 classification evaluation.

An Inception v3 network's architecture is developed one step at a time, as described below:

1. **Factorized Convolutions**: This decreases the amount of parameters in a network, which improves computational efficiency. It also monitors the network's efficiency.

2. **Smaller convolutions**: substituting smaller convolutions for larger convolutions results in much quicker training. A 5x5 filter, for example, has 25 parameters; two 3x3 filters, in place of a 5x5 convolution, have just 18 (3*3 + 3*3) parameters.

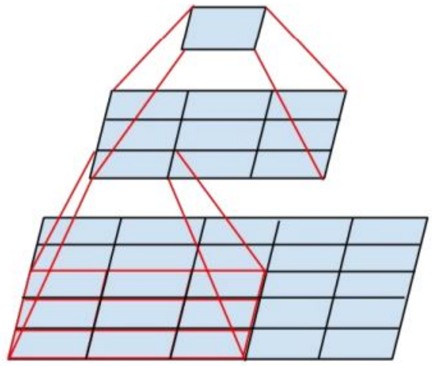

Figure 4: Mini-network replacing the 5 × 5 convolutions.

3. **Asymmetric Convolutions**: Instead of a 3x3 convolution, a 1x3 convolution followed by a 3x1 convolution might be used. The number of parameters would be significantly greater than the asymmetric convolution described if a 3x3 convolution was substituted with a 2x2 convolution.

4. **Auxiliary classifier**: a tiny CNN is placed between layers during training, and the loss it incurs is contributed to the main network loss. Auxiliary classifiers were utilised for a deeper network in GoogLeNet, whereas an auxiliary classifier works as a regularizer in Inception v3.

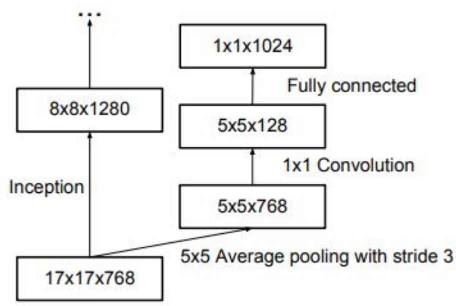

Figure 5: Auxiliary classifier on top of the last 17×17 layer

5. **Grid size reduction**: is commonly accomplished by pooling processes. However, a more effective strategy is presented to overcome computational cost barriers

VGG16 is Visual Geometry Group which is the deep and standard architecture of a convolution Neural Network with various layers. The depth of the network depends on the number of layers, VGG-16 or VGG-19, each of which contain 16 and 19 layers respectively. The concept behind this innovatory model of object recognition is the architecture of VGG. This architecture was initially developed as VGGNet, a deep neural network, and has proven to perform beyond baselines that have been set for various datasets and tasks. Additionally, it continues to be one of the most used architectures for image recognition. VGG-16 is the VGGNet model of convolutional neural network which supports 16 layers. This model has a 92.7% accuracy when tested in the top-5 test in ImageNet, a database which contains approximately 14 million images. This model involves the replacement of one larger kernel sized filter and instead uses multiple 3x3 filters consecutively which is considered to be a significant improvement when compared to AlexNet.

VGG16 is a 16 layered convolution neural network model which was initially proposed in a paper of Oxford University by K. Simonyan and A. Zisserman. This model allows you to lead a pre-trained network model which was trained using ImageNet Database with approximately a million images from it. The trained network then has the ability to classify images of objects such as keyboard, pencil, mouse or even animals into their separate categories which amount to a total of 1000. Hence the network has learned a significant amount of feature representation via the use of a broad scope of images. The input dimensions of the images that are used as an input to the network are 224 x 224.

The most distinguished feature of VGG16 is the substantial amount of hyper parameters which work based on convolution layers with a filter size of 3x3 which also contains a stride of 1 along with the say max-pool layer and padding of a filter with the size of 2x2 and a stride with the size of 2. This unique arrangement provides an architecture with consistency which leads to a total of two fully connected layers along with an output for soft-max. The number 16 in the VGG16 related to the weights of 16 different layers which creates a huge network with approximately 138 million parameters.

With the VGG network containing small convolutional filter, the VGG-16 model is made up of 13 convolutional layers and three fully connected layers. A brief description of the architecture is considered and summarized:

1. **Input**: The input side of the images that are taken as input in the model is 224×224.

2. **Convolutional Layers**: There is a minimal reception field of leverage on the model of 3x3 which is also the smallest possible size that has the ability to capture the up or down and left or right. Additionally, a convolution filter with the size of 1x1 words as the linear transformation of any input. Furthermore, the ReLU unit follows and is considered as a significant innovation when compared to the AlexNet as it reduces the time of training. The ReLU, which has a full form of rectified linear unit activation function, will only produce the output same as the input if the input is positive, if the input is negative, the output will be zero.

3. **Hidden Layers**: ReLU is used is all of hidden layers in the network. There is no leverage of the LN (Local Response Normalization) since it leads to an increase in the consumption of memory and increase in the total training time. It also does not lead to any improvement in the accuracy.

4. **Fully-Connected Layers**: The three fully connected layers in the network consist of two of them having 4096 channels each, while the third has 1000 channels.

After AlexNet [46] success in the classification competition at LSVRC2012, deep **Residual Network** [48] was undoubtedly the most revolutionary breakthrough in the computer vision/deep learning community in recent years. ResNet enables training with hundreds or even thousands of layers while maintaining impressive performance. Since AlexNet, the state-of-the-art CNN architecture have become increasingly complex. The VGG network [43] and GoogleNet [42] both featured more convolutional layers than AlexNet, which only had 5.

However, just adding layers on top of one another would not necessarily work. Because of the vanishing gradient problem, it is challenging to train deep networks because, when the gradient is back-propagated to previous layers, repetitive multiplication may lead the gradient to become infinitely tiny. As a result, the network's performance becomes saturated or even starts to decline quickly as it gets deeper.

The core idea of ResNet is to introduce an "identity shortcut connection" that skips one or more layers, as shown in the Figure 6 and Figure 7.

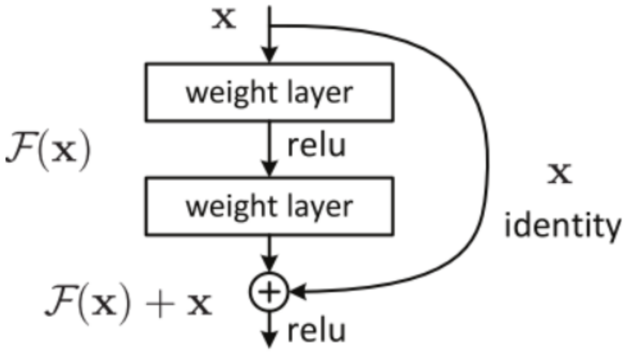

Figure 6: A residual block

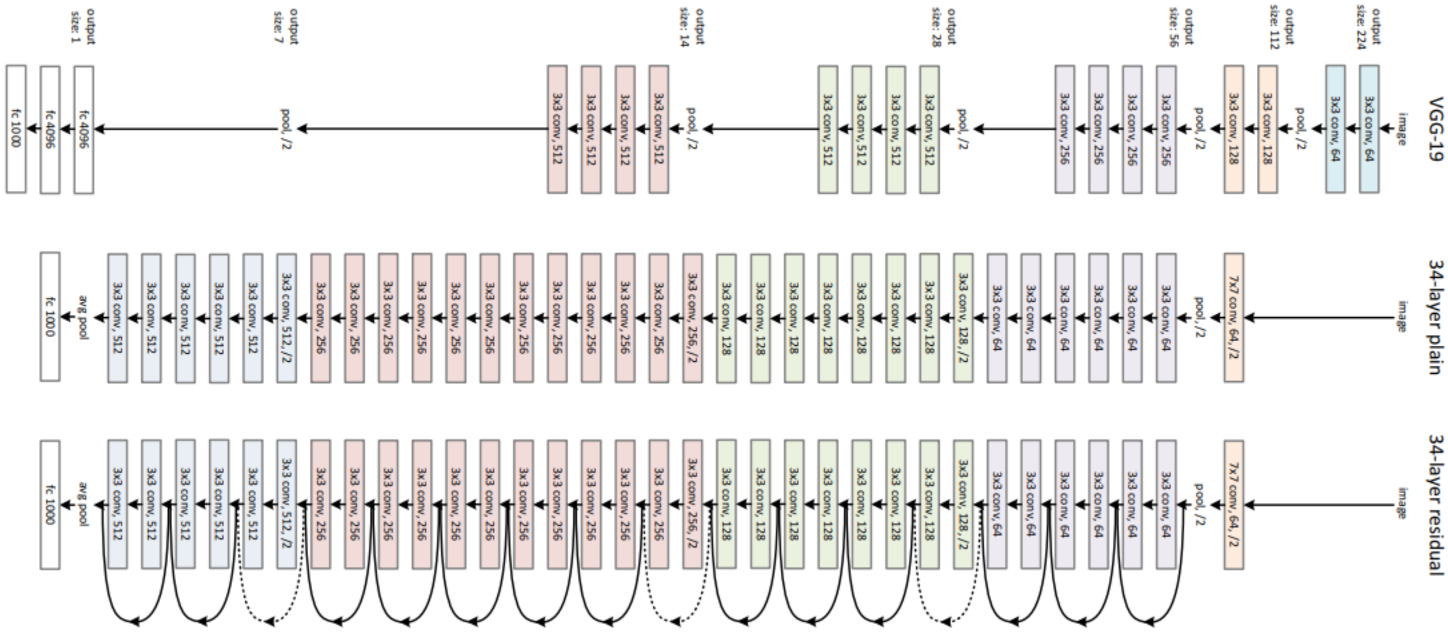

Figure 7: ResNet Architecture

The residual block was improved by the authors of [48], who also introduced a pre-activation variation of residual block [51], in which gradients can pass freely through shortcut connections to any other prior layer. There are 5 stages in the ResNet-50 model, each having a convolution and an identity block. Each identity block and each convolution block each have three convolution layers. There are around 23 million trainable parameters in the ResNet-50. One MaxPool layer, one Average Pool layer, and 48 Convolution layers make up the ResNet50 version of the ResNet model.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 8: ResNet50 Architecture

So as we can see in the Figure 8 the Resnet 50 architecture contains the following element:

- A convoultion with a kernel size of 7x7 and 64 different kernels all with a stride of size 2, gives us 1 layer.
- There is a pooling with also a stride size of 2 next.
- In Next convolution is 1x1, 64 kernel following this a 3x3, 64 kernel and lastly, a 1 * 1, 256 kernel. These three layers are repeated thrice which gives 9 layers in this step.
- Next comes a kernel of 1x1, 128 after that a kernel of 3 *x3, 128 and lastly a kernel of 1x1, 512 this step was repeated 4 times which gives 12 layers in this step.
- After that there is a kernel of 1x1,256 and two more kernels with 3x3,256 and 1x1,1024 and this is repeated 6 times which gives a total of 18 layers.
- Additionally, there is again a 1x1, 512 kernel with two more of 3x3,5 12 and 1x1, 2048 and this was repeated thrice whichh gives a total of 9 layers.
- After that, there is average pool, ending it with a fully connected layer containing 1000 nodes and at the end a softmax function so this gives 1 layer.

Hence, it gives us a 1 + 9 + 12 + 18 + 9 + 1 = 50 layers Deep Convolutional network of ResNet50 architecture.

2.4 Hybrid Approach in Breast Cancer Detection

The most crucial aspect of building a computer-aided diagnostic (CAD) system is integrating appropriate feature extraction and pattern classifiers so that they can work together to create an effective and efficient CAD system. The interest has increased, over recent years, in various applications which involve computer-aided diagnosis systems. This is done through the development of a deep learning model using a computer aided diagnosis systems and a hybrid-architecture (shown in Figure 9. This involves the use of CNN architecture to extract features from the input data and hence predict an output class with the use of Machine Learning algorithms like Support Vector Machine [52]. The classification layer includes the softmax function for image classification. An alternative study [53] was conducted which used a different method instead of the softmax function, the support, and proceeded to the conclusion that comparatively better results can be produced with the use of SVM in an artificial neural network (ANN) architecture instead of the softmax function.

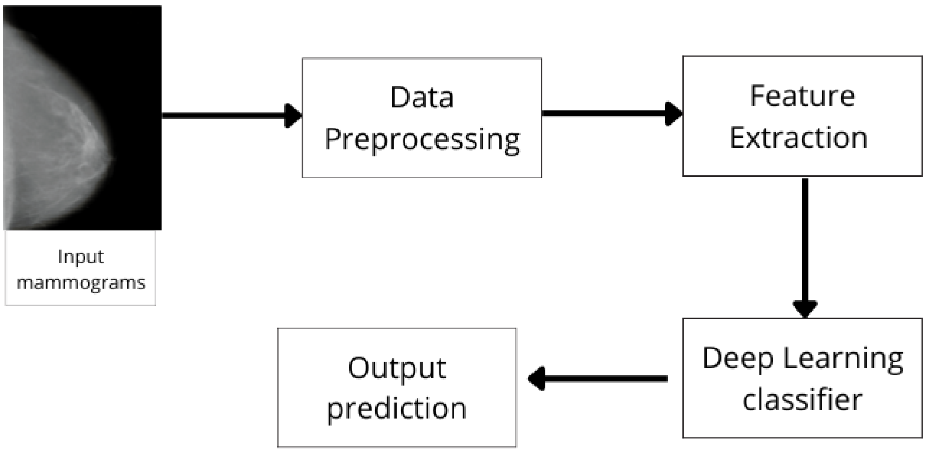

Figure 9: A flowchart of hybrid approach in CNN

To categorise breast cancer images, a hybrid technique is suggested by Inzamam et al. that combines and optimises the qualities of handmade and deep features. Pre-trained models VGG19 and InceptionV3 are serially integrated with HOG and LBP features. The suggested method's classification performance is evaluated using PCR and ICR [54].

There are various conventional practices used in medical imaging analysis and the integration of machine learning into these practices can have a significant improvement in the produced results. Deep learning methods work directly by taking raw data as input and after processing it, produce the output in the form of classification. By applying the De Novo technique on CNN, the feed forward Neural Network, the CNN architecture is trained from scratch. In breast cancer detection, the input provided to the CNN is a raw image which is then passed through the layers of the neural network. A hybrid technique has also been suggested which combines and optimizes the qualities of handmade and deep features.

Chapter 3

3 Dataset

3.1 Mammograms used for Breast Cancer Detection

Popular imaging techniques for breast cancer screening and detection are described in this section. There are a number of imaging techniques used for this but Mammograms [55], Ultrasound [56], Magnetic Resonance Imaging [57], and Histopathology [58] are the most often employed.

A mammogram is a low-dose X-ray of the breast, as shown in Figure 1. Regular mammograms are a means for reducing the risk of death caused by breast cancer as it is a means to early detection during the most initial stages when it cannot even be felt, this is when the disease is easiest to treat. For breast cancer the best form of screening test, for both women who have symptoms of cancer or for those who don't, are mammograms. A mammogram has the ability to detect breast cancer at the earliest possible stage, they can detect approximately two years before any lump can be felt. This is the stage when treatment is easily possible. For older women aged 50 years or above, mammograms are the ideal choice for detecting breast cancer as it has a sensitivity of 87%. Microcalcifications are tiny deposits of calcium which can prove to be the earliest signs of breast cancer. When compared with ultrasound images, mammograms come on top as ultrasound images fail to capture these microcalcifications while a mammogram can easily detect them. However, when compared to an MRI, the MRI reports better results of breast cancer detection as it can easily differentiate between anomalies which may look suspicious on a mammogram as an MRI is significantly more sensitive and can detect breast cancer much earlier than a mammogram. Nonetheless, the reason of preferring mammograms is that using MRIs for routine screening can be extremely expensive.

3.2 Curated Breast Imaging Subset of DDSM (CBIS-DDSM)

An open research approach has been adopted by the non-medical computer vision community to perform the analysis of breast imaging. Standard data sets have been provided for the algorithms of evaluation [59]. Hence the standardized and update version of the Digital Database for Screening Mammography (DDSM) is the CBIS-DDSM [1]. A total of 2,620 scanned film mammography studies are contained in the DDSM database which includes cases of malignant or benign tumors as well as normal cases with verified pathology information. What makes the DDSM a useful and essential part in the testing and development of a system for decisions is the scale of its database along with ground truth validation. A subset of the DDSM is a part of the CBIS-DDSM collection which has been selected and curated by a skilled and trained mammographer. The images are converted to DICOM format after decompressing. The data also includes pathological diagnosis along with bounded boxes and updated ROI segmentation [60].

What makes the DDSM an extremely useful tool, in the testing and developing of decision supporting systems, is the scale if the database as well as the truth validation. All images in the dataset have been converted, after being compressed, into the DICOM format. It also includes bounding boxes, pathological diagnosis and updated ROI segmentation for the training data. The total nuber of images included in the dataset are 10,239 while the number of subject amount to 6671. The total size of all the images in the dataset adds up to 163.6 GB.

The mammography community have been provided with a number of well-curated public datasets, one of which is the DDSM. The images in the DDSM are stored in compression files which are not standardized and hence require decompression code which may not be able to function on modern computers. A large number of images, hundreds and thousands, are also not widely available. There is also no precise segmentation available for the ROI annotations which have been provided for the abnormalities in the DDSM. Significantly improved ROI segmentation and easily accessible data is provided in the CBIS-DDSM. Advancement in mammography using decision support system research has been made possible with this resource since the dataset is of considerably improved quality in comparison with the DDSM. The format of the CBIS-DDSM is also more accessible with its updated metadata extraction code along with the decompressed images. What makes it even better and easier for use is the improved the testing and training splits that help in the evaluation.

Therefore, considering these factors, researches are often forced to implement algorithms for segmentation of the dataset. The CBIS-DDSM collection is a solution to this problem and hence contained a standardized and curated version of the DDSM.

The dataset has been divided into these 4 separate folder with each of them containing a csv file for reference:

- Calcification Training Set
- Calcification Test Set
- Mass Training Set
- Mass Test Set

The distribution of dataset in train and test as given in CBIS-DDSM is shown in Table 1.

	Benign Cases	Malignant Cases
Calcification Training Set	329 cases (552 abnormalities)	273 cases (304 abnormalities)
Calcification Test Set	85 cases (112 abnormalities)	66 cases (77 abnormalities)
Mass Training Set	355 cases (387 abnormalities)	336 cases (361 abnormalities)
Mass Test Set	117 cases (135 abnormalities)	83 cases (87 abnormalities)

Table 1: CBIS-DDSM Dataset distribution [1]

Breast Calcification is essentially the deposit of calcium within the breast tissue which appear on a mammogram in the form of white flecks or spots and are especially prevalent after age 50. While they may usually be benign, some specific patters, such as irregularly shaped tight clusters, can lead to an indication of either breast cancer or precancerous changes to breast tissue.

A lump developing in the breast is known as a breast mass and can vary in texture and size while also possibly causing pain and are often not discovered until an imaging or physical exam. Most of these are benign.

There are 3 classes in the dataset: benign, benign without callback and malignant. In our case, we have considered benign and benign without callback a single class named as benign. A benign tumor has regular, distinct and smooth edges while a malignant tumor has irregular edges and grows faster. A malignant tumor is contagious and can spread to other parts of the body. Whereas, a benign tumor can grow in size, but it will not invade nearby tissue or spread in other parts of the body.

Calcium deposits within the breast tissue is known as Breast Calcification. They appear as white spots or flecks on a mammogram. Breast calcifications are usually seen on mammograms, and they're especially prevalent after age 50. Although breast calcifications are usually benign, certain patterns of calcifications - such as tight clusters with irregular shapes and fine appearance — may indicate breast cancer or precancerous changes to breast tissue.

A breast mass is a lump that develops in the breast. Breast lumps vary in size and texture and may cause pain. Some are not found until a physical or imaging exam. Most breast lumps are benign. When it comes to detection of breast cancer, mammograms come out on top as compared to other techniques such as MRIs as they can easily capture microcalcifications which can prove to be the earliest signs of breast cancer. While an MRI is significantly more sensitive, it is also extremely expensive and can prove to be a difficult option for many for regular screening. The dataset used in this research is the standardized and update version of the Digital Database for Screening Mammography (DDSM) which is the CBIS-DDSM with a total of 2,620 scanned film mammography studies which includes cases of malignant or benign tumors as well as normal cases with verified pathology information. The dataset has been divided into these 4 separate folder with each of them containing a csv file for reference. There are 3 classes in the dataset: benign, benign without callback and malignant. We are focussing on classifying the mammograms in either benign or malignant.

Chapter 4

4 Methodology

4.1 Structuring Dataset

The CBIS-DDSM collection is a solution to this problem and hence contained a standardized and curated version of the DDSM. A manifest file for images was downloaded from the source [60] in .tcia format which were extracted in dicom format through NBIA Data Retriever. A DICOM image file is an outcome of the Digital Imaging and Communications in Medicine standard. Specifically, image files that are compliant with part 10 of the DICOM standard are generally referred to as “DICOM format files” or simply “DICOM files” and are represented as “.dcm" [61]. The images are then converted into png format. PNG (Portable Networks Graphics) supports greater color depth and because it is usually a lossless compression, it enables better image quality.

At first we labelled ROI images given in the dataset according to the pathology which is given in the csv files mentioned in 3.2. But after inspection, we were unable to extract information from the ROI images. Hence, we labelled the full mammograms and created 8 folders out of it:

1. calc_benign_train.csv
2. calc_malignant_train.csv
3. mass_benign_train.csv
4. mass_malignant_train.csv
5. calc_benign_test.csv
6. calc_malignant_test.csv
7. mass_benign_test.csv

8. mass_malignant_test.csv

After that, we merged the benign and malignant folders and created **train_csv** and **test_csv**; with proper labelling of the images belonging to benign and malignant class.

To get baseline results, we augmented the dataset in two ways:

- Histogram Equalization
- Image Sharpening

4.2 Data Augmentation

- **Histogram Equalization**. It is a digital image processing method to improve image contrast as shown in Figure 10. It does so by constructively spreading out the the most frequent intensity values, i.e. expanding out the image's intensity range.

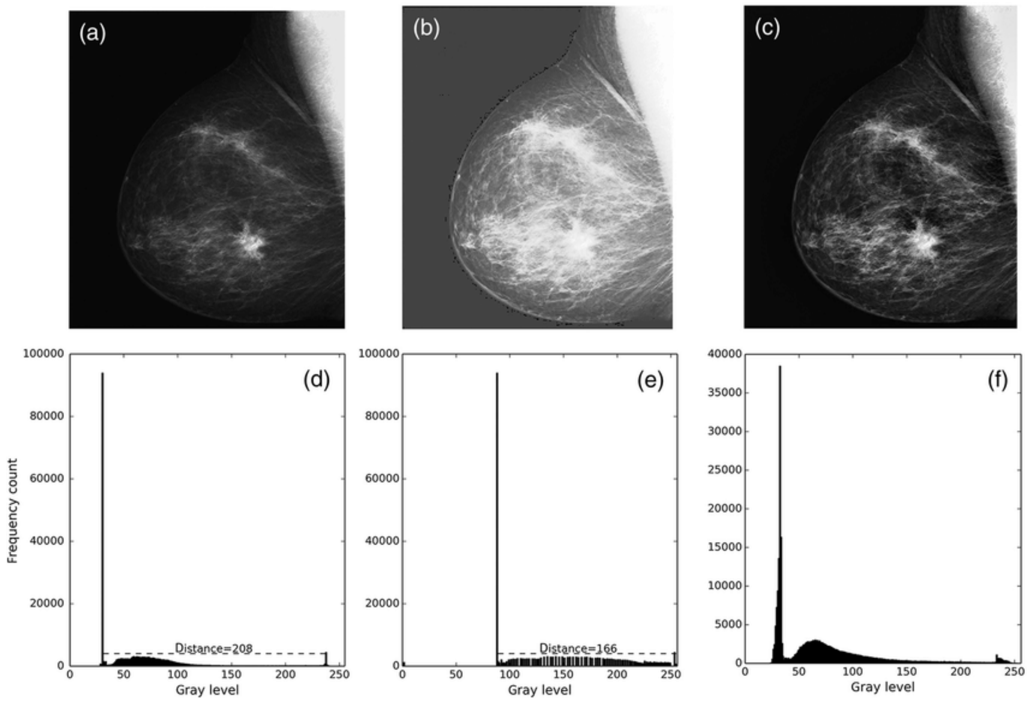

Figure 10: Ways of histogram equalization [3]

The goal of this strategy is to give the cumulative probability function associated with the image a linear trend.The cumulative probability function (cdf) is used in the processing of histogram equalisation . The cdf is defined as the sum of all probabilities in its domain and is calculated as follows:

$$cdf(x) = \sum_{n=-\infty}^{x} P(k)$$

- **Image Sharpening** (shown in Figure 11). Kernels, also known as masks and convolution matrix, are used in the process of image sharpening. The technique of sharpening, image detection and others, require that a kernel should be applied to the pixels of image. This is why this method is also knows as Convolution- the process of kernel being applied to the image.

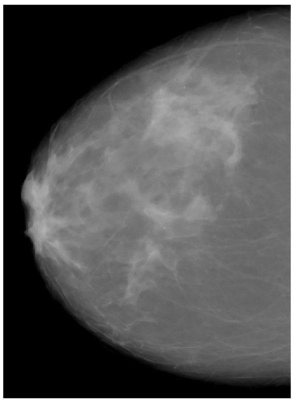

Figure 11: Result of applying 3x3 kernel to sharpen the image

There is a general formula for convolutions, which is as follows:

$$g(x,y) = \omega * f(x,y) = \sum_{dx=-a}^{b} \sum_{dy=-b}^{b} \omega(dx,dy) f(x+dx, y+dy)$$

The type of kernel used in the operation of image affects the image processing task. This is because, the kernel is responsible for bringing a change in formation of pixel and pixel intensity when it is multiplied with the pixels in the original image.

The kernel we used is:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

The reason we chose image sharpening as our main image augmentation technique was because: to eliminate blurring caused by camera equipment, to bring attention to certain regions, and to improve readability. Moreover, training Denovo models took less computational resources and gave significantly good results. However, pretrained CNN architectures took fairly more computational resources and performed poorly on our dataset.

4.3 Convolutional Neural Network Architecture

4.3.1 De Novo models

In order to build a Convolutional Neural Network, we have used Tensorflow, a framework for building Deep Learning models. From Tensorflow's libraries, we implemented Keras Conv2D class to define the layers and their hyperparameters. During the learning process in deep learning, a hyperparameter is a value that for a parameter that is used to regulate the learning process.

In Convolutional Neural Networks, there is a **Convolution layer** which is the foundation layer for the Convolutional Neural Network. It has a set of kernels which are learnt throughout the training. *filters* parameter is an integer value which determines the number of filters that convolution layer will learn from. Additionally, it also tells the number of resultant filters in the convolution. Next, *kernel_size* determines the dimension of the kernel. In our case, we have kept it 3 throughout. The *padding* parameter can be either 'valid' or 'same.' When the parameter is set to "valid", the input volume isn't zero-padded, and the spatial dimensions are allowed to shrink naturally by convolution. Lastly, *activation_function* is a basic function that converts its inputs into outputs with a specific range of values. The final value given out by a neuron is determined by this function. We have set this value to be *relu* which is Rectified Linear Unit Activation Function. If the relu function receives any negative input, it returns 0; however, if the function receives any positive value x, it returns that value. ReLu function is shown graphically in Figure 12 and can be written as: $f(x) = max(0,x)$

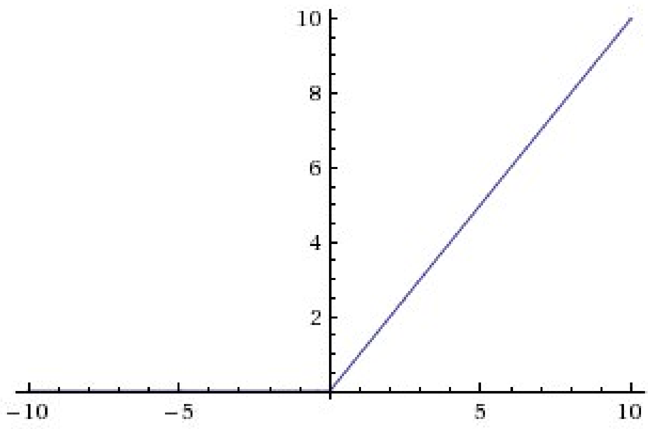

Figure 12: ReLu Function

After the convolution layer, there is a **MaxPooling2D Layer**. This layer downsamples the input along its spatial dimensions (height and width) by obtaining the maximum value for each channel of the input over an input window (of size determined by *pool_size*). *strides* are used to adjust the dimensions of each window. We have used pool_size to be (2,2) and strides at default setting, i.e. 2.

Dropout Layer can be used with convolutional layer after pooling layer. It is a hyperparameter which determines the likelihood of the layer's outputs being dropped out, or, conversely, the probability of the layer's outputs being preserved. Its value varied since it was tuned to improve the results of our models.

Flatten Layer flattens the input and it does not bring an impact on the batch size,

The regular deeply interconnected neural network layer is the **Dense Layer**. It is the most preferred and often used layer. The following operation is performed on the input by the dense layer, and the output is returned.

$output = activation(dot(input, kernel) + bias)$, where:

- **input** represent the input data
- **kernel** represent the weight data
- **dot** represent numpy dot product of all input and its corresponding weights
- **bias** represent a biased value used in machine learning to optimize the model

- **activation** represent the activation function.

Source: Keras Dense Layer Document

The final Dense layer has *sigmoid activation function*. The sigmoid function is shown in Figure 13 and is stated as: $S(x) = \frac{1}{1+e^{-x}}$

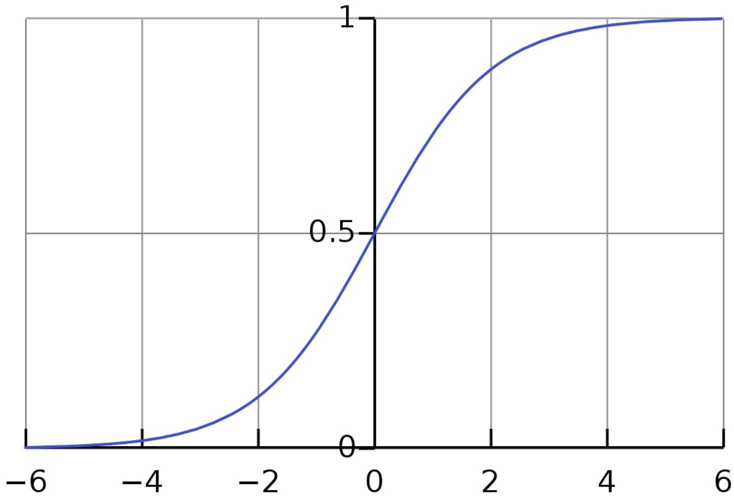

Figure 13: Sigmoid Function

As shown in Figure 13, the function's input is converted to a value between 0.0 and 1.0. Inputs that are significantly bigger than 1.0 are changed to 1.0, and values that are significantly smaller than 0.0 are reverted to 0.0. The function's shape for all potential inputs is an S-shape ranging from zero to 1.0. Since our focus is binary classification, i.e. benign and malignant, using sigmoid activation function is the best choice; where the model will predict 0 if the output class is benign and predict 1 if the output class is malignant.

4.3.2 Transfer Learning

Transfer learning is a machine learning algorithm which allows an existing model to be reused to solve a problem by using the knowledge from the previous model to help perform a new task. There are numerous benefits to using transfer learning which include better efficiency by utilizing saved resources. Unlabeled datasets can also be trained using the pre-trained model.

Transfer learning allows reducing the amount of training data that Is required for new learning models with the use of a pre trained model. When large sets of data are not available to any organization, with transfer learning the medel can be trained on the available dataset and then applied to any similar unlabelled data. With transfer learning, training does not have to begin from scratch every time, this allows an increase in efficiency as well as allowing the saved time and resources from the pre trained model to be used to train multiple models through the knowledge held by the previous model. Knowledge from various pre-trained models can be taken by researchers and blended into training an finer and newer model for any specific task or problem. A much more powerful and accurate model can result from the sharing of knowledge between the models.

For models that require training in real world scenarios and environments, transfer learning can be used for digital simulations which are also less time consuming and less expensive. Real life actions and environments can be recreated in simulations to allow the model to train.

We started off by training our dataset on **Inceptionv3 model**, whose total parameters were 22M. In these total parameters, trainable parameters were 72K which performed poorly on our dataset.

Next, we trained **VGG-16 model**, whose total parameters were and trainable were .

Lastly, we trained **ResNet50 model** whose total parameters were and trainable were . The baseline architectures are the same as the above plain nets, expect that a shortcut connection is added to each pair of 3×3 filters as shown in Figure 6.

To conclude our methodology, we have downloaded dicom images of CBIS-DDSM and converted them into png format. We performed data wrangling in which we did data munging that split the data into train and test images, further to get outperformed results we have applied a different set of augmentation techniques like the histogram equalizer and Image Sharpening technique. After the preprocessing phase, we moved to the training phase and we had different options at your disposal, to either use a pre-trained model or train our own model on the dataset. We chose to use the second option first and trained our own model using De novo architecture and found the baseline result, keeping that baseline result we used State-of-the-art (SOTA) Pre-trained model (i.e Transfer learning) like VGG16, InceptionV3 and Resnet50, and compared the result with the different tool of performance measure.

A manifest file with mammogram images in DICOM format was downloaded and after converting them to PNG 8 folders were created of the mammograms which contained separate sections for test and train data. The dataset was then augmented in two ways, Histogram Equalization and Image Sharpening. Histogram Equalization is a digital image processing method to improve image contrast while Image Sharpening is a technique of sharpening and image detection applied to the pixels of images. Tensorflow, a framework for building Deep Learning models, has been used for building a Convolutional Neural Network (CNN) as well as the implementation of Keras Conv2D class to define the layers and their hyperparameters. The layers of CNN were used to create our model which include the Convolution layer, MaxPooling2D Layer., Dropout Layer, Flatten Layer and Dense Layer. Inception-v3 is a transfer learning architecture for scaling up convolutional networks, in the context of the Inception architecture and has been used. VGG16 architecture has also been used for the purpose of object detection and classification.

Chapter 5

5 Results

Two augmentations techniques were used: Histogram Equalization and Image Sharpening method (refer to Section 4.2) on Denvo models and Transfer learning models. Results on Denovo models are shown in Table 2.

Image Agumentation	Conv. layers	Dropout value	Optimizer	Accuracy
Image Sharpening	4	0.2 and 0.3	RMSProp	68.8%
Image Sharpening	5	0.1	RMSProp	74.1%
Image Sharpening	5	0.2	Adam	71.3%
Histogram Equalization	4	0.1 and 0.2	RMSProp	70.1%

Table 2: Results on Denovo models

Results on the pretrained Convolutional Neural Network Architectures are shown in Table 3.

Image Agumentation	CNN Architecture	Optimizer	Accuracy
Image Sharpening	Inception-V3	RMSProp	42.7%
Image Sharpening	VGG16	RMSProp	59.7%

Table 3: Results on Transfer Learning models

Comparison graphs of accuracy and loss on our models and pretrained Convolutional Neural Networks are shown in Figure 14 and Figure 15.

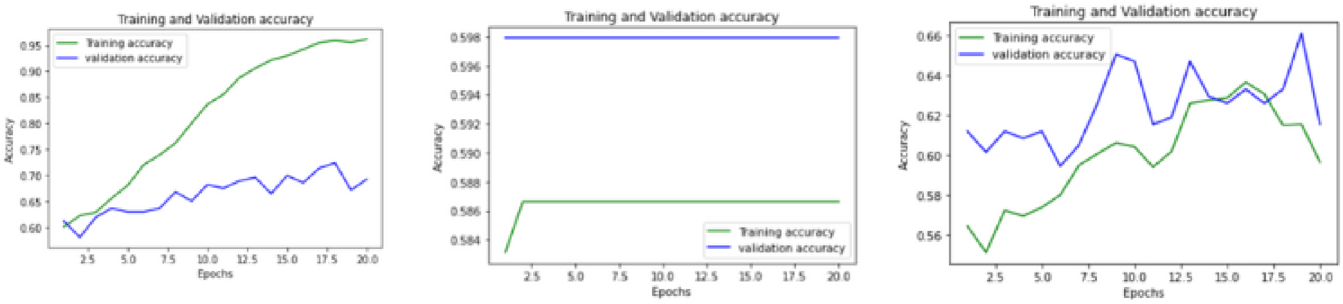

Figure 14: Training accuracy of De Novo, VGG16 & Resnet50 model

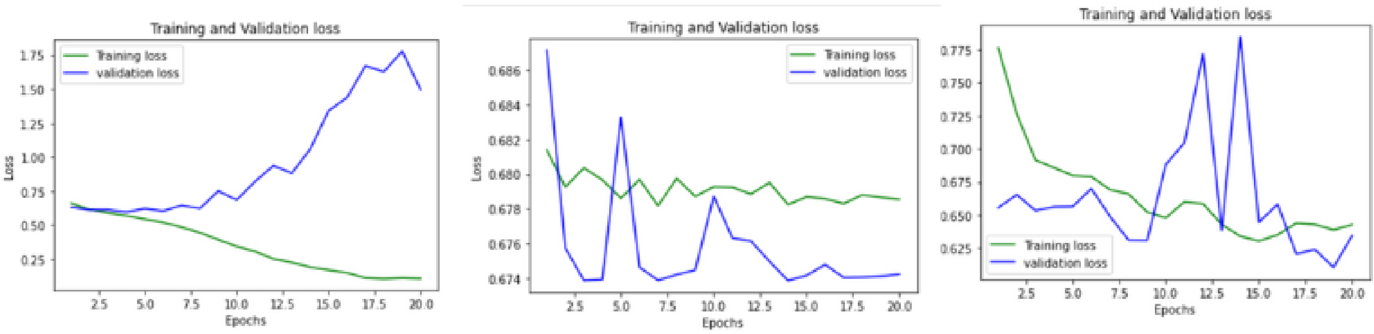

Figure 15: Training loss of De Novo, VGG16 & Resnet50 model

For further evaluation, Using the second model in Table 2, we trained it on adding cropped images of the dataset

Chapter 6

6 Conclusion & Future Work

Breast cancer is a serious public health issue and one of the leading causes of death among women. To decrease the mortality caused by this terrible disease, early diagnosis and identification, adequate control mechanisms, and cure are required. Breast cancer is diagnosed using a variety of imaging modalities, including mammograms, ultrasound, magnetic resonance imaging, and histopatholic images, among others. Traditionally, pathologists and radiologists examine breast pictures manually and make choices based on the opinion of other medical specialists. However, physically examining a large number of breast photos for suspected breast cancer detection is a time-consuming and inefficient procedure that frequently results in false positive or false negative results. As a result, an automated system is constantly needed to speed up the image processing process and assist radiologists in early breast cancer diagnosis. In addition to these automated methods, radiologists can get a second opinion and make more solid, dependable, and accurate conclusions about breast cancer diagnosis [24].

We started off by downloading the CBIS-DDSM dataset [1] and converting it to our desired png format. The dataset already had a 80-20 train-test split and we performed two image augmentations technique before training our deep learning models. Convolutions Neural Network (CNN), which is one of the most popular and frequently used Deep Learning architecture for medical imaging analysis (particularly breast imaging analysis). We focused on image sharpening and histogram equalization to augment the breast images and we passed these images to our Denovo models and Transfer learning models. Our denovo models performed fairly better than the pretrained architectures (results in Table 2 and Table 3) and took less compute.

Black box decisions (such as recommendations / decisions made by AI-based automated systems) are usually not preferred in medical diagnosis (such as breast cancer diagnosis) because radiologists / physicians are primarily interested in knowing and understanding how the decision was made and on what factors it was based [62]. The automated AI based systems for disease prediction / diagnosis (decision support systems) are frequently considered as a danger or loss of control by physicians due to their less explainable nature and a variety of other considerations such as losing control over autonomous decisionmaking [63]. Physicians ranked explainability of AI algorithms high many years ago [64], calling it one of the most essential features of AI-based automated decision support systems.

Today, many academics propose "opening the dark box of artificial intelligence" and focusing on the use of interpretable models for making high-stakes judgments. As a result, in order to earn physicians' trust in AI algorithms' judgments and to justify their dependability, a thorough explanation of the decisions is required, particularly when the algorithms are employed to forecast disease.

References

[1] A. H. D. R. Rebecca Sawyer Lee, Francisco Gimenez, “Curated breast imaging subset of ddsm [dataset],” *The Cancer Imaging Archive*, 2016.

[2] A. W. A. W. G. M. H. F. N. M. A. A.-g. F. Z. G. R. Ghulam Murtaza, Liyana Shuib and N. A. Azmi, “Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges.” *Artificial Intelligence Review*, pp. 1–66, 2019.

[3] A. Toet and T. Wu, “Efficient contrast enhancement through log-power histogram modification,” *Journal of Electronic Imaging*, vol. 23, p. 063017, 12 2014.

[4] J. J. Pavithra S., Vanithamani R., “Computer aided breast cancer detection using ultrasound images,” *Materials today Proceedings*, vol. 33, pp. 4802–4807, 2020.

[5] R. M. L. H. X. C. H. D. H.D. Cheng, X.J. Shi, “Approaches for automated detection and classification of masses in mammograms,” *Pattern Recognition*, vol. 39, pp. 646–668, 2006.

[6] J. L. J. T. Rachael F.Brem, Megan J. Lenihan, “Screening breast ultrasound: Past, present and future,” *American Journal of Roentgenology*, vol. 204, pp. 234–240, 2014.

[7] A. A. L. L. M. Y.Gao, K. J. Geras, “New frontiers : An update on computer-aided diagnosis for breast imaging in the age of artificial intelligence,” *American Journal of Roentgenology*, vol. 212, p. 300–307, 2019.

[8] G. H. Y. LeCun, Y. Bengio, “Deep learning,” *Nature*, vol. 521, pp. 536–444, 2015.

[9] A. M. A. Shrestha, “Review of deep learning algorithms and architectures,” *IEEE Access*, vol. 7, pp. 53 040–53 065, 2019.

[10] K. L. S. Targ, D. Almeida, “Resnet in resnet: Generalizing residual architectures,” *In International Conference on Learning Representations, ICLR 2016*, p. 1–4, 2016.

[11] K. H. X. Zhang, J. Zou and J. Sun, “Accelerating very deep convolutional networks for classification and detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 1943–1955, 2016.

[12] H. C. S. et al, “Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1285–1298, 2016.

[13] D. CeArley, B. Burke, S. Searle, and M. J. Walker, “Top 10 strategic technology trends for 2018,” *The Top*, vol. 10, pp. 1–246, 2016.

[14] A. S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on mri,” *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019, special Issue: Deep Learning in Medical Physics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0939388918301181

[15] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, “Deep learning for health informatics,” *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[16] Z. M. Nayeri, T. Ghafarian, and B. Javadi, “Application placement in fog computing with ai approach: Taxonomy and a state of the art survey,” *Journal of Network and Computer Applications*, vol. 185, p. 103078, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804521000989

[17] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, “Convolutional neural networks for breast cancer detection in mammography: A survey,” *Computers in Biology and Medicine*, vol. 131, p. 104248, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521000421

[18] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of research and development,*, vol. 3, p. 210–229, 1959.

[19] Y.-W. C. H. L. G. B. K. J. B. S. June-Goo Lee, Sanghoon Jun and N. Kim, "Deep learning in medical imaging: general overview," *Korean journal of radiology*, vol. 18, p. 570, 2017.

[20] Z. A. Bradley J Erickson, Panagiotis Korfiatis and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, p. 505–515, 2017.

[21] A. C. M¨uller and S. Guido, "Introduction to machine learning with python: a guide for data scientists," *O'Reilly Media, Inc.*, 2016.

[22] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, "Comparative analysis of breast cancer detection using machine learning and biosensors," *Intelligent Medicine*, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667102621000887

[23] T. Xiaoli, "The role of artificial intelligence in medical imaging research," *BJR open*, vol. 2, 2019.

[24] S. M. Shah, R. A. Khan, S. Arif, and U. Sajid, "Artificial intelligence for breast cancer analysis: Trends directions," *Computers in Biology and Medicine*, vol. 142, p. 105221, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522000130

[25] N. Yassin, S. Omran, E. El Houby, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review," *Computer Methods and Programs in Biomedicine*, vol. 156, 12 2017.

[26] A. F. M. Agarap, "On breast cancer detection: An application of machine learning algorithms on the wisconsin diagnostic dataset," in *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, ser. ICMLSC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 5–9. [Online]. Available: https://doi.org/10.1145/3184066.3184080

[27] A. Sharma, S. Kulshrestha, and S. Daniel, "Machine learning approaches for breast cancer diagnosis and prognosis," in *2017 International Conference on Soft Computing and its Engineering Applications (icSoftComp)*, 2017, pp. 1–5.

[28] S. Sharma and R. Mehra, "Conventional machine learning and deep learning approach for multiclassification of breast cancer histopathology images—a comparative insight," *Journal of digital imaging*, vol. 33, p. 632–654, 2020.

[29] S. G. J. L. Z. Z. Tsung-Han Chan, Kui Jia and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE transactions on image processing*, vol. 24, p. 5017–5032, 2015.

[30] X. Z. G. W. Yushi Chen, Zhouhan Lin and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected topics in applied earth observations and remote sensing*, vol. 7, p. 2094–2107, 2014.

[31] B. F. K. Jr, "Guest editorial: discovery and artificial intelligence," 2017.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[33] Hecht-Nielsen, "Theory of the backpropagation neural network," in *International 1989 Joint Conference on Neural Networks*, 1989, pp. 593–605 vol.1.

[34] G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials*, vol. 13, pp. 27–31, 1994.

[35] V. K. Daniel Svozil and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and intelligent laboratory systems*, vol. 39, pp. 43–62, 1997.

[36] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, p. 47, 2007.

[37] T. V. N. Yasin Yari and H. T. Nguyen, "Deep learning applied for histological diagnosis of breast cancer," *IEEE Access*, vol. 8, p. 162432–162448, 2020.

[38] R. A. Khan, A. Crenn, A. Meyer, and S. Bouakaz, "A novel database of children's spontaneous facial expressions (liris-cse)," *Image and Vision Computing*, vol. 83-84, pp. 61–69, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885619300137

[39] Y. Liu and X. Yao, "Ensemble learning via negative correlation," *Neural networks*, vol. 12, p. 1399–1404, 1999.

[40] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014. [Online]. Available: https://arxiv.org/abs/1409.4842

[41] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, "Breast cancer detection using extreme learning machine based on feature fusion with cnn deep features," *IEEE Access*, vol. 7, pp. 105 146–105 158, 2019.

[42] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," 2016. [Online]. Available: https://arxiv.org/abs/1602.07261

[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[44] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016. [Online]. Available: https://arxiv.org/abs/1610.02357

[45] H.-c. Shin, H. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, 02 2016.

[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*,

F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015. [Online]. Available: https://arxiv.org/abs/1512.00567

[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[50] D. K. . C. S. W. Ayana, G., "Transfer learning in breast cancer diagnoses via ultrasound imaging." vol. 13, p. 738, 2021.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *CoRR*, vol. abs/1603.05027, 2016. [Online]. Available: http://arxiv.org/abs/1603.05027

[52] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152.

[53] A. Alalshekmubarak and L. S. Smith, "A novel approach combining recurrent neural network and support vector machines for time series classification," in *Innovations in Information Technology (IIT), 2013 9th International Conference on*. IEEE, 2013, pp. 1–66.

[54] R. M. S. J. H. S. M. A. M. . A. M. H. Nasir, I. M., "An optimized approach for breast cancer classification for histopathological images based on hybrid feature set," *Current medical imaging*, vol. 17, pp. 136–147, 2021.

[55] N. I. Y. W. H. Mehrdad Moghbel, Chia Yee Ooi and N. Memari., “A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography,” *Artificial Intelligence Review*, pp. 1–46, 2019.

[56] H. B. M. S. Ehsan Kozegar, Mohsen Soryani and T. Tan, “Computer aided detection in automated 3-d breast ultrasound images: a survey,” *Artificial Intelligence Review*, pp. 1–23, 2019.

[57] A. W. A. W. G. M. H. F. N. M. A. A.-g. F. Z. G. R. Ghulam Murtaza, Liyana Shuib and N. A. Azmi, “Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges.” *Artificial Intelligence Review*, vol. 64, pp. 1–66, 2019.

[58] C. C. Monjoy Saha and D. Racoceanu, “Efficient deep learning model for mitosis detection using breast histopathology images.” *Computerized Medical Imaging and Graphics,*, pp. 29–40, 2018.

[59] G. F. H. A. e. a. Lee, R., “A curated mammography data set for use in computer-aided detection and diagnosis research,” *Sci Data*, vol. 4, p. 170177, 2017.

[60] S. K. F. J. K. J. K. P. M. S. P. S.-M. D. P. M. T. L. P. F. Clark K, Vendt B, “The cancer imaging archive (tcia): Maintaining and operating a public information repository,” *Journal of Digital Imaging*, vol. 26, pp. 1045–1057, 2013.

[61] D. R. Varma, “Managing dicom images: Tips and tricks for the radiologist.” *The Indian journal of radiology imaging*, vol. 22, pp. 4–13, 20112.

[62] S.-A. Pearson, A. Moxey, J. Robertson, I. M. Hains, M. Williamson, J. Reeve, and D. Newby, “Pearson sa, moxey a, robertson j, hains i, williamson m, reeve j, newby ddo computerised clinical decision support systems for prescribing change practice? a systematic review of the literature (1990-2007). bmc health serv res 9: 154,” *BMC health services research*, vol. 9, p. 154, 09 2009.

[63] E. Liberati, F. Ruggiero, L. Galuppo, M. Gorli, M. Gonzalez Lorenzo, M. Maraldi, P. Ruggieri, H. Polo Friz, G. Scaratti, K. Kwag, R. Vespignani, and L. Moja, "What hinders the uptake of computerized decision support systems in hospitals? a qualitative study and framework for implementation," *Implementation Science*, vol. 12, 09 2017.

[64] R. L. Teach and E. H. Shortliffe, *An Analysis of Physician Attitudes Regarding Computer-Based Clinical Consultation Systems*. New York, NY: Springer New York, 1987, pp. 68–85. [Online]. Available: https://doi.org/10.1007/978-1-4613-8674-2_6