BREAST CANCER CLASSIFICATION USING MODIFIED CONVOLUTION NEUTRAL NETWORK TECHNIQUE

BY

EMMANUEL, Lawrence Omonigho
MENG/SEET/2017/7551

DEPARTMENT OF TELECOMMUNICATION ENGINEERING
FEDERAL UNIVERSITY OF TECHNOLOGY MINNA

NOVEMBER, 2021

ABSTRACT

Cancer is the abnormal growth of cells in the body tissue which can spread out rapidly in any part of the body. Some common types include lung cancer, colorectum cancer, prostate cancer and breast cancer, among which breast cancer is the most predominant. Breast cancer forms and grow uncontrollably in the surrounding tissue of the breast. The disease has resulted in several deaths most due to lack of early diagnosis and treatment. In breast cancer diagnosis, accurate detection and classification of the disease have been a major challenge over the years. A lot of time is wasted in making diagnoses due to huge volume of images to analyze as a result of numbers of increasing cancer cases. Some tissues have similar characteristics and formation often occurs in clusters varying between 0.05mm – 1mm in size making it difficult to locate, as a result, classification accuracy rate tends to decrease. Giant strides have been made in improving detection accuracy through application of machine learning algorithms like support vector machines, decision trees which have slightly improved system accuracy but limited when operating on raw image data. It requires features from the image to be first extracted before being fed into the model. Therefore, there is a need to design a system that can automatically learn features and make accurate prediction. The research focus on the use of a more recent approach in medical image analysis using a modified Convolutional Neutral Network (CNN) technique to learn, detect and classify the presence of breast cancer whether they are malignant (abnormal tissue) or benign (normal tissue) with a high precision. Transferring learning approach is adapted instead of building model from the scratch a method which has been proven to work satisfactory in breast cancer classification tasks. The techniques allow one to custom model tailored to a particular task using popular architecture as a baseline structure. The model uses the AlexNet architecture which was modified and tailored to our task. The work uses reflection and rotation as a form of augmentation technique to increase the amount of dataset that is used to train the model. The dataset undergoes a processing operation in order to enhance the low quality of the images before being fed into the model as training sources. The performance of the proposed model on test dataset is found to be; 95.80%, 95.00%, 80.00%, 92.30% and 93.63% for accuracy, sensitivity, specificity, precision and F1score respectively. The results show significant improvement in classification accuracy when compared with existing literature using deep learning techniques and MIAS breast cancer dataset. This work will help doctors in making accurate classification and reduced time wastage that is usually associated with the manual ways of analyzing breast cancer.

TABLE OF CONTENTS

Content	Page
Title Page	i
Declaration	ii
Certification	iii
Acknowledgements	iv
Abstract	v
Dedication	vi
Table of Contents	vii
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
CHAPTER ONE	1
1.0 **INTRODUCTION**	1
1.1 Background to the Study	1
1.2 Statement of the Research problems	4
1.3 Aim and Objectives of the Research	5
1.4 Scope of the Study	5
1.5 Justification for the Study	5
1.6 Thesis Outline	6
CHAPTER TWO	7
2.0 **LITERATURE REVIEW**	7
2.1 Breast cancer	7
2.2 Early Breast Cancer Detection Systems	7
2.3 Machine Learning techniques to medical image analysis	8
2.4 Deep Learning techniques to medical image analysis	13
2.5 Deep Learning challenges and advancement	20
2.6 Summary of existing literature	27

CHAPTER THREE 29

3.0 **RESEARCH METHODOLOGY** 29

3.1 Dataset Acquisition 30

3.2 Image processing using MATLAB 31

3.3 Convolution Neutral network Model Design 35

3.4 General Design Decision 46

3.5 Evaluation of model performance 49

CHAPTER FOUR 52

4.0 **RESULTS AND DISCUSSION** 52

4.1 Presentation of Results and Discussion 52

4.2 Results for Data Acquisition 52

4.3 Results for Breast Cancer classification using CNN 54

4.4 Comparison with existing literature 59

CHAPTER FIVE 61

5.0 **CONCLUSION AND RECOMMENDATIONS** 61

5.1 Conclusion 61

5.2 Recommendations 62

5.3 Contribution to knowledge 62

References 63

APPENDIX A (Some Breast Cancer Classification Simulation Codes) 68

APPENDIX B (Publication) 72

LIST OF TABLES

Table	page
3.1 : Modified CNN architecture	43
3.2 : Training parameters of the CNN system model	46
3.3 : Showing the flow of execution in MATLAB (R2019a)	49
4.1 : Performance metrics parameters	57
4.2 : Performance metrics of model parameters	60

LIST OF FIGURES

Figure	Page
1.1: International Agency for Research on cancer	2
2.2 : Machine Learning-based CAD system	10
2.3 : Support Vector Machine maximal margin classifier	11
2.4 : Artificial neural network model with one hidden layer	12
2.5 : Diagram showing different Neural Network architecture	14
2.6 : Diagram of DBN architecture	15
2.7 : Diagram of RBM architecture	16
2.8 : Architecture of CNN based CAD model	17
2.9 : Typical CNN structure	18
2.10 : Example of Maxpooling operation	19
2.11 : Typical AlexNet CNN architecture	21
2.12 : Breast segmentation task with the original and segmented images	23
3.1: Detailed processes of the system model	29
3.2: The three basic stages of the system model	30
3.3: Mammogram image adopted from MIAS database	31
3.4: MIAS dataset class distribution	32
3.5: Workflow for image augmentation in datastore	34
3.6: Rotation Transformation	34
3.7: Stacking of the CNN model layers	35
3.8: Convolution operations using kernal size 3x3	36
3.9: Maxpooling operations of 3x3 with stride of 2	37
3.10 : Image resizes operation	38
3.11 : Convolution operation with stride of 2 and zero padding	39
3.12 : Convolution operation with stride of 1 and 2 padding	39
3.13 : Convolution operation with stride of 1 and 1 padding	40
3.14 : Convolution operation with stride of 1 and 1 padding	40
3.15 : Flatten operation	41
3.16 : Image classification using Softmax classifier	41
3.17 : Modified CNN architecture used for the model	42
3.18 : Back propagation operation	45

4.1 : Samples from original MIAS dataset 52
4.2 : Pixel intensities ranging from 0 to 255 before normalization 53
4.3 : Pixel intensities ranging from 0 to 1 after normalization 53
4.4 : Augmented data samples 54
4.5 : Plot showing Training and Validation Accuracy 54
4.6 : Training and Validation Loss function 55
4.7 : MATLAB plots showing training progress, validation and percentage accuracy 56
4.8 : MATLAB plots showing training progress, validation loss and accuracy 56
4.9 : Performance metrics showing precision, Recall and F1score of the model 58
4.10 : The model ROC curve 59
4.11 : Plot showing the comparison between studies and proposed model 60

LIST OF ABBREVIATIONS

Abbreviation	Meaning
ANN	Artificial Neutral Network
AE	Autoencoder
BCDR	Breast Cancer Digital Repository
BI-RADS	Breast Imaging Reporting and Data Systems
CAD	Computer Aided Detection
CBIS	Curated Breast Imaging Subset
CBIS	Curated Breast Imaging Subset
CNN	Convolution Neutral Network
CUDA	Compute Unified Device Architecture
DBN	Deep Belief network
DDSM	Digital Database for Screening Mammography
DT	Decision Trees
ELM	Extreme Learning Machine
FCRN	Full Complex valued Relaxation Network
GLCM	Gray Level Co-occurrence Matrices
GoogLeNet	Inception Network
GPU	Graphical Processing Units
KNN	K-Nearest
LSTM	Long Short-term Memory
MIAS	Mammographic Image Analysis Society
NB	Naive Bayes
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
ResNet	Residual Neutral Network
RBM	Restricted Boltzmann Machine
RF	Random Forest
RNN	Recurrent Neutral Network
VGG	Visual Geometry Group
YOLO	You Look Once

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the Study

Cancer causes an abnormal growth of cells and often spread out rapidly in other part of the body tissues. It is usually named after the part of the body in which it occurs. At the early stage of development, it forms a lump or mass, often referred to as tumor (Saeed *et al*., 2021). The lump produces no significant pain when it is treated early. Hence, conducting screening is crucial for early detection of the disease formation (Saira *et al*., 2018). Cancer occurs in both men and women but women are more likely susceptible to it, especially the breast cancer, which is considered as the most common type of cancer all over the world (Simon *et al*., 2020). According to the estimate made by American Cancer Society (ACS) about 1,898,160 new cases of invasive cancer were diagnosed for both women and men and 608,570 cancer deaths was recorded as reported on U.S Breast Cancer Statistics, (Rebecca *et al*., 2021). Each year, approximately 14.1 million people suffer from breast cancer worldwide which results to 8.2 million deaths. It is expected that by the year 2025, there will be about 19.3 million newly reported cancer cases globally.

Most of the newly reported cancer cases occurred in developing nations, (Bray *et al*., 2018) which is expected to be on an increasing rate due to lack of awareness of the risk factors such as population growth, age and other factors such environmental toxins, medical history, family history, genetics, exposure to radiation and infection. The accurate cancer statistics of Nigeria is relatively unknown due to lack of adequate funding by cancer registries, although some research studies have been conducted in different parts of the country. Jedy-Agba *et al*., (2018) reported Age Standardized incidence Rate (ASR) for all invasive cancer as 58.3/100,000 men and138.6/100,000 women while 66.4/100,000 men,

130.6/100,000 women are noninvasive cancer rate as conducted in Abuja and Ibadan based cancer registries, respectively.

In 2018, the World Health Organization (WHO) conducted a comprehensive study on cancer statistics of Nigeria, including both sexes and age, stated that there are about 115,950 new cancer cases and 70, 327 recorded death cases. The study also shows a glance of other types of cancer cases in the country represented in a form of pie chart in the Figure1.1.

Figure 1.1: International Agency for Research on Cancer GLOBOCAN (Bray *et al*., 2018)

One of the ways of preventing an abnormal growth of cells (benign tumor) from becoming a full-blown cancer (malignant tumor) is by early detection, (Burcu *et al*., 2018). A lot needs to be done to educate women about the importance of self-examination and the significance of early disease detection. Over the years, doctors have made giant strides in early diagnosis and treatment of the disease and in reducing breast cancer mortality rate by using medical diagnostic imaging tools such as computer-aided detection, Digital

mammography, Magnetic resonance imaging and ultrasound. Among which digital mammography procedures is considered to be faster and more accurate in detection, (Susan *et al*., 2017). A mammogram is an x-ray of breast indicating changes in breast tissues that can result to cancer. It helps in detecting very small lumps that can rarely be seen by the human eyes. In the past years, researchers have tried to automate the procedure by reducing the huge task of the radiologist in breast cancer diagnosis. No matter the expertise of the radiologists in examining mammograms, other limiting factors such as eye fatigue, interpretational errors, distractions need to be minimized as such could lead to false interpretations making the radiologist to give description that can harm the patients. For instance, breast biopsies are usually prescribed to patient diagnosed as cancerous. However, over 40 - 60% of extraction of cells in breast tissue is diagnosed as noncancerous, which betray the necessity for correct mammography diagnosis (Ragab *et al*., 2019). On the other hand, cancer to be reported while in reality there is no malignant tissue this could lead to unnecessary treatment admitted to the patient all together as reported by (Elter and Horsch 2019).

Recent studies show that the use of Computer-Assisted Detection (CAD) software can greatly reduce the number of false interpretations and thereby increasing the accuracy of digital mammography diagnosis by Shen *et al*., (2020). Accuracy improvement in detection and classification of breast cancer is still a major challenge as reported in the work of (Saira *et al*., 2018). The focus behind this research is to design a model that can accurately detect and classify breast cancer with higher degree of accuracy using modified Convolution Neural Network Technique. This can help the radiologist to make accurate diagnoses. This can ultimately, prevent late treatment due to false negatives and as well unnecessary treatments in cases of false positives.

1.2 Statement of the Research Problem

Breast cancer is common among women, leading to the major causes of death. Most of the death cases result from lack of early detection and diagnosis. Over the years, researchers have made great effort in early diagnosis and treatment of the disease using medical diagnostic tools like digital mammography. Mammogram images are formed by varying dissimilar factors like background interference, lightning variation and illumination changes (Suresh *et al*., 2018). The tumor formation often occurs in clusters varying between 0.05mm

– 1mm in size which may also be located in a dense tissue, this posed a huge challenge to the radiologist in identifying and recognizing the cancer region due to the large volume of images as a result of increasing cancer cases (Teshale *et al*., 2019). As a result, detection accuracy rate tends to decrease. Giant strides have been made toward improving classification accuracy through application of machine learning algorithms like support vector machines, K-Nearest Neighbor and decision trees. The approach tried to automate the process by using a separate feature extraction method which is followed by classification algorithm, in theory, this approach slightly increase the accuracy of screening mammogram (Falconi *et al*., 2019). However, these newly developed algorithms could not effectively operate on a raw data because it required important features like edges, shapes, textures and colors to be first extracted from the image by human before being fed into the models (Wang

et al., 2020). Most recent approach involves the use of deep learning techniques it can automatically learn the features from the mammograms directly. It has greatly increased the accuracy of mammogram screening for early signs of breast cancer detection. However, the major limiting factors of these techniques are that it requires high computing resources, large amount of data to train, making them hard to optimize. Hence,

the research seeks to use a modified Convolution Neutral Network Technique for breast cancer classification.

1.3 Aim and Objectives of the Research

The aim of this study is to improve on classification accuracy in determining the likelihood of cancerous areas from mammogram images using modified Convolution Neutral Network Technique. To achieve this, the following objectives are to:

I. Acquire dataset from Mammographic Image Analysis Society (MIAS) database website and process the dataset using MATLAB

II. Design a deep learning model architecture using modified Convolution Neutral Network for breast cancer classification

III. Evaluate the performance of the model using system accuracy, sensitivity, specificity and F1-score

IV. Compare the research work with existing research in literature using deep learning and MIAS breast cancer dataset

1.4 Scope of the Study

The scope is limited to breast cancer classification using deep learning in MATLAB environment which does not require physical development.

1.5 Justification for the Study

Breast cancer is one of the most common forms of cancer amongst women with statistics indicating that 1 in 7 females will be diagnosed with breast cancer in their lifetime. Early detection of breast cancer through screening tests such as mammograms is an efficient way to maximize patient survival rate by treating the disease. However, no matter the expertise of radiologists examining mammograms, external factors such as fatigue, human error and interpretation error need to be minimized (Hepsag *et al*., 2017). Error-

prone mammogram interpretations by radiologists can lead to decisions that can ultimately harm the patients. To that end, using Computer-Assisted Detection (CAD) software can help minimize the number of wrong interpretations and increase the accuracy of mammography screening is justifiable as an area of interest.

1.6 Thesis Outline

The remaining part of the thesis is organized as follows; Chapter two reviews Breast cancer, Early Breast Cancer detection systems, Machine Learning applications to medical image analysis, and Deep Learning applications to medical image analysis and also related works of literature on Breast cancer classification using Convolutional Neutral Networks. Chapter three presents the research system model, and methodology of Breast cancer classification using Convolution Neutral Network techniques. Chapter four presents and discusses the MATLAB simulation results of the research and chapter five states the conclusion and recommendations of the research.

CHAPTER TWO

2.0 LITERATURE REVIEW

This chapter will explore relevant existing literatures that have been carried out in the breast cancer diagnosis which is outline as follow; Breast cancer, Early Breast Cancer detection systems, Machine Learning applications to medical image analysis, and Deep Learning applications to medical image analysis using Convolutional Neutral Network (CNN) and also give a summary of the reviews.

2.1 Breast Cancer

Breast cancer is an uncontrollable growth in breast cells forming a lump which can either be benign when it is a normal tissue or malignant (cancerous tissue), (Saeed *et al*., 2021). The most popular type of cancer among women is the breast, lung and colorectal cancer among which breast cancer account for 30% of all the cancer cases resulting to a lot of deaths, (Suresh *et al*., 2018). In a report by the World Health Organization, predicted that by the year 2025, there will be 19.3 million new cases which are expected to be increasing annually, (Siegel & Selvathi 2020). In women within the age of 40 years and above breast cancer is widely spread, fortunately, the likelihood of women's death cases at a premature stage is about 3% (Gbenga *et al*., 2017). Hence, early detection of breast cancer can increase patient survival rates, (Maleika *et al*., 2021). Statistics have shown that the survival rate of breast cancer within the first and second stages could reach up to 98% and 93%. However, the rate significantly drops to 73% and 22% with the third and fourth stages respectively, (Saravanan, 2019).

2.2 Early Breast Cancer Detection Systems

Early signs of breast cancer before the appearance of any symptoms like lumps, lesion or tumor are usually detected by test screenings. In breast cancer diagnosis, the main tool used

for screening is the mammograms; these are carried out around the breast area, this scan reveals backgrounds in black and calcifications or masses in a dense white. Most times if an area looks suspicious mammograms are followed by Magnetic Resonance Imaging (MRIs) for detailed analysis of the area. Biopsies can be conducted to confirm if any of the suspicion areas have a potential presence of cancer, Biopsy is the extraction of tiny part of breast tissue for further analyses to obtain a definite results by (Martin and Laura 2019). Biopsies are usually invasive in nature so it ideal for patients to use imagery tools like mammograms for early breast cancer detection. However, the accuracy of this relies on the competence of radiologists, (Osareh and Shadgar 2019). Mammograms are 2D images representation of the breast which leads to the superposition of the breast tissue making it even more difficulty for a radiologist to make accurate analyses. In 1970s, the CAD software was introduced to assist radiologists in their interpretations. However, in the 1990s CAD systems consisted of modeling of pixels to construct ruled based systems that did not offer much more knowledge than the expert radiologist (Litjens *et al*., 2017).

2.3 Machine learning techniques to medical image analysis

Machine learning algorithm can be classified into supervised and unsupervised learning, in supervised learning all the sample in the dataset are labeled (Tang and Rangayyan 2019). In supervised learning, the aim is to determine the optimal parameter θ to minimize a loss function defined as, $(\bar{\ },)$ which corresponds to the error between $\bar{\ }$ and . Classification and regression are the major applications of supervised machine learning.

2.3.1 Supervised Machine Learning based Systems

Supervised machine learning techniques started replacing these CAD systems in the late 1990s, where advanced algorithms are used to recognize the hidden pattern that could not

be interpreted by the radiologists, (Litjens *et al*., 2017). These methods were becoming very popular and accepted over the statistical approaches like regression for classification task problems (Huang, 2019), especially with large datasets. This era marked the shift from system that is design by humans to systems which were trained on medical imagery datasets. However, these newly developed algorithms could not effectively operate on a raw data because it required important features like edges, shapes, textures and colors to be first extracted from the image by human before being fed into the models (Yu *et al*., 2020).

2.3.2 Unsupervised Machine Learning based Systems

Unsupervised learning means the data sample are unlabeled only the input features x are known while the label y are predicted (Litjens *et al*., 2017). The algorithm will separate them into different groups by creating a cluster in the dataset. Anomaly detection, clustering, data visualization are main applications of unsupervised learning.

2.3.3 Machine learning Algorithms

Over the years, many algorithms have been developed and tested leading to improvement in detection accuracies in breast cancer diagnoses (Yue *et al*., 2018). The commonly use machine learning algorithm consist of Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (kNN), Artificial Neural Networks (ANN) and Decision Trees (DT). The model building process in machine learning CAD system is illustrated in the Figure 2.2.

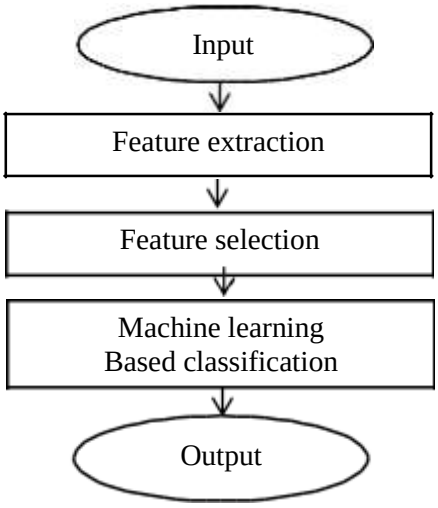


Figure 2.2: Machine Learning-based CAD system adopted from (Lian Zou 2019). The procedure of using a pre-trained ML-based CAD model can be described mathematically as follows:

Step I: the input (I_x) of the suspicious region is quantified with scalar variable ($E(I_x)$) by using feature extraction

Step II: feature dimension is decreased by the application of feature selection (S) algorithm to ($E(I_x)$) resulting in ($S(E(I_x))$

Step III: the output label (y) is predicted using machine learning classifiers which can be formulated as $y = F(S(E(I_x)))$. (Tang *et al*., 2019).

(a) Decision Trees

The decision algorithm worked by creating a tree of different types of nodes performed by sending a specimen down in each tree and mapping samples input feature, attributes and values. A classification decision is made after each node tests one of the feature vectors attributes branching out to a deeper note based on the value (Ling *et al*., 2015).

(b) Support Vector Machine

This process involves finding the hyper plane that divide two classes used to separate non-linear data. An SVM is made of a maximal margin classifier which linearly separates the plane by applying kernel trick. The kernel matrix operates by mapping the input features vectors to a higher dimension using dot product. SVMs become most popular machine learning algorithm in medical imagery analysis due to kernel matrix operation (Yue *et al*., 2018). The visual representation of maximal margin hyper plane linearly separating tumors is shown in Figure 2.3

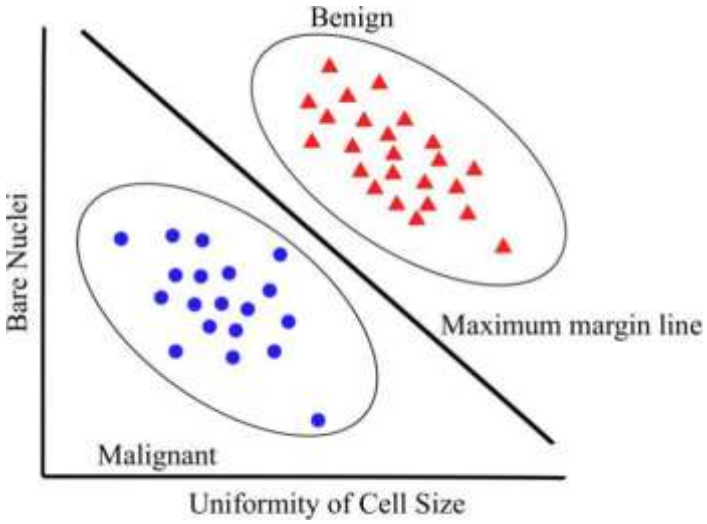


Figure 2.3: Support Vector Machine maximal margin classifier, adopted from (Yue *et al*., 2018)

(c) Artificial Neural Networks

The concept of ANN began with the idea that every aspect of learning or intelligence can be simulated by a machine (Szegedy *et al*., 2017). In 1957 McCarthy developed a programming language that allows one to create flexible programs that present basic operations but this development was too few to be tested and was referred as dark period of ANN because the creating intelligence machine was still uncovered. Within a few decades after this period, there was a significant development on the subject such as disease

diagnosis which gives the basis of modern day general human-level Artificial intelligent (AI). In the 1980s, AI began gaining more applications in practical projects like solving real life problems. As technology advances, research in the field of AI is paving way to a new development known as deep learning.

ANNs are layer structures inspired by the human brain. It is also referred to as Multi-layer Perceptron (MLP), (Wu, *et al.*, 2018). This shows a great potential when applied to medical image analysis.

The human brain is similar to a machine which constitute complex processing neurons combined together to perform task every day. Thus, it is on this basis that the ANN was developed to learn by example like the human brain, (Huang 2019). It can also be designed for a particular task like image and pattern recognition by adjusting the relationship between and among structured layers. The basic structure of AI normally constitute of three main kinds of layers: input, hidden, and output layer nodes, especially when used as a classifier where the input features is matched

to the output.

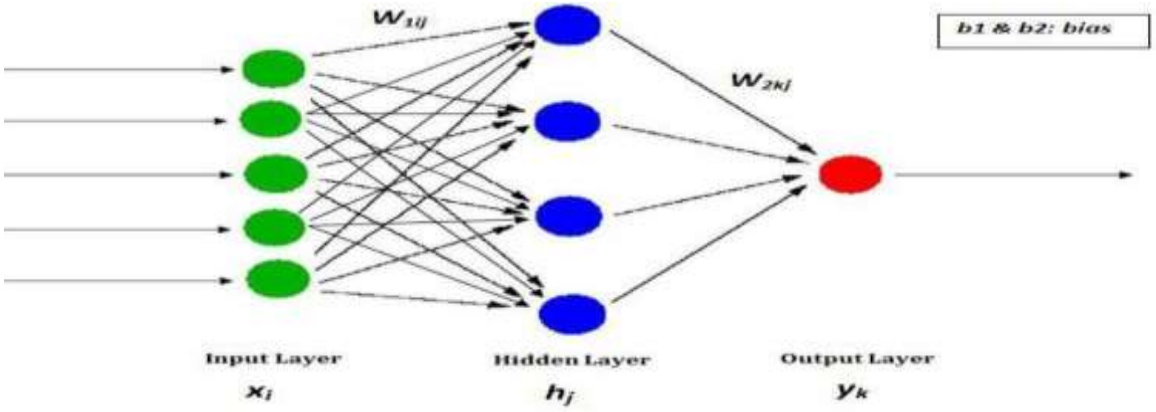


Figure 2.4: Artificial neural network model with one hidden layer, source: (Huang 2019)

The input variables in Figure 2.4 denoted as $X_{i,}$ are mapped unto the hidden layer $h_{j,}$ for further processing and then to the output variables, Y_k, by the following expression:

$$h = (\sum_{1} + _{1,}) \quad (2.1)$$

$$_k = (\sum_{2} \quad h \quad + _2) \quad (2.2)$$

Where f and g are the activation functions, W_{1ji} and $W_{2kj,}$ are weight matrices, and b_{1j} and b_{2k} are bias parameters.

The ANNs minimizes the loss *L* function by the principle of back propagation techniques, the loss function is sent back through the network while weights are adjusted in order to reduce the loss.

Logically, the next stage of development of breast cancer detection systems was for the model to learn how to extract these features automatically, this idea paved way to advances in deep learning techniques (Yala *et al*., 2019). However, due to high computational requirements like Graphical Processing Units (GPU) and very large datasets these models have been limited until recent years, (Litjens *et al*., 2017).

2.4 Deep learning techniques to medical image analysis

Most recent approach in medical image analysis is the use of deep learning techniques, it can automatically learn the features from the mammograms directly. It has greatly increased the accuracy of mammogram screening for early signs of breast cancer detection.

2.4.1 Deep learning Architectures

The commonly used neural network architecture are Autoencoder (AE), Convolution Neural Network (CNN), Deep Belief Network (DBN), Recurrent Neural Network (RNN) and Restricted Boltzmann Machine (RBM) among which the Convolution Neutral Network (CNN) is the most used method for medical image analysis due to its great potential in

image processing. The classification of different types of deep Learning architectures are presented in Figure 2.5.

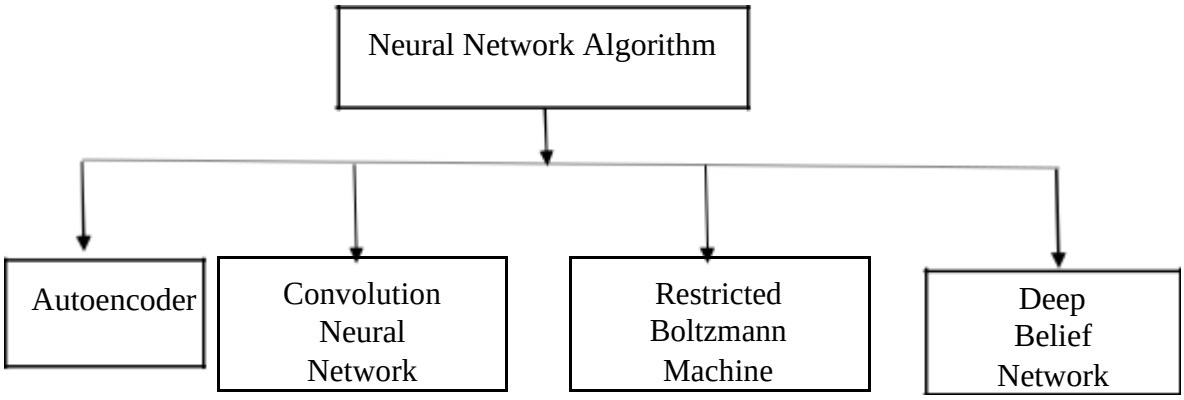


Figure 2.5: Different Neural Network architecture

2.4.2 Autoencoder (AE)

It is an auto-learning algorithm used to reduce the dimension of data by a method of coding. The idea is the data is first converted into an abstract representation and later formatted by the encoder function to produce back the original data, (Bengio *et al*., 2014). One key merit of this model is that it has the capability of continuous feature extraction and equally filtering noise information. AE algorithm perform basic functions: the first is to learn features using unsupervised learning and the second is to optimize the network using supervised learning method, (Ling *et al*., 2015).

2.4.3 Deep Belief Network (DBN)

The idea behind the development of DBNs was from Bayesian probabilistic model, consisting of multiple layers of stochastic and latent variables. The DBN architecture is made of layers in a top-down approach, whereas the training process is also into two different stages: the first stage is known as the pre-training stage which involves the feature extraction in an unsupervised manner and the second stage is the fine tuning of the network. The performance of this model can further be improved by adjustment of network

parameters. However, performance is highly limited when input data are clamped. In order to overcome this problem, Huang (2019) proposed a greedy layer-by-layer learning algorithm to optimize the weights of a DBN to the size of the network. This algorithm has significantly reduced the training time of dataset. The diagram of the model is shown in Figure 2.6.

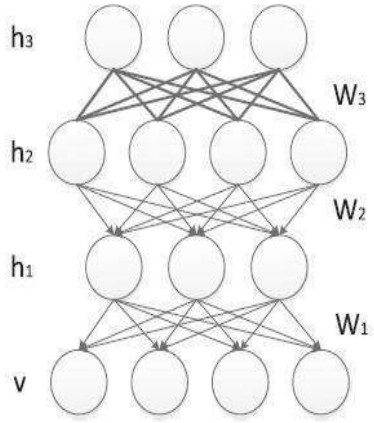


Figure 2.6: DBN architecture, adopted from (Deng *et al*., 2012)

2.4.4 Restricted Boltzmann Machine (RBM)

It is a stochastic model that can be used to learn the probability distribution with respect to their inputs. It is also used to simplify the arrangement of network layers and to improved model efficiency. The RBM algorithm architecture consists of a visible unit and the hidden layer with no connection between units from the same layer as shown in Figure 2.7. The idea behind the algorithm is the data generated from RBM is trained by performing Gibbs sampling and updating units in other layers until state of equilibrium is achieved.

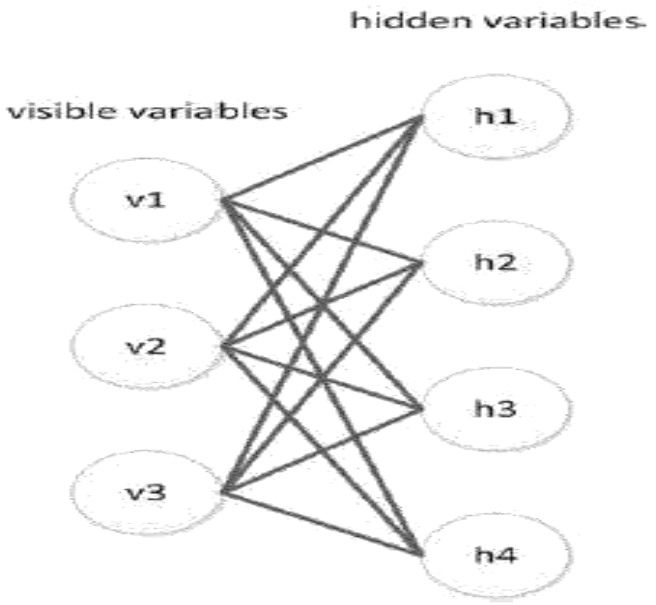


Figure 2.7: Restricted Boltzmann Machine architecture, adopted from (Dina *et al*., 2019)

2.4.5 Convolution Neural Network

ANNs slightly differ from CNNs in terms of connections, CNNs are partially connected. CNNs are generally 2-dimensionals of several channels data processing algorithms which has shown excellent performance in images and videos processing. The idea is greatly inspired by the human visual cortex organization. In the work of (Shin *et al*., 2016), they observed that the cell arrangement in the animal visual cortex in charge of light detection is in a sub-region of the visual field. In CNNs architecture, the complexity of the network is greatly reduced compared to the standard neural network due to the fact that the general matrix multiplication in neural network has been replaced with a convolution. Furthermore, the feature extraction in the standard machine learning algorithm is completely truncated since raw images can be directly imported to the CNNs. It is also noteworthy that CNN model requires less pre-processing, reduced number of parameters in the network by taking

advantage of spatial relationship. Due to technological advancement in computation algorithm, CNNs have been successfully applied in diverse areas like face detection, behavior recognition, and speech recognition and image classifications. The basic schematic of CNN based model is depicted in the Figure 2.8.

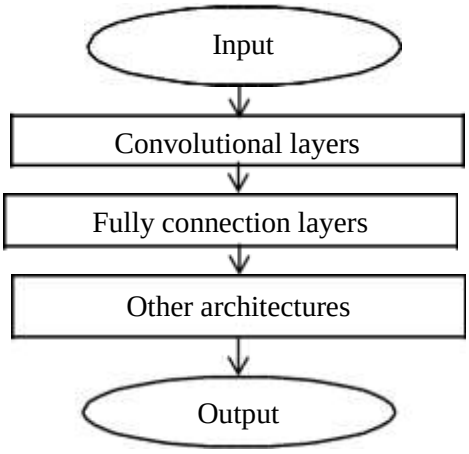


Figure 2.8: Architecture of CNN based CAD model

(a) Convolution Neural Network structure

CNNs architecture is multi-layer neural network which consists of the activation layer, pooling and convolution layer. The size of the input image is reduced to a simple format which can be processed by the fully connected layers by using the pooling and convolutional layers, (Shen *et al*., 2020). Figure 2.9 show a typical example of CNN structure.

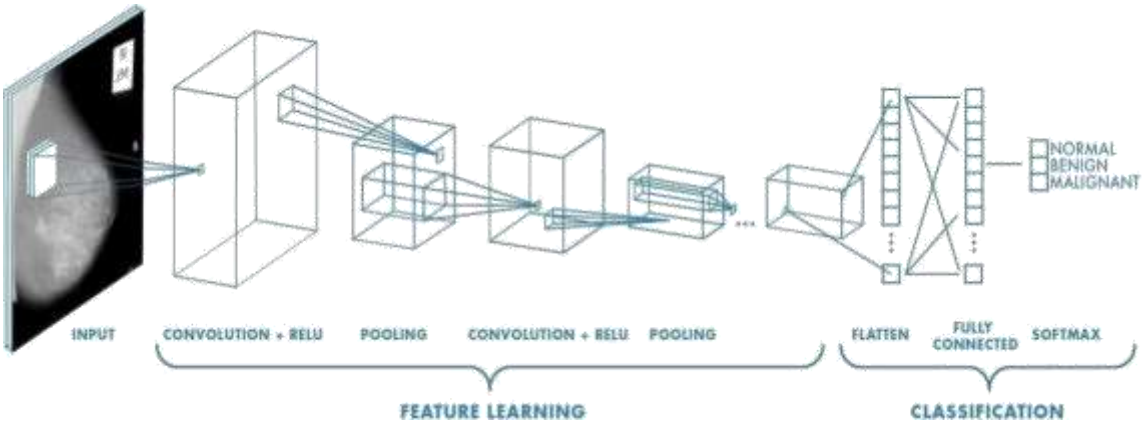


Figure 2.9: Typical CNN structure adopted from (Shin *et al*., 2016).

The commonly used CNNs models are ResNet, GoogLeNet, LeNet, VGG, YOLO, AlexNet, faster RNN and Long Short-Term Memory (LSTM). In CNN based CAD model, instead of learning a different set of parameters at each location as the case in machine learning algorithm, it only needs to learn one set of them and apply it to the whole image. This approach often leads to reduction in computational time, thereby improving system performance.

(b) Convolution layers

Convolution layers connects to only pixels in their receptive fields and do not connect to every pixel in the image. This make CNNs to first focus on low-level features and gradually assembled into high-level features as they get deeper, (Shen *et al*., 2020).

The mathematical operation that govern convolution, is the movement of filter or (kernel) *f* over an image *I* to calculate a weighted sum as shown in Equation 2.3. The weights are learned during the training stage by using optimization techniques like gradient descent. These kernel allow a layer of neurons to highlight the area of the image that activate the filter the most which can be detected in the layers input

$$\text{ii}(\ ,\ )=(\text{ii}\cdot\ )(\ ,\ )=\sum.\sum\text{ii}(\ ,\ )\cdot(\ -\ ,-\ ) \qquad (2.3)$$

The training of CNN model involves the use of error gradient algorithm which reduces error by calculating the error between the expected and actual values of the output and then adjusts its weight (Krizhevsky *et al*., 2012).

(c) Activation Function

Activation functions are functions in a neural network that decide if a particular neuron can be fired or not, (Szegedy *et al.,* 2015). It performs such operation by computing the weighted sum of input and biases. Activation function can be grouped into linear or non-linear depending on the function it represents. Examples of activation functions are Sigmoid, Tanh and Rectified linear Unit (ReLU).

(d) Pooling layer

This technique is used to downscale the image as it moves through the network which is done by calculating an aggregate of its inputs based on a function. There are basically main types; maximum pooling which returns the maximum value and average pooling returns the average values. Example of Maxpooling is illustrated in Figure 2.10

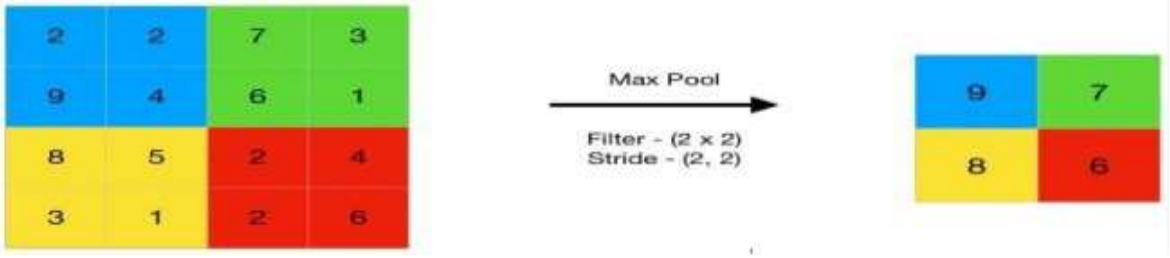


Figure 2.10: Example of Maxpooling operation

(e) Training Neural Network

In ANN model, the dataset used undergoes training and testing operation which is done in order to find the optimum model parameters. Training operation usually involves the use of validation dataset for tuning the model parameters to avoid over fitting effects (Wang *et al*., 2020). ANN model is application specific, once the network is configured for a particular

application, that network is ready to be trained. At the beginning of the operation, the initial weight is usually randomly chosen. The training process is tested by exposing it to sets of existing data with known outcome.

2.5 Deep learning challenges and advancement

A lot of factors limit the implementation of deep learning models, in order to achieve an acceptable level of performance, the models must be trained on a large set of data. However, labeled datasets require huge amount of time and processing resources like the Graphical Processing Units (GPUs) which are not always available (Krizhevsky *et al.*,2012).

Up until recent years, where the huge success in deep learning is link to the development of GPUs. The GPUs are very powerful processors that have the capability of processing large amount of data in parallel. This was initially designed for video gaming by the Nvidia being one of the main manufacturers of GPUs. The GPUs are good in huge matrix operations making them an excellent choice for deep learning but difficult to use as a low level perspective (Caulfield, 2009). However, an open sources computing libraries is made available through the CUDA online platform.

(a) Regularization techniques

Overfitting of data is one of the problems in deep learning. Data overfit in deep learning is the ability of the model to learn the data too well that there is no significant changes in weight parameters anymore. To solve this, a new concept was introduced known as regularization techniques. This is the technique used to reduce the complexity of a model, which is done by penalizing the loss function.

(b) Transfer learning

This technique allows one to transfer attribute gained in a larger dataset to be used in a similar domain which is common use when only a small amount data is available (Falconi *et al*., 2019).

This technique allows one to custom CNN model tailored to a particular task using popular CNN architecture as a baseline structure. There are many CNN architectures available in Keras like ResNet50, GoogLeNet, AlexNet, VGG19, InceptionV3 which are designed for general classification of 1000 different categories (Krizhevsky *et al.,* 2012). Figure 2.11 shows a typical AlexNet architecture, it contain five (5) convolution layers and three (3) fully connected layers for classification.

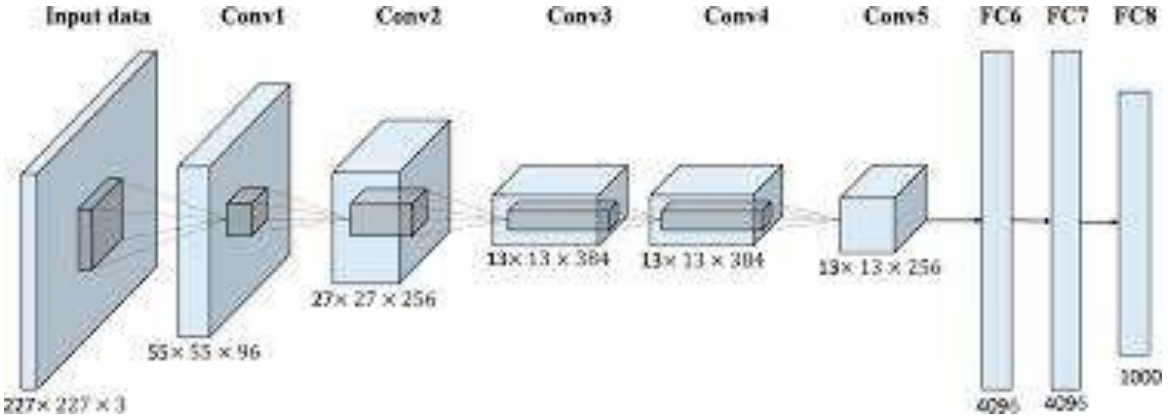


Figure 2.11: Typical AlexNet CNN architecture, image retrieved from

(Diaz *et al*., 2018)

This technique can easily be implemented to any classification task by replacing the dense output layer which makes the actual prediction with the layer containing one neuron per class to predict since the actual models are initial design to classify millions of images across thousands of different classes.

In Diaz *et al*., (2018), the authors, uses two ResNet-50 models to be tested on two different dataset one on Curated Breast Imaging Subset of Digital Database for Screening

Mammography (CBIS-DDSM) and the other on ImageNet, results show that the model using transfer learning obtained 84% accuracy as compared to 75% accuracy clearly demonstrating the merit of using CNNs with pre-trained weight parameters.

(c) Dropout technique

This technique is implemented by randomly ignoring neurons in any layer during the forward and backward propagation. This help to minimize neurons from too much co-adapting with their neighbors weight thereby causing overfitting.

(d) Data augmentation

This is another technique used to solve the problem of small datasets, done by increasing the amount of data in the dataset. In image processing, data augmentation is creating new training data from the existing training data. This is done by creating transformed copies of the images in the training dataset that belong to the same class as the original image. Data augmentation is domain specific and the type of a particular technique must be chosen carefully in context and knowledge of the problem domain.

2.5.1 Review of related works

Over the years, many algorithms have been developed and tested leading to improvement in detection accuracies in breast cancer diagnoses.

(a) Machine learning applications to Breast Cancer classifications

Machine learning tasks that are common to medical imagery analysis are detection and classification (Litjens *et al*., 2017). In machine learning task, these two terms detection and classification can be used interchangeably, depending on the type of dataset. The classification is binary when there only two classes and multiclass for more than two (Lee *et al*., 2017). In Digital Database for Screening Mammography (DDSM) the classification task is a multi-class where images are categorized as benign (non-cancerous tissue),

normal, malignant (cancerous tissue). On the other hand, segmentation task is the classification of each pixel belonging to a particular class to without a distinction (Geron, 2019). The segmentation task in machine learning consists of several processing blocks like the feature extraction, feature selection and feature segmentation. (Moura and Guevara 2013). Segmentation task can highlight masses like calcifications, fibroadenomas. An illustration of segmentation task is as shown in Figure 2.1 revealing masses in the image (Punitha *et al*., 2018).

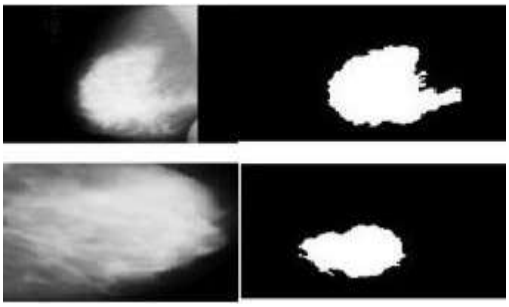


Figure 2.12: Breast segmentation task with the original and segmented images, adopted from (Punithaet *et al*., 2018)

In Ghongade and Wakde (2017), the authors combined two different machine algorithms the Random Forest (RF) and Extreme Learning Machine (ELM) technique to detect breast cancer abnormalities in mammograms images using MIAS dataset. The classifiers worked by sending irregular samples down to each node assigning a weighted attribute value to it and at the end a vote of all trees indicate classification into either benign tumor or malignant tumor. Result shows that RF classifier works well with large datasets and high dimensional data but the combination of RF-ELM gives a better result. The overall detection accuracy achieved for the work is at 89%.

In Sarosa *et al*., (2018), the authors compared the effects of machine learning algorithms on different dataset. When SVMs is coupled with grey-level co-occurrence matrices (GLCM)

as the feature extraction techniques applied to CBIS-DDSM datasets gives 63.03% accuracy, on the other hand, when SVMs with texture-based features was applied on a smaller dataset like MIAS 92% accuracy was obtained. This clearly depicts the importance of feature selection and image processing techniques for machine learning algorithms. In Karashif (2018), the authors first removed noise from the MIAS dataset using 2D median filter. They combined two different image segmentation algorithms edge-based and region based segmentation on MIAS dataset before applying multiple machine learning models. The SVM outperforms others with higher accuracy of 90% and F1-score of 95%. In Cong *et al*., (2017), the authors combined three different machine learning algorithms the KNN, SVM and NB for breast cancer classifications, the work is to know the effect of integrated classifiers with respect to individual classifiers. Results indicate that the integrated classifiers outperforms with higher classification accuracy and sensitivity when compared to single classifiers.

(b) Deep learning applications to Breast Cancer classifications

Recent approaches in breast cancer diagnosis make use of deep learning techniques to detect and classify images by the presence of possible lump or masses formation. One of the most popular deep learning algorithms used in breast cancer classification is the Convolution Neutral Network (CNN) due to its great potential in image processing (Saira *et al*., 2018). In breast cancer detection, there are basically two forms of mass tumor, the benign tumor which often appears round or oval shape and malignant tumor (cancerous) which is partially rounded and irregular outline white in colour (Tang *et al*., 2009). Over the years, researchers have proposed several methods for breast mass classification using mammogram images, (Sharka and Al-Sharkawy 2011) but one of the most important parameters for consideration is the classifier accuracy. Dina *et al*., (2019) worked on breast

cancer detection using deep CNNs and support vector machines for classifying mammography images into benign and malignant mass tumors. The aim of the work was to extract features from the image using two different techniques: the region of interest and threshold, and the last fully connected layer of the DCNN architecture were replaced with support vector machine. The work was implemented using two different databases, the digital database for screening mammography (DDSM) and Curated Breast Imaging Subset (CBIS). The result shows that the overall accuracy was improved to 87.2% when support vector machine was used as a classifier in combination with DCNN as against 71.01% when feature extraction was done manually.

In a work proposed by Al-antari *et al*., (2017), a Computer-Aided Diagnose (CAD) system was used with a combination of CNN. The work is broken down into four stages: feature extraction using CNN, mass detection with confidence model and mass classification using fully connected neural network and it is implemented using public available mammograms images from DDSM on YOLO-based CAD system. The image pre-processing was done using the multi-threshold peripheral equalization techniques to remove the noise background from the images. The YOLO-based CAD system performs two basic operations which are to find the mass locations and classify them according to their features into benign or malignant. A total of 600 breast images was used for both training and testing purposes containing 300 benign images and 300 malignant images, the results for both detection and classification was evaluated and about 85.52% overall accuracy was achieved.

In Jiao and Geo (2016), the authors proposed and demonstrated a CAD system based on convolution neural network to classify benign and malignant masses of breast cancer. The developed model was trained using the combination of two different CNN layers to extract

high and low deep features from mammogram images published a research paper on masses detection in mammograms images using object detector based on a deep convolution neural network. This was implemented on RetinaNet which consist of a backbone network and two sub networks, the convolution feature extraction is done on the backbone network while the sub network performs box bounding regression. One of the advantages of using RetinaNet architecture is that it is simpler and easy to implement for image classification and regression. Duraisamy & Emperumal (2017), presented a study on tumor tissues classification using the Chan-Vese level set method to first extract the contour in the mammogram images before been fed into DCNN model to learn specific features from the contour of mammograms images. The classification of the mammogram images into ten different classes by a Full Complex-valued Relaxation Network (FCRN) classifier was because the suspicious areas in the ROI are so dense with varying characteristics. The FCRN will learn and generalize them into normal, benign or malignant-calcification. Result shows that this classifier achieved 85% overall accuracy and performed better than most classifiers using the same system parameters, Jiang (2017) work on breast mass lesions classification in mammographic images using CNN of GoogLeNet and AlexNet on Breast Cancer Digital Repository (BCDR) database. The combination of three different algorithms used for various features extraction (histogram-oriented gradients, gray-level co-occurrence matrix and histogram gradient divergence) and their performances was compared and evaluated with area under the curve and receiver operating curve. In Peng Shi *et al*. (2019), the authors combined both vertical and horizontal gradient magnitude weight-arrays based on CNNs architecture to extract features from mammogram images using MIAS dataset. Lastly, the output of four Breast Imaging Reporting and Data Systems (BI-RADS) density classes is obtained. The CNN model

achieved 83.6% accuracy. In Yu *et al.* (2020), the authors classified breast cancer based on two classes benign and malignant using MIAS dataset. They combined median filter and adaptive histogram equalization to enhance the image before being fed into VGG 16 model for classification, the work achieved 87.80%, 94.74% and 91.14% for recall, precision and F1-score respectively and total system accuracy of 89.06%.

2.6 Summary of the existing literature

The summary of existing literature presents the strength and weakness of related works on Breast cancer classification using Deep Learning techniques.

In Saira *et al*. (2018), the authors work on Breast cancer detection in mammogram images using CNN with MIAS dataset. The work uses a selective feature extraction approach based on morphology operation which significantly improved classification results the study uses the original dataset without augmentation, to achieve satisfactory detection accuracy a large dataset is required in deep learning. The detection accuracy is 65.0%.

In Peng Shi *et al*. (2019), the authors worked on small dataset for BI-RADS density classification of mammography images using MIAS dataset based on lightweight Convolution Neural Network. The study uses a density classification approach which is highly related to breast cancer risk factors, but limited in the use of separate image segmentation algorithm which is usually time consuming for deep learning classification, 83.6% accuracy was achieved.

In Saeed *et al*. (2021), the authors worked on Breast cancer diagnostics on deep learning classification for mammogram images using MIAS dataset. The study uses a balanced test data for two classification level but limited in the use of separate segmentation algorithm, 89.06% accuracy was achieved.

In Emperumal and Duraisamy (2017), the authors work on Breast cancer detection in deep learning using MIAS dataset. The study uses chan-vese level approach to extract features from weak boundaries in images before training but limited in the use of separate segmentation algorithm, 93% accuracy was achieved in the work.

The most recent approach is the use of deep learning techniques which can automatic detect and classify the presence the disease.

2.6.1 Research gap

CAD systems have been developed since the 1970s to assist radiologists in their interpretation of mammograms starting with primitive expert systems. The major challenge with these systems is the inability to locate formation of tumor in dense tissues. The size is usually small compared with the size of the mammographic image. For example, a tumor size might range from 0.005mm – 1mm and the image size might be 1024 x 1024 pixels dimension. The expert systems were replaced by supervised machine learning algorithms but required that features to be first extracted from the images manually. The performance of the algorithms largely depends on the quality of the images which could not operate on raw mammograms. Therefore, deep learning algorithms have gained traction recently, notably the CNNs to automatically learn which features to extract from the images, thus preserving their 2D spatial property. However, CNNs require large amounts of data to avoid overfitting, which is not always available especially in medical imagery. In literature, the most popular labeled dataset in medical image analysis are MIAS and Digital Database for Screening Mammography (DDSM). Researchers have used these datasets for breast cancer diagnosis. However, these datasets are also limited in size but with advances in data augmentation techniques significant results have been achieved in recent times.

CHAPTER THREE

3.0 RESEARCH METHODOLOGY

This chapter will provide the logical procedure, description of conceptual block diagram and methods employed in this study. It covers the criteria that are used to design the models, along with the rationale behind the choice of dataset.

The design of the deep learning model implemented for the project is shown Figure 3.1. It represents the proposed breast cancer classification system block diagram. The design processes can be broken down into three (3) basic stages such as image pre-processing (Data loading, Image normalization and Data augmentation), DL model design and Image classification as shown in Figure 3.2. The brief description about the proposed technique is determined as follows.

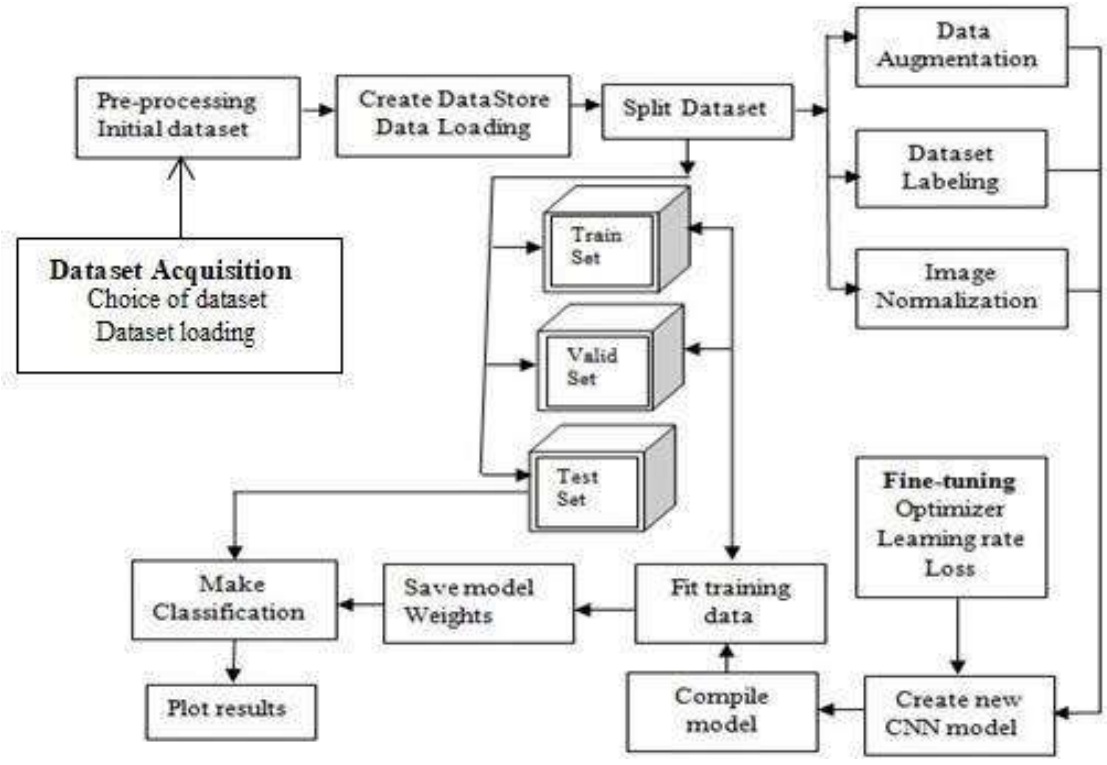


Figure 3.1: Detailed processes of the system model

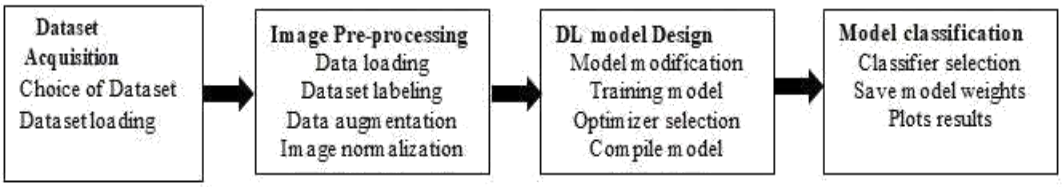


Figure 3.2: The three basic stages of the system model

The first stage in the block diagram is to acquire dataset from a source, which constituent a vital component of the research.

3.1 Dataset acquisition

The design of deep learning model requires the use of life sample information to obtain the attribute and patterns of the data enable it to detect new sample cases and to evaluate its performance therefore, the choice of dataset become a vital factor in designing the model. After widely exploring existing literatures on the subject matter, the Mammographic Image Analysis Society (MIAS) dataset was chosen, the reason being that the sample data contain locations both normal and abnormal cases. It is also an open-source dataset making them easily accessible for research purposes. Although, the data size is relative small, makes it useful for initial prototyping. However, with advances in data augmentation, more data can easily be generated and fed into the model.

The MIAS dataset was developed by the UK Research group for the purpose of medical research. It comprises 322 Portable Gray Map (PGM) images format, 161 pair of patients. Each image has labeled features with background tissue characteristics and class of abnormality whether cancerous tissue (malignant) or non-cancerous (benign). It also contains the locations of these abnormalities such as the x and y coordinates and radius measurement of the pixels, (Vishrutha and Ravishankar 2014). These features are represented in Figure 3.3.

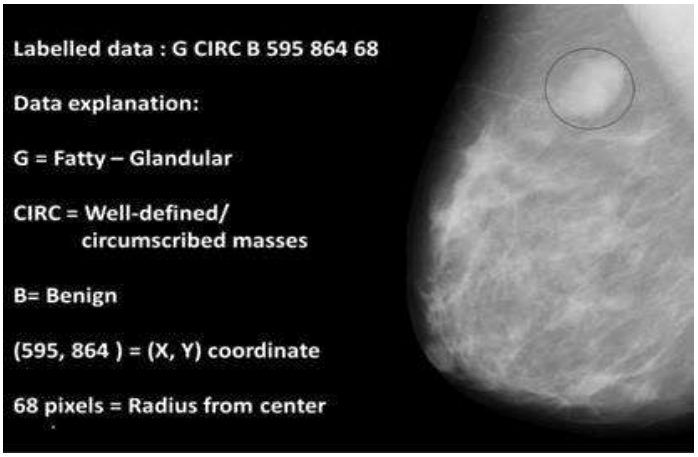


Figure 3.3: Mammogram image adopted from (Cong *et al*., 2017)

3.2 Image processing using MATLAB

This involves loading the dataset into the memory and processing the image label for the classification task. The original image comes with high resolution making training on these datasets not feasible because it will require more computational resources. Accordingly, only 20% or less of this data contains cancerous tissue which can affect performance accuracy of the classifier. In order to overcome this, the images are first resized to (64x64) during import to scale operation. It was also observed that the pixel intensities correspond to integer values ranging from 0 to 255 and the weight parameter in the model are small, having such discrepancies can slow down the training process, and to avoid this problem data normalization technique was applied to make the pixel intensities range from 0 to 1.

3.2.1 Dataset split

The dataset is split between a training set and testing set to avoid any form of data random movement between classes. In a machine learning the practices is to split dataset in 80%/20% rate as reported in breast cancer papers (Yue *et al*., 2018). However, in this work 30%/70% was used for training and testing. The reason is that in the class distribution of the samples in the dataset, there are more normal tissues than benign and malignant

tissues combined together and adopting the 80%/20% will make the training phase to overfit, to overcome this, a 30%/30%/40% was used for training, validation and testing respectively. This also helps to overcome the problem of samples imbalance in different classes which can affect the classification accuracy of the model. This is as shown in Figure 3.4, plotting the class distribution of MIAS dataset to determine whether some classes appear more than other classes

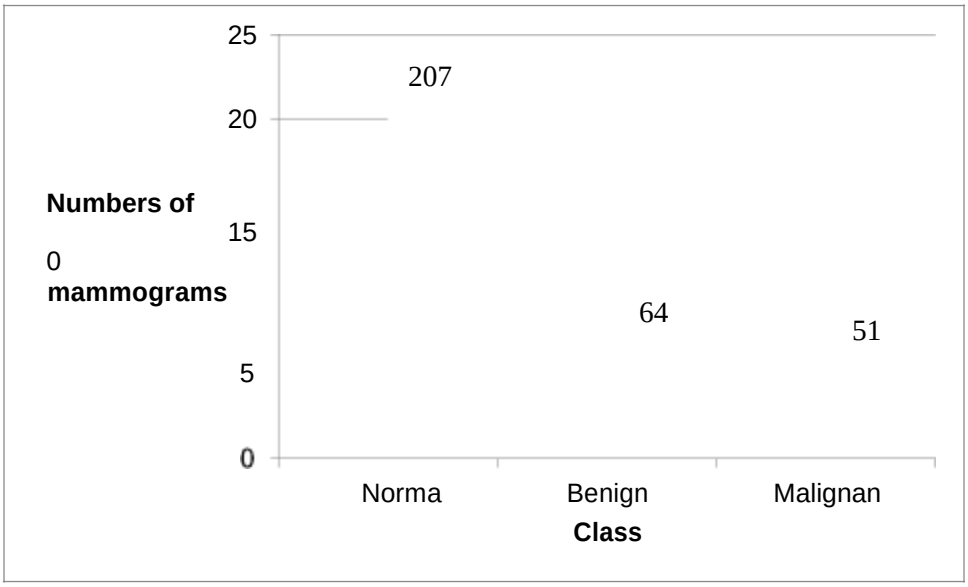


Figure 3.4: MIAS dataset class distribution

3.2.2 Data loading

The MIAS dataset cannot be loaded directly through a single import format rather it is loaded in a datastore in mini-batches before being fed to the sequential CNN model, a practice which speed up processing time.

At the end, data augmentation is applied to increase the amount of data needed to effectively train the CNN model. To obtain a satisfactory performance, a large number of training images is needed to train the deep neural networks. The principle of data augmentation is applied through degree transformation, for each refined breast tumor image another new sample is generated. In tumor classification problem, data augmentation

technique is the most commonly used approach; tumor tissue can be represented in any particular orientation without image deformation.

3.2.3 Data augmentation

In image processing, data augmentation is creating new training data from the existing training data. This is done by creating transformed copies of the images in the training dataset that belong to the same class as the original image. Data augmentation is domain specific and the choice of the particular data augmentation techniques must be chosen carefully in context and knowledge of the problem domain. There are many different augmentation techniques use in image processing like cropping, scaling, translation, color, reflection and rotation. In medical imaging the integrity of the image is of important so the choice of augmentation technique should be based that, in this study, the combination of flipping and rotation was used as a form of augmentation technique, the reason being that during both operations the features and intensities of the image is maintained without any form of distortion.

The sample images are retrieved from the datastore containing the dataset and loaded in mini-batches into the model while the transformation angles are being applied, each of the 322 down sample images was flipped (reflection operation) giving different copies of the images. The down sample images and the reflected images were then augmented through rotation by the application of different transformation angles ranging from 0 to 360 degrees as shown in Figure 3.7. The transformation angles used in this study are 90, 180, 270 and 360 degrees; at these angles augmented copies of images is obtained without deformation of pixel intensity. The down sample images and reflected copies were each rotated through the angles. The images were augmented by a factor of eight generating about 2,576 images for training samples.

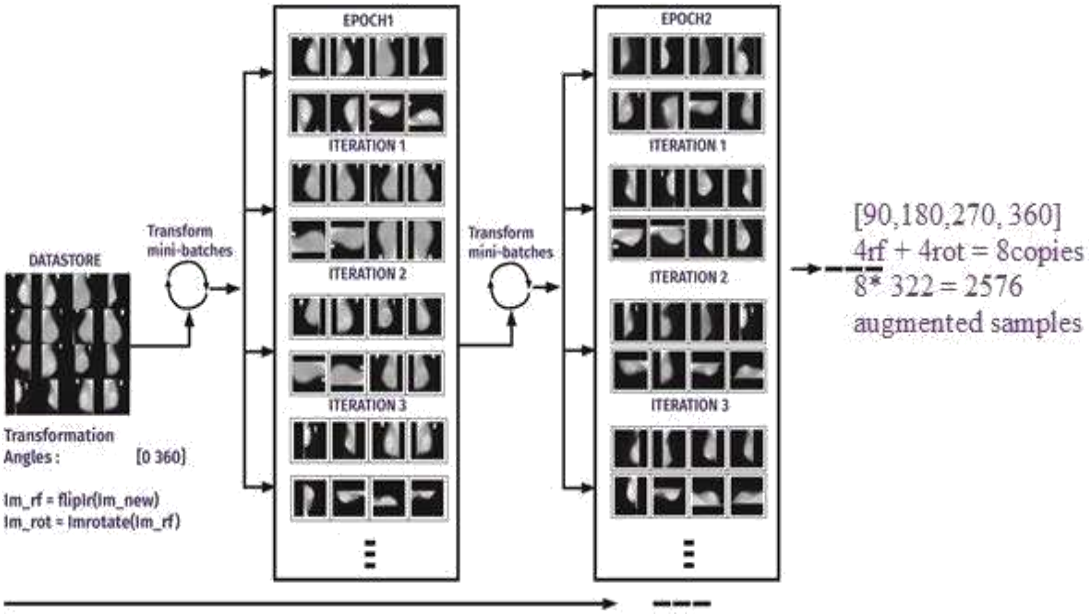


Figure 3.5: Workflow for image augmentation in datastore

Diagrams showing different transformations using MATLAB are as shown in Figure 3.6.

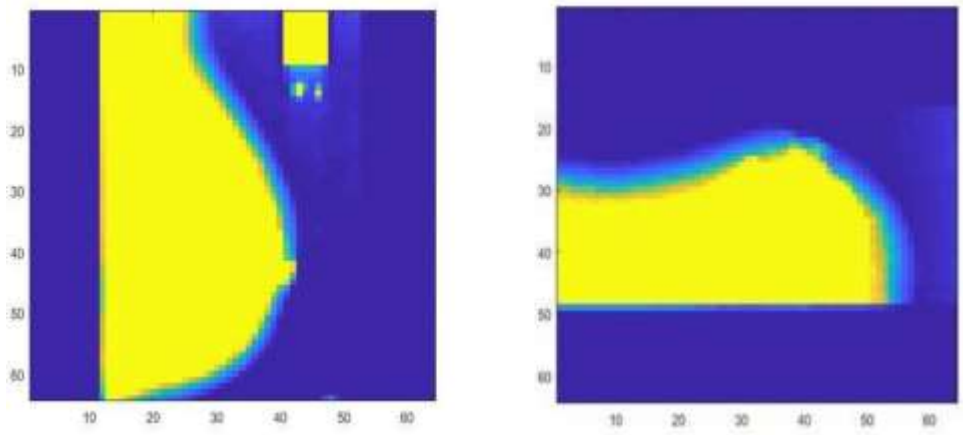


Figure 3.6: (a) Rotation of 90-degree transformation (b) Rotation of 180-degree transformation

The processing operations was implemented using MATLAB R (2019)a software, the choice of this software for data augmentation is that it accepts the dataset image format directly, saving us the need to make conversion to any other format. The augmented images were then exported to Python platform where the CNN model is implemented.

3.3 Convolutional Neutral Network (CNN) Model Design

There are several deep learning algorithms for data processing but CNNs have shown excellent performance in medical image processing which is the reason behind the choice of using CNN for this work.

The CNN model structure consists of three basic layers: convolution, activation and pooling layers and a typical stacking of the layers is shown in Figure 3.7.

Figure 3.7: Stacking of the CNN model layers

The CNN model will automatically learn the underlying patterns of the images from the training set loaded in the memory before making classification. Due to the type of dataset used for this project, transfer learning approach is adapted instead of creating a CNN model from the scratch a method which has been proven to work on breast cancer detection tasks (Falconi *et al.,* 2019). The AlexNet CNN architecture is of interest, the reason being that the structure is very stable and converges very fast when used in a relatively large dataset, it is also easy to implement and requires less computational resources as compared to others.

3.3.1 Model modification

We modified the AlexNet architecture by removing the last three layers of a standard AlexNet architecture and replaced with fully connected layer (FC), soft-max layer and classification layer respectively. In the output layer, a Softmax activation function transforms the output of the last hidden layer into probabilities for each class sum up to 1 (Litjens *et al*., 2017).

3.3.2 Convolution layer

The structure consists of four convolution layers for detecting features or pattern in the image, to compute output image, each neuron performs a dot product of weighted sum or filter (kernel matrix) to the local region which is connected to the input image. Mathematically, the operation is governed by equation (3.1)

$$( ) = (\quad -1^{(l)} \cdot ( )) \tag{3.1}$$

Where;
Ω(j)= Input image
( )= Weight parameters
* = Convolution operator

= Activation function (ReLU)

During the convolution operation, each kernel moves across the length and height of the input image which is guided by padding. The process is shown the Figure 3.8.

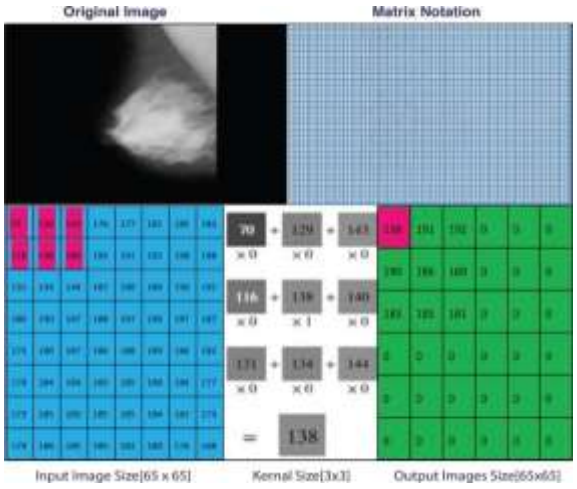


Figure 3.8: Convolution operations using kernel size 3 x 3

3.3.3 Activation Function

In this study, ReLU is used as the activation function, the choice for this is that it offers better performance and generalization in deep learning compared to others like Sigmoid and Tanh function, (Zeiler, *et al*., 2013). The ReLU function performs a threshold operation to each input element of the image pixels where values less than zero are set to zero and thereby eliminating the vanishing gradient problem observed in the earlier types of activation function.

The ReLU is given by:

$$f(x) = \max(0, x) = \begin{cases} 0 \\ \geq 0 \end{cases} \tag{3.2}$$

3.3.4 Pooling layer

The pooling layers are used to reduce the image size to minimize the computational time. In this model, a Maxpooling of [3 3] with stride of 2 was used. The operation is shown in Figure 3.9

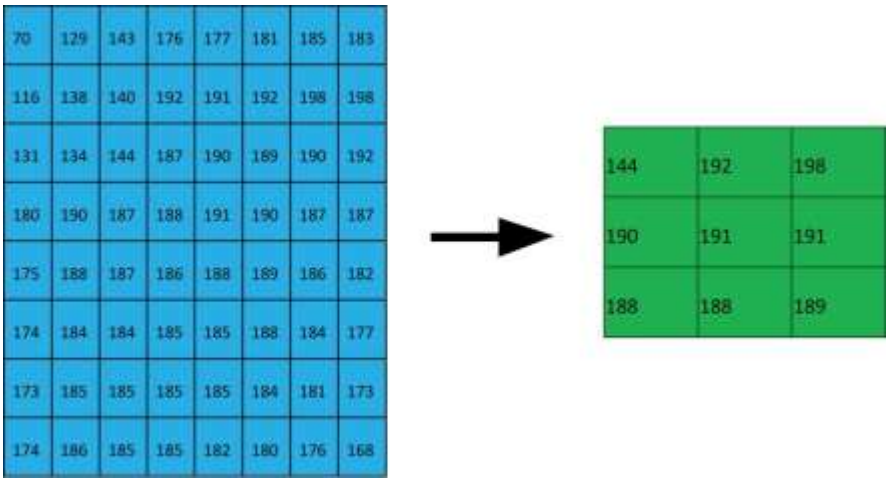


Figure 3.9: Maxpooling operation of 3 x 3 with stride of 2

3.3.5 Output image size

The size of the output image is obtained by:

$$\text{Size of output image} = \left[\frac{W + 2P - F}{S} \right] + 1 \quad (3.3)$$

Where:

W = Image size

P = padding (controls the spatial dimensions of the input volume of image)

F = filter (kernel) size, S = stride

Equation (3.3) is used to obtain the shape and size of the image during each stage of convolution operations from the input to the output layer where classification is made. The original image size was resized to [64x64x3] during import to scale operation as presented in Figure 3.10.

Input image size: [64x64x3]

Figure 3.10: Images resize operation

Layers 1and 2: detect large features and patterns that differ in color and intensity in the mammogram images.

Layers 3, 4 and 5: detect smaller features having irregular shape and size without definite patterns.

Layers 6, 7 and 8: these are last three layers that combined all the features detected by the previous layers and used the parameters learnt to make detailed classification.

Layer 1: CONV with 96 filter, size 5x5, stride 2, zero padding 0

- Output size = ((64 - 5)/2) +1 = 31 is size of image outcome
- Memory: 31 x 31 x 96
- Weights (parameters): 5x 5 x 96 x3

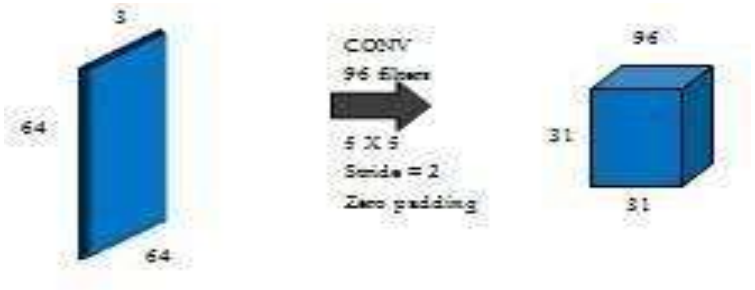


Figure 3.11: Convolution operation with stride of 2 and zero padding

Layer 1**:** 31 x 31 x 3

Layer 2: CONV with 256 filters, size 5 x 5, stride 1, and padding 2

- Outcome Size = (((31+4)- 5)/1) +1 = 31 is size of image outcome
- Memory: 31 x 31 x 256
- Original size is restored because of padding
- Weights (parameters): 5x 5 x 256 x3

Figure 3.12: Convolution operation with stride of 1 and 2 padding

Layer 2: 31 x 31 x 3

Layer 3: CONV with 384 filters, size 3 x 3, stride 1, and padding 1

- Outcome Size = (((31+2)- 3)/1) +1 = 27 is size of image outcome
- Memory: 27 x 27 x 384
- Weights (parameters): 3x 3 x 384 x3

Figure 3.13: Convolution operation with stride of 1 and 1 padding

Layer 3: 27 x 27 x 3

Layer 4: Same as layer 3

- Layer 5: CONV with 256 filters, size 3 x 3, stride 1, padding 1
- Outcome Size = (((27+2)- 3)/1) +1 = 27 is size of image outcome
- Memory: 27 x 27 x 256
- Weights (parameters): 3x 3 x 256 x3

Figure 3.14: Convolution operation with stride of 1 and 1 padding

Layer 5: 27 x 27 x 256 = 186,624 pixels are fed to FC

Layer 6: Fully connected with 4096 neuron

- Memory: 4096 x 3
- Weights:4096 x (27x 27 x 256)

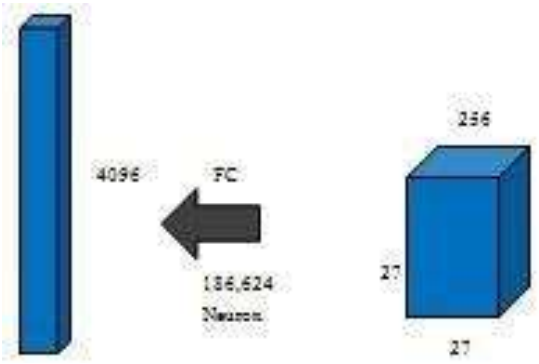


Figure 3.15: Flatten operation

Layer 6: Fully connected with 4096 neuron

Layer 7: Fully connected with 2 neurons

- Classification layer: output into two classes (Benign Vs Malignant)
- Memory: 2

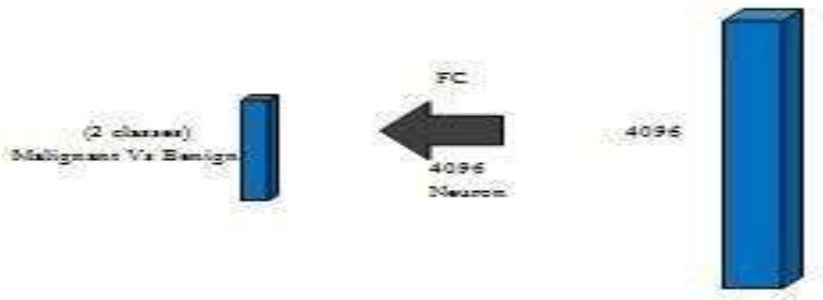


Figure 3.16: Image classification using Softmax

This is the final layer where classification is made by assigning probability values of the class using the features learned from the flatten layer. The modified CNN model is in Figure 3.17.

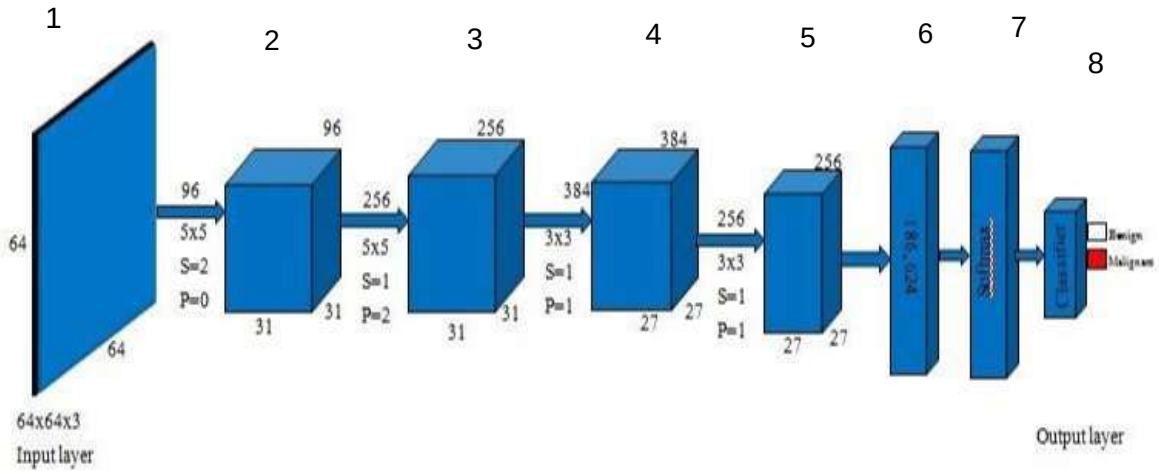


Figure 3.17: Modified CNN architecture used for the model

The modified structure parameters are presented in Table 3.1

Table 3.1: Modified CNN Architecture

	Input image (64 x 64)	
	Conv1 96, 5 x 5 convolution with Stride =2 and zero padding	
3	ReLu	Rectifier linear unit
4	Cross channel normalization	cross channel normalization with 5 channels per element
5	Max pooling layer	3x3 max pooling with stride [2 2] and zero padding
6	Conv2	256, 5 x 5 convolution filter, with Stride=1, and Padding = 2
7	ReLu	Same as previous
8	Cross channel normalization	Same as previous
9	Max pooling layer	Same as previous
10	Conv3	384, 3 x 3 convolution filter with Stride =1, and Padding = 1
11	ReLu	Same as previous
12	Conv4	Same as Conv3
13	ReLu	Same as previous
14	Conv5	256, 3 x 3 convolution filter with Stride=1, and Padding=1
15	ReLu	Same as previous
16	Max pooling layer	Same as previous
17	FC1	Fully connected layer with 186624 neurons
18	ReLu	Same as previous
19	Dropout	Dropout layer of 50%
20	FC2	Fully connected layer with 186624 neurons
21	ReLu	Same as previous
22	Dropout	Dropout layer of 50%
23	FC3	Fully connected layer of 2 neurons
24	Softmax	Softmax
25	Classification layer	Output layer into two classes (Benign Vs Malignant)

3.3.6 Selection of optimization algorithm

The CNN model was trained using the stochastic gradient descent with momentum (SGDM) optimizer due to the ease of convergence and its ability to contribute gradient step of precious iteration to the current iteration of the training phase makes it a good choice for this model. The SGD updates the current weight using the current gradient $\partial L/\partial w$ multiplied by some factor called the learning rate α, which indicated in the equation (3.4) instead of depending only on the current gradient to update the weight; SGDM replaces the current gradient with momentum as shown in equation (3.5). Figure 3.18 shows the back propagation operation which is based on mini batch where a subset of points are sent through the network from the output back to the input while weight values are being updated until a point of convergence. The step carried out during the training is as follows;

I. Initialize the weight parameters;

II. Pass a mini batch subset of points of inputthrough network in a forward propagation to get the loss function

III. Compute the loss in (ˆ,)

V. Update the weights from the end of the network to start (Back propagation) using equation (3.4).

= − (3.4)

VI. Repeat step II till point of convergence. A point where update in weight has no significant change in the network.

Where +1 = (3.5)

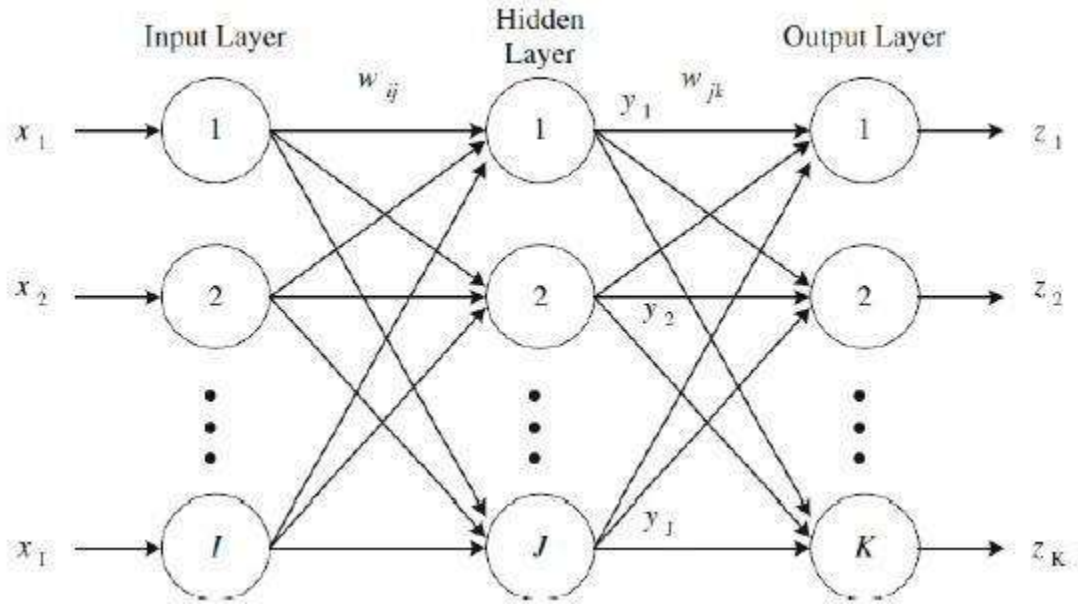


Figure 3.18: Back propagation operation, image adopted from (Salim *et al*., 2016)

Ideally, a grid search approach would have been preferred to fine-tune the model hyperparamter but due to the number of hyperparamter to fine-tune this method was unrealistic as the Scikit-learn prevented it use. Instead, a manually trying different approach like dropout values, input image sizes, learning rate, weight decay were used. To best utilized the base model weights of the ImageNet since we adapt transfer learning approach, during the training stage, all the layers from the base model are initial pause to enable only the custom model to fit the mammogram images, secondly, the learning rate was lower to a very small value (0.10) allowing the base model to slightly change its weight to adapt to the mammogram dataset. To ensure that the model generalizes well to unseen data from the testing set, the training set is further split to form a validation. The

validation set is used to make prediction at the end of each epoch by calculating the loss and accuracy. The training stops when the specified condition is satisfied. Some of the parameters used for this study are given in Table 3.2.

Table 3.2: Training parameters of the CNN system model

S/N	Parameters	Values
1	Maximum Epoch	100
2	Initial Learning rate	0.10
3	DropoutLayer	0.05

3.3.7 Classification layer

A Softmax function was used at the final layer for classification. This work by assigning probability values of the i^{th} class using the features X_L learned from the flatten layer.

$$= \frac{\quad}{\Sigma \quad} \tag{3.6}$$

Where; = *ith* element of the input image

X_L = Learned features from the flatten layer

Σ ( ) = Standard exponential function applied on the learned features to make the value close to zero.
( ) = Normalization term making the output vector sums to 1

3.4 General Design Decisions

MATLAB and Python are chosen due to the experience with the language and open access to good machine learning libraries such as such as Tensorflow, Scipy, Keras, SciKit-Learn, Numpy for array, and Matplotlib for data visualization and torchvision for image dataset and models.

3.4.1 Code Design

The code is designed by being split into functions spread across multiple MATLAB modules, which are all organized into different directories based on the kind of task they are designed to carry out (see Appendix A).

3.4.2 Flow of execution

The code is design in such way that it allows all the images to be stored in a particular folder directory which are arranged based on the kind of task to perform. All the images are stored in Datastore to manage the collections of image data; the imageDatastore command allows MATLAB to reads data directly when needed.

3.4.3 Preparing image to use as input

To classify image with CNN model, each image needs to have the size specified by the network input layer. All these operations are carried out in the data_operation directory, where operations like resize, retrieving image paths and labels, and application of transformations for data augmentation are performed.

3.4.4 Label images in Datastore

The label needed for training is automatically determined by the label folder names by specifying the label source option using the following format:

3.4.5 Dataset Splits

SplitEachLabel function is used to divide the images in datastore already created into training, testing and validation. The validation data is not used to train the network but to evaluate its performance.

3.4.6 Model Training

The pre-trained network is used through the keras Application API. The fully connected layers are dropped by adding include_top = False to the imported model and replaced with a new fully connected network layer, softmax activation function and the output layer is finally added to the end of the sequential model.

3.4.6 Set training algorithm options

Create a variable containing training algorithm options. The goal of training is to minimize the loss function which is the measure of how badly the network performs in the training data. During training, the validation data is use to test the network performance, which are set to stop training when the validation loss does not further improve after the maximum number of epoch is reached.

3.4.7 Use trained network to perform classification

After the network finished training, all the parameters learnt are saved to be re-used for final model evaluation on the test data. The classification is made through the classified function by passing in new images and their ground truth labels. The summary of the flow of execution is as shown in Table 3.3

Table 3.3: Showing the flow of execution in MATLAB (R2019a)

S/N	Command	Functions
1	[trainDigitData,testDigitData] = splitEachLabel(digitData, ... trainingNumFiles,'randomize');	Dataset Splits
2	options = trainingOptions('sgdm','MaxEpochs',10, ... 'ValidationData',testDigitData, ... 'ValidationFrequency',10, ... 'InitialLearnRate', 0.01...)	Set training algorithm
3	convnet = trainNetwork(trainDigitData,layers,options); YTest = classify(convnet,testDigitData); TTest = testDigitData.Labels; % Calculate the accuracy. accuracy = sum(YTest == TTest)/numel(TTest);	Make classification
4	printf('The recognition rate is %1.3f\n',accuracy);	Print results

3.5 Evaluation of model performance

In evaluating the model performance the following metrics are used; classifier accuracy, Sensitivity, Specificity, Precision, Recall and F1 score. The following terminology is used to define the metrics and mathematical formulas used:

TP = True Positives (positive case correctly predicted as positive)

TN= True Negatives (negative case correctly predicted as negative)

FP= False positives (negative case incorrectly predicted as positive)

FN= False negatives (positive case incorrectly predicted as negative)

Classifier accuracy

It is the ratio of number of correct predictions to the total number of input samples

$$\text{Accuracy} = \frac{+}{\ldots} 100 \qquad (3.7)$$

Sensitivity

It is also referred to as True Positive Rate (TPR) which corresponds to the proportion of positive data points that are correctly predicted as positive with respect to all positive data points. It measure the disease presence.

$$\text{Sensitivity} = \frac{\quad}{+} \ 100 \tag{3.8}$$

Specificity

It is referred to as False Positive Rate (TPR) corresponds to the proportion of negative data points that are incorrectly predicted as positive with respect to all negative data points. It measure the disease absence.

$$\text{Specificity} = \frac{\quad}{+} \ 100 \tag{3.9}$$

F1 Score

It is used to measure test accuracy, it tells how precise the classifier is as well as how robust it is in prediction. The greater the F1 Score, the better the model performance. It tries to find the balance between precision and recall.

$$F1 = 2* \frac{\quad}{\quad} \tag{3.10}$$

Precision

It is the number of correct positive results divided by the number of positive results predicted by the classifier. This is number of correct positive prediction.

$$\text{Precision} = \frac{\quad}{+} \ 100 \tag{3.11}$$

Recall

It is the number of correct positive results divided by the number of all samples that should be positive. This is number of positive cases that are correctly predicted.

$$\text{Recall} = \frac{}{} 100 \tag{3.12}$$

CHAPTER FOUR
4.0 RESULTS AND DISCUSSION

4.1 Presentation of Results and Discussion

The chapter presents the simulation results of the research on Breast cancer classification. The results are broadly categorized into two phases; breast cancer classification using the original MIAS dataset and augmented samples in both cases, the model performance were evaluated. The entire MATLAB codes used for the simulation are bulky hence some of the codes are presented in appendix A.

4.2 Results for Dataset acquisition

This section presents result for dataset acquisition which is obtained from the UK based medical research group referred to as Mammographic Image Analysis Society (MIAS) using the web site address: http://peipa.essex.ac.uk. The dataset contains 322 gray scale image samples in Portable Gray Map (PGM) with 1024 x 1024 pixels dimension. A randomly selected samples from the original MIAS dataset is as shown in Figure 4.1

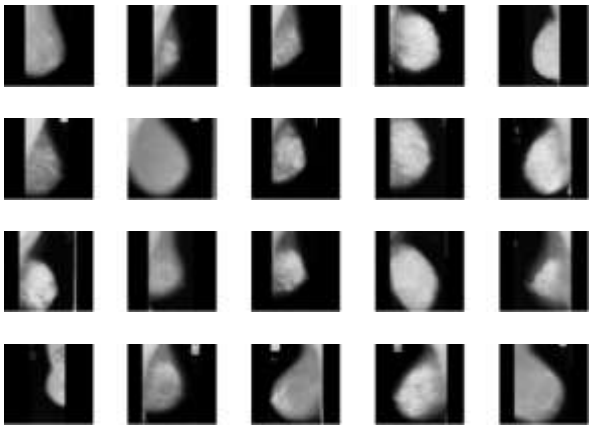
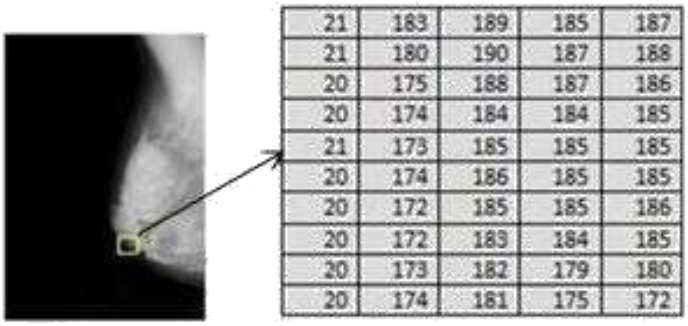


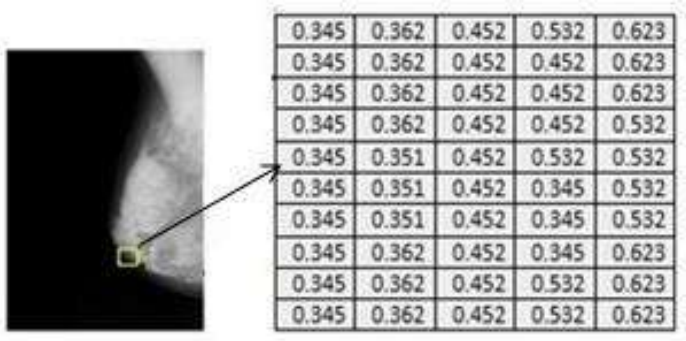
Figure 4.1: Samples from MIAS dataset

The dataset was further processed and normalized as shown in Figures 4.2 and 4.3



21	183	189	185	187
21	180	190	187	188
20	175	188	187	186
20	174	184	184	185
21	173	185	185	185
20	174	186	185	185
20	172	185	185	186
20	172	183	184	185
20	173	182	179	180
20	174	181	175	172

Figure 4.2: Pixel intensities ranging from 0 to 255 before normalization



0.345	0.362	0.452	0.532	0.623
0.345	0.362	0.452	0.452	0.623
0.345	0.362	0.452	0.452	0.623
0.345	0.362	0.452	0.452	0.532
0.345	0.351	0.452	0.532	0.532
0.345	0.351	0.452	0.345	0.532
0.345	0.351	0.452	0.345	0.532
0.345	0.362	0.452	0.345	0.623
0.345	0.362	0.452	0.532	0.623
0.345	0.362	0.452	0.532	0.623

Figure 4.3: Pixel intensities ranging from 0 to 1 after normalization

It was observed that the pixels intensities correspond to integer values ranging from 0 to 255 and the weight parameters in the model are relative small, having such discrepancies can slow down the training speed, data normalization was applied to make the pixels intensities range from 0 to 1. Data augmentation was also applied to the dataset to increase the amount of data sample need for model training to achieved better system accuracy. The dataset was increased from 322 samples to 2,576 samples after application of augmentation techniques, randomly selected samples are shown in Figure 4.4. The result obtained is the deliverable for objective one.

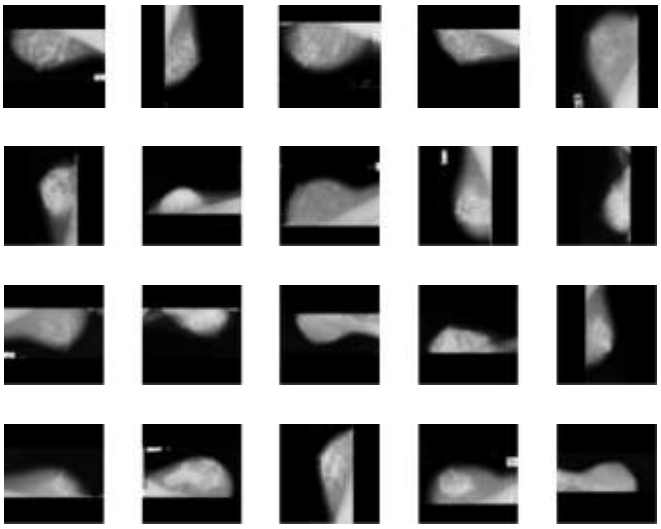


Figure 4.4: Augmented data samples

4.3 Results for Breast cancer classification using Convolutional Neutral Network

The results are presented in two different phases;

4.3.1 Phase I: Breast cancer classification using original MIAS dataset

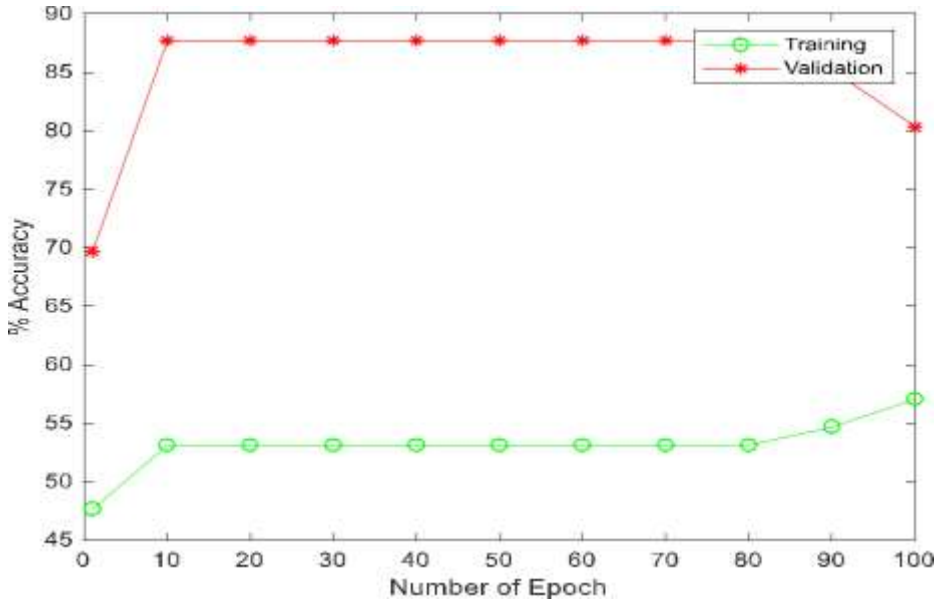


Figure 4.5: Plot showing Training Accuracy and Validation

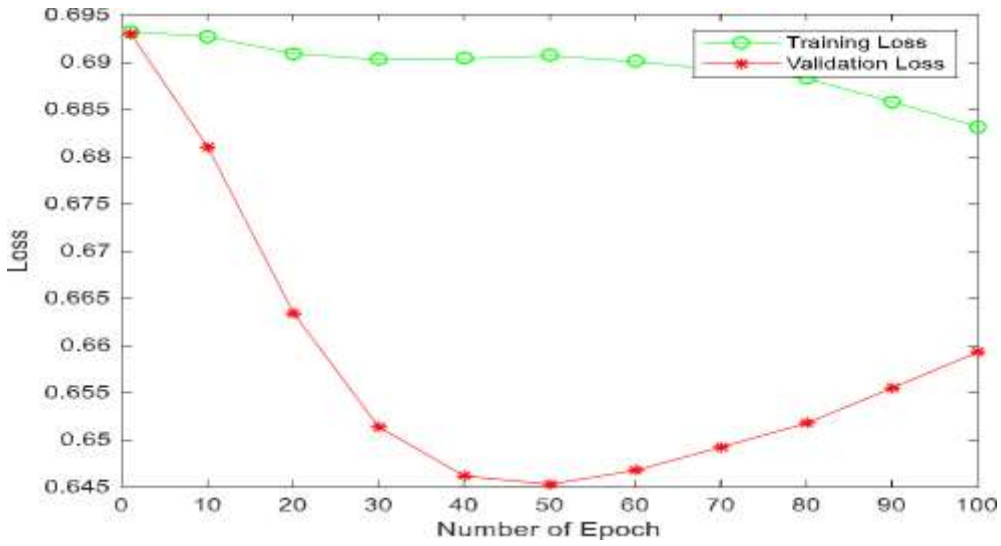


Figure 4.6: Training and Validation Loss function

The obtained results using single CPU for simulation is thus presented in Figures 4.5 and 4.6.

In Figure 4.5 at the beginning of training phase there is a slight increase in the validation to a point where it become constant, at this point the model stop learning and gradually start to overfit. The training loss curves in Figure 4.6 indicate a slight decrease as the training progresses while the validation loss curve shows a steady decrease to a point and begins increasing again at a point called inflection point which is where the model stops learning and after that point shows the dynamics of overfitting. The gap between the validation loss and the training loss is very large indicating an overfitting. This occurs when there is either a limited dataset or model being trained for too long.

MATLAB plots showing training progress with values are reported in Figure 4.7.

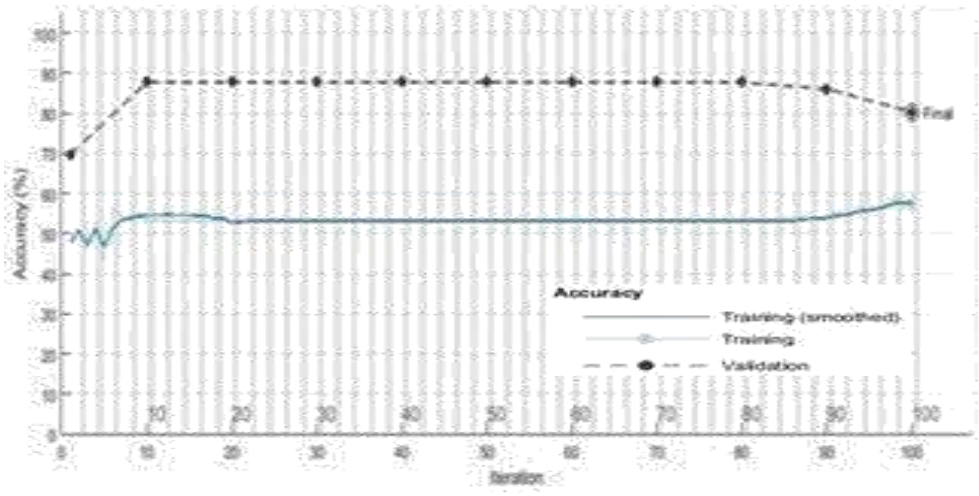


Figure 4.7: MATLAB plots showing training progress, validation and percentage accuracy of the model with original samples from the dataset.

In this case, the average classification accuracy of the model is about of 87.70%.

4.3.2 Phase II: Breast cancer classification using augmented samples dataset

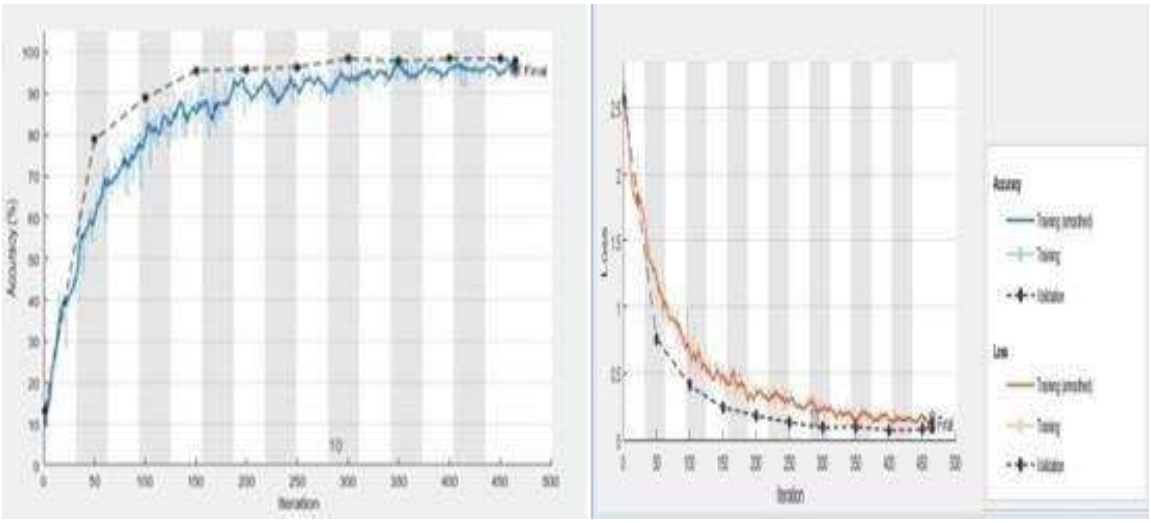


Figure 4.8: MATLAB plots showing training progress, validation loss and percentage accuracy of the system model

The second case involves augmentation of the original dataset using the techniques previously discussed. With augmentation, obtained results show some level of improvement and this is thus presented as in Figure 4.8. The learning curves in this case

shows a constant decrease to a point of stability and the gap between the validation loss and the training loss is relative small given an indication for a good fit of the model. At this point, continued training of the model might result to an overfit. The learning curves in Figure 4.8 indicate a significant increase in classification accuracy of about 95.80%.

The Receiver Operative Characteristic (ROC) curve describes how good a model is at predicting accurately positive class. This shows the relationship between the True Positive Rate (TPR) known as sensitivity (recall). The TPR tells what proportions of the breast cancer samples are correctly classified and False Positive Rate (FPR) or (1- specificity) indicates breast cancer samples that are incorrectly classified. The relationship between these metrics is presented in Table 4.1 and Figure 4.9.

Table 4.1: Performance metrics parameters

Methods	Database Type	Precision (%)	Recall (%)	F1score (%)
Model without augmentation	MIAS	70.00	68.31	69.14
Proposed model with augmentation	MIAS	92.30	95.00	93.63

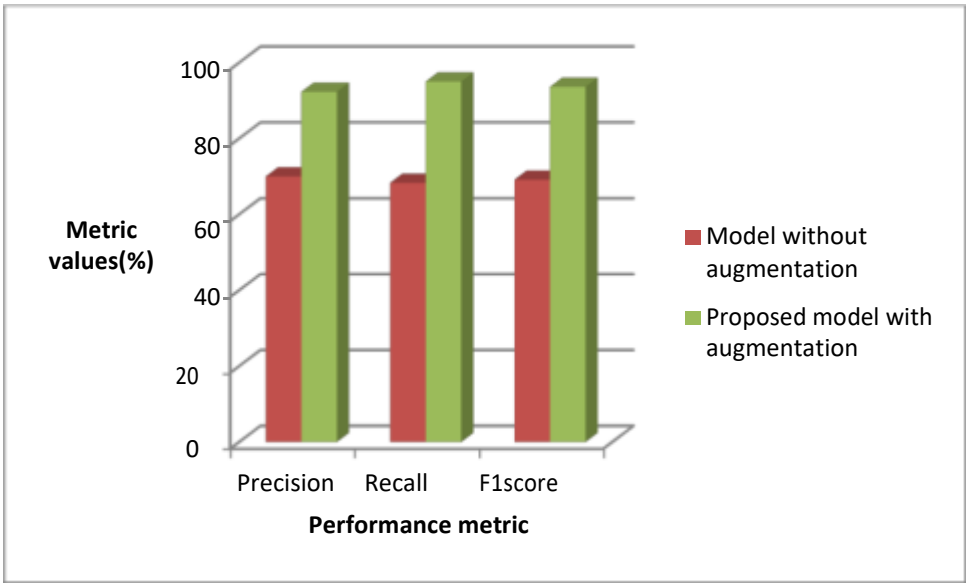


Figure 4.9: Performance metrics showing precision, Recall and F1score of the model

In Figure 4.8, the precision, recall and F1score of model without augmentation are 70%, 68.31% and 69.14% respectively. With augmentation of the dataset the value become 92.3%, 95% and 93.63% for precision, recall and F1score. The high rate percentage increment shows the effect of augmentation on the dataset.

The ROC curve of the model summaries all the confusion matrices that each of the threshold produced. The threshold was set at the optimum point of 0.5, as one move along the ROC curve upward it shows that there is more proportion of accurately predicted samples than incorrectly predicted samples. At a high point (0.958) of correctly predicted samples on the TPR axis indicate a lower point of 0.15 incorrectly predicted samples on the FPR axis as shown in Figure 4.8. The straight line indicates a direct proportionality relationship between true positive rate and false positive rate, at any point on the line mean that the proportion of correctly classified breast cancer samples (malignant) is the same as the proportion of incorrectly classified samples that are not breast cancer (benign).

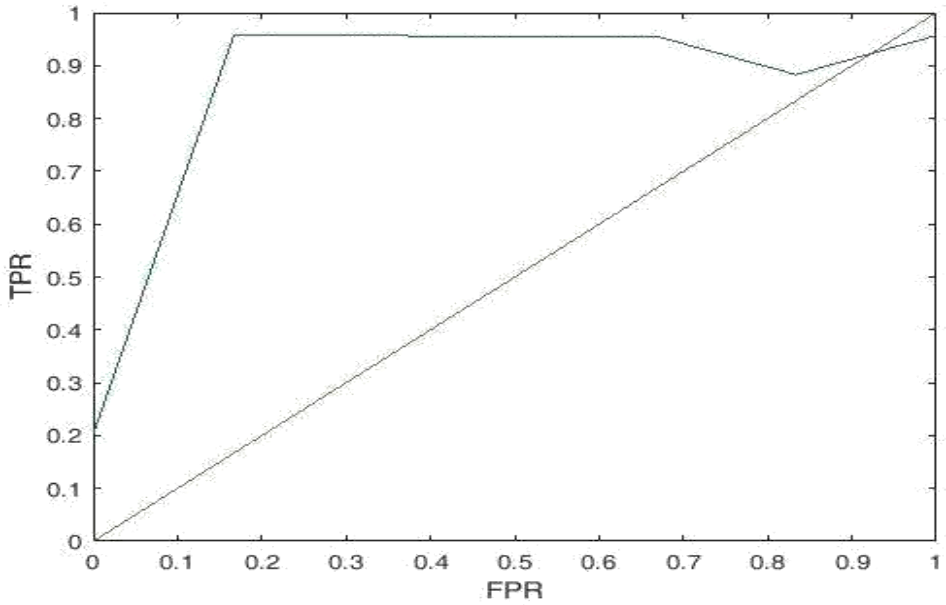


Figure 4.10: The ROC curve of the system model

This is the deliverable for both objectives two and three.

4.4 Comparison with the work of Peng Shi *et al*. (2019)

In comparison with the results in the work of (Peng Shi *et al*., 2019) using lightweight Convolution Neural Network for breast cancer classification with MIAS dataset the model accuracy on the test data is 83.6% while the proposed model was able to achieved 95.8% overall classifier accuracy. The result indicates an improvement of about 12.2%. In medical image analysis 12.2% improvement in classification accuracy is a huge contribution toward the ravaging breast cancer disease as this can help the radiologist to make accurate prediction of the presence or absence of the disease. To better illustrate the results, Table 4.1 and Figure 4.9 show the comparison and improvement ratio in results between the study and proposed model. The studies are compared based on system accuracy alone because the study do not include the results of other evaluation metrics such as sensitivity, specificity and F1- score. However, the proposed model out performs in term of model accuracy.

Table 4.2: Performance metrics of model parameters

Methods	Aug	Database Type	Sensitivity (%)	Specificity (%)	Accuracy (%)
Peng Shi.,(2019)	Y	MIAS	-	-	83.60
Augmentati	N	MIAS	68.31	69.90	87.70
on Proposed	Y	MIAS	95.00	80.00	95.80
model					

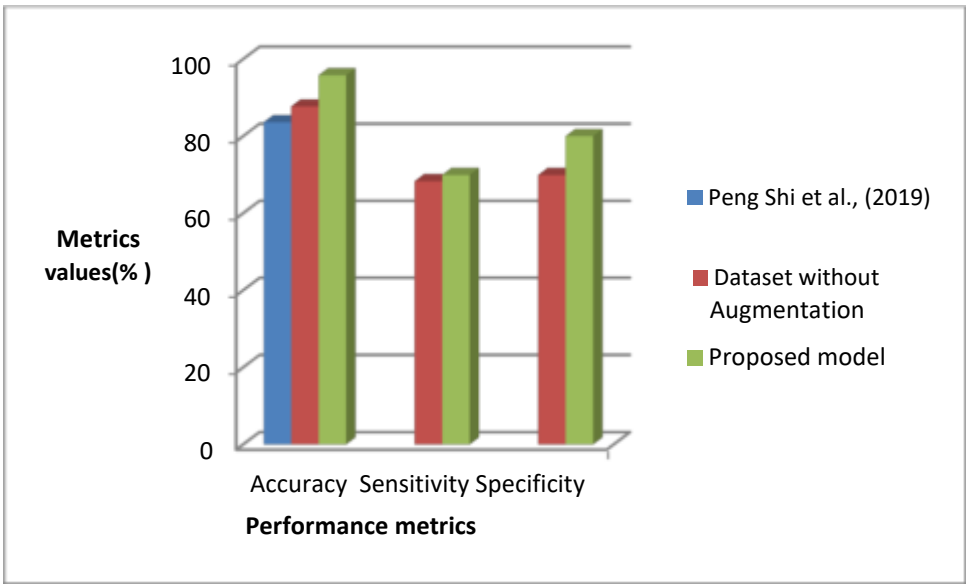


Figure 4.11: Plot showing the comparison between studies and proposed model

In Figure 4.9, the metric parameters of the models is plotted and presented as follow; 83.6%, 87.7% and 95.80% gives system accuracy of (Peng Shi et al., 2019), system model without augmentation and proposed model respectively. This shows that the proposed model outperforms the existing model. This is the deliverable for objective four.

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Breast cancer disease is still ravaging there is need for more research in this areas to help save life. This research proposed Breast cancer classification using modified Convolutional Neutral Network Technique and it was able to achieve 95.80%, 95.00%, 80.00%, 92.30%, and 93.63% for model accuracy, sensitivity, specificity, precision and F1score respectively. It was designed using eight (8) layers with the parameters presented in Table 3.1. It was implemented using MATLAB R2019a platform and deep learning toolboxes. The dataset was resized to a dimension of 64 by 64 pixels to avoid having inconsistent input sizes and to reduce computation complexity. The pixel intensities of the images in the dataset have a large integers value ranging from 0 to 255, using such large input values can slow down the training process considering that weight of the neural network model is small this could lead to lower accuracies because of the discrepancy of the image intensities. Therefore, to solve this problem we applied normalization technique to reduce the pixel intensities to a small value ranging from 0 to 1. It was also observed that the dataset come with class imbalance between samples (normal, benign and malignant) which training the model with such can affect the classification accuracy of the model. To overcome this problem, we split the dataset using the ratio of 30%/70% for training set and the rest for validation and testing set. This approach also helps to avoid over fitting of the training model.

The implementation of the work requires the use of different convolution filter sizes of (5 x 5) and (3 x 3) for the first four layers respectively. The program was simulated on two cases: dataset with augmentation and without augmentation. In both cases, a learning rate of 0.10 and maximum epoch of 100 were used. The SGDM training function (optimizer)

was used with a PC specification of 8GB RAM, Core i5, 2. 5GHz.The obtained results showed that the proposed model using modified Convolution Neutral Network Technique outperformed with better classification accuracy than the existing literature using deep learning techniques for breast cancer analysis. This is made possible by leveraging on augmentation techniques and modification of hyper parameters of the deep learning model.

5.2 Recommendations

The work can be done on more dataset and run on different algorithms and compose to see which is better. Also, more work can be explored on the combination of traditional augmentation techniques and Generative Adversarial Networks (GAN) due to the great potential benefit on data augmentation.

5.3 Contribution to knowledge

i. Breast Cancer classification system using modified Convolution Neural Network Technique is the key contribution of the study.

ii. A comparative analysis of the system using Accuracy, Sensitivity, Specificity, Precision, and F1score for the purpose of future reference.

REFERENCES

Al-antari, M. A. Al-masni, and Y. M. Kadah, (2017) “Hybrid model of computer-aided breast cancer diagnosis from digital mammograms,” Journal of Scientific and Engineering Research (JSAER), 4(2): 114 - 126.

Bengio Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., &, Y., (2014).Generative adversarial nets”, in Advances in neural information processing systems, 10(4): 2672–2680.

Bray, Jacques, F., Isabelle Soerjomataram, Rebecca L Siegel, Torre, & Ahmedin, J.,(2018) Global cancer statistics: Globocan estimates of incidence & mortality worldwide for 36 cancers in 185 countries. CA: A Cancer journal for clinicians, 68(6): 394–424.

Burcu, B., Ilkim, E., & Elif, K. (2018). Classification of Mammogram images by machine Learn Techniques, 3rd International Conference on Computer science and Engineering, IEEE

Caulfield, B. (2009), `What are the Difference between CPU vs a GPU, NVIDIA Blog', https://blogs.nvidia.com/blog/2009/12/16/whats-the-difference-between-a-cpu-and-a-gpu/. [Online] Accessed: 2020-07-29.

Cong J, Wei B, He Y, Yin Y, & Zheng, Y.,(2017) A selective ensemble classification method combining mammography images with ultrasound images for breast cancer diagnosis. Computer Mathematical Methods

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li and Li Fei-Fei (2012), ImageNet:A large-Scale hierarchical image database, in `Institute of Electrical & Electronics Engineers (IEEE)', pp. 248{255.

Dina Abdulhafiz, Clifford Yang, Roda Ammar & Sheida Nabavi, (2019).Deep Convolution Neural Networks for mammagraphy: advances, challenges and applications. 7th IEEE International Conference on Computational Advances in Bio and Medical Science (ICCABS) Orlando, FL, USA. pp. 79 -103

Diaz, O., Marti, R., Llado, X. and Agarwal, R. (2018), Mass detection in mammograms Using pre-trained deep learning models, in E. A. Krupinski, 14th International Workshop on Breast Imaging (IWBI 2018)', Vol. 10718, SPIE, p. 12.URL: https: // www. spiedigitallibrary. org/ conference-proceedings-of-spie/ 10 718/ 2317681/ Mass-detection-in-mammograms-using-pre-trained-deep-learning-models/10 .1117

Emperumal & Duraisamy (2017). Computer-aided mammogram diagnosis system using Deep learning convolutional fully complex-valued relaxation neural network classifier,” Inst. Eng. Technol., vol. 11, pp. 656–662.

Elter, M. and Horsch, A. (2019), CADx of mammographic masses & clustered structure Microcalcications A review Medical Physics 36(6), 2052{2068.

Falconi, L. G., Perez, M. & Aguilar, W. G. (2019), Transfer Learning in Breast Cancer Mammogram Abnormalities Classication with Mobilenet and Nasnet, International Conference on Systems, Signals, and Image Processing 2019-June, 109114.

Gbenga A., Hamidinekoo, E. Denton, A. Rampun, K. Honnor, & R. Zwiggelaar,(2017). Deep learning in mammography and breasthistology, an overview and future trends, Medical Image Analysis, vol. 47, pp. 45–67.

Geron, A. (2019), Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 2nd O'Reilly Media.

Ghongade, R. D., & Wakde, (2017). "Detection and Classification of Breast Cancer from Algorithm," IEEE. pp.3 – 6

Hepsag, P. U., Ozel, S. A. and Yazici, A. (2017), Using deep learning for the mammograph classification, in `2nd International Conference on Computer Scienceand Engineering, UBMK 2017', Institute of Electrical and ElectronicsEngineers Inc., 15(6):418-423.

Huang, T. J., (2019). Imitating the brain with neurocomputer a new way towards Deep Artificial general intelligence," Int. J. Autom. Comput., vol. 14(5), pp. 520–531.

Jedy-Agba, F., Curado, & Ogunbiyi, (2018). Cancer Incidence in Nigeria report from population based cancer registries, West African J. Radiol., vol. 36, pp. 271–279.

Jiang, F.,(2017). Breast Mass Lesion Classification in Mammograms by Transfer Learning, ICBCB, vol. 23, pp. 8–11.

Jiao X. Z., & Gao, (2016). A deep feature based framework for breast masses density classification, J. Neurocomputing, vol. 197, pp. 221–231

Karasfi, (2018) "Computer aided detection/diagnosis of breast cancer in mammography And ultrasound: a review," Clinical Imaging, vol. 37, no. 3,pp. 420–426.

Krizhevsky, A., Sutskever, I., & Hinton, G. E., (2012). Imagenet classification with Deep convolutional neural networks//Advances in neural information processing systems. vol 2(5)pp. 1097-1105.

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M. and Rubin, D. L.(2017). A curated mammography data set for use in Computer-Aided Detection and diagnosis research, Scientic Data 4.

Lian Zou, (2019). A technical review of convolutional neutral network-based Medical Mammographic breast cancer diagnosis, Computational and Mathematical Methods in Medicine, vol,5 ID 6509357, page 16.

Ling, L. D., Kang, S. J., & Zen, (2015). Deep learning for acoustic modeling in speech Parametric speech generation: a systematic review of existing techniques and future trends," IEEE Signal Process, vol. 32 (3), pp. 35–52.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, 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

Martin & Laura J. (2019).WebMD: Breast Biopsy for Breast Cancer Classification Diagnosis https://www.webmd.com/breast-cancer/breast-biopsy.[Online] Accessed: 2020-06-22.

Maleika Heenaye, Mamode Khan, Nazmeen Boodoo- Jahangeer, Wasiimah Dullull, Shaista Nathire, Xiaohong Gao, G. R. Sinha, Kapil, Kumar Nagwanshi, (2021). Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network.

PLoS ONE 16(8): e0256500. https://doi.org/10.1371/journal.pone.0256500

Moura, M. A., & Guevara L´opez, (2013). An evaluation of image descriptors combined With clinical data for breast cancer diagnosis," International Journal of Computer Assisted Radiology and Surgery, vol. 8, no. 4, pp. 561–574.

Osareh, A. & Shadgar, B. (2019). Machine learning techniques to diagnose breast cancer, in 5th International Symposium on Health Informatics and Bioinformatics, HIBIT, pp. 114,120.

Peng Shi, Chongshu Wu, Jing Zhong, & Hui Wang, (2019). Deep learning from Small Dataset for BI-RADS Density Classification of Mammography Images, 10th International Conference on Information Technology in Medicine and Education (ITME), 978-1-7281-3918-0/19, DOI 10.1109/ITME.2019.00034

Punitha, S., Amuthan, A. & Joseph, K. S. (2018). Benign and malignant breast cancer segmentation using optimized region growing technique', The Future Computing and Informatics Journal 3(2), 348{358.

Ragab, D. A., Sharkas, M., Marshall, & Ren, (2019). Breast cancer detection using deep Convolutional Neural Networks and Support Vector Machines," pp. 1–23.

Rebecca, L., Siegel, A., & Isabelle, S. (2021). Cancer Statistics, in Cancer Journal for Clinician.https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.2 1 6 54.

Saeed, M. D., Adel, M., Anwar, A., Farag, O., Kotb, A., Adel, M., Tawfik, A., Mostafa, H., Darweesh, M. S., Adel, M., Anwar, A., Farag, O., & Kotb, A. (2021). Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images Early breast cancer diagnostics based on hierarchical machine learning classification for mammographyimages.*CogentEngineering*,*8*(1). https://doi.org/10.1080/23311916.2021.1968324

Saeed M. Saeed Darweesh, Mostafa Adel, Ahmed Anwar, Omar Farag, Ahmed Kotb, Adel, A., Ayman Tawfik & Hassan Mostafa (2021), Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images" Cogent Engineering vol 8, 1968324 https://doi.org/10.1080/23311916.2021.1968324

Saira Charan, Muhammed, Jaleed & Khan,(2018). "Breast Cancer Detection in Mammograms using Convolution Neural Network," IEEE Int. Conf. Math. Eng. Technol., vol.1, pp.2–6.

Salim T. Yousif, Suhail I. Khattab, Raghdan Zuhair Al-saffar, (2016). Prediction of soil Compaction parameter using Artificial Neural Network, https://www.researchgate.net/publication/303406340

Saravanan, S. K., (2019). Review on classification based on Artificial Neural Networks, Int. J. Ambient Syst. Appl., vol. 2 (4), pp. 11–18.

Sarosa, S. J. A., Utaminingrum, F. and Bachtiar, F. A., (2018), Mammogram Breast Cancer Classification Using Gray-Level Co-Occurrence Matrix and Support Vector Machine, in 3rd International Conference on Sustainable Information Engineering and Technology, SIET 2018 - Proceedings', Institute of Electrical and Electronics Engineers Inc., pp. 54{59.

Sharkas, & Al-Sharkawy, (2011). Detection of microcalcifications in mammograms using support vector machine, UKSin 5th Eur. Symp. Comput. Model. Simul.vol.23, pp.598

Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R. B. & Sieh, W., (2020). Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography, Scientic Reports 9(1). URL: http: // arxiv. org/ abs/ 170 8. 09427http: // dx. doi. org/ 10 .10 38/ s41598-019-48995-4

Siegel, M. J., & Selvathi, D., (2020). Cancer statistics, CA Cancer J. Clin, vol.66, pp. 7–30.

Shin, H. R., Roth, M., & Gao, (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging. Vol. 35(5), 1285-1298.

Simon, H., Yaecob, G., & Gebrekirstos, H. (2020). Breast Cancer Detection using Convolutional Neutral Netorks.AI4AH, ICLR. vol. 2.

Susan, Selvathi, D., & Aarthy, A. (2017). Mammography: Benefits, Risks, What You Know, Breast cancer community.

Suresh, R., Rao, A. N., & Reddy, B. E. (2019). *Detection and classification of normal and abnormal patterns in mammograms using deep neural network*. *October 2018*, 1–12. https://doi.org/10.1002/cpe.5293

Szegedy, C. W., Liu, Y. Jia, P., Sermanet, S., Reed, D., Anguelov, D., & Erhan, V. R., (2015) Going deeper with convolutions, in Conference on Computer Vision and PatternRecognition (CVPR), IEEE, vol. (2), pp. 1–9.

Tang J, Member S, Rangayyan RM, Xu J., & El Naqa, I. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: recent advances ". IEEE Transactionson Information Technology in Biomedicine 13:236-251.

Tang E. N., Rangayyan, J., (2019). Computer aided detection and diagnosis of breast cancer with mammography: recent advances," IEEE Trans. Inf. Technol. Biomed., vol. 13, pp.251.

Teshale Mulatu D., Tilayie F. G., & Lemma N., (2019). Breast cancer screening in practice and its associated factors among women in kersa district, eastern ethiopia. The Pan African Medical Journal, 33.

Wang, N. Yassin, S. Omran, E. M. F. El Houby, & H. Allam, (2020). Machine Learning techniques for breast cancer computer aided diagnosis using different image modalities: systematic review," Computer Methods and Programs in Biomedicine, vol. 156, 25–45.

Vishrutha, V. and Ravishankar, M. (2014), Early detection and classification of breast cancer, in Advances in Intelligent Systems and Computing', Vol. 327, Springer Verlag, pp. 413{419. URL: https: // link. springer. com/ chapter/ 10 .10 07/ 978-3-319-11933-5{ }45

WHO, (2018). Global Cancer Statistics. Retrieved from International Agency for Research on Cancer website: http://www.iarc.fr

Wu, A. Jalalian, S. B. T. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, & B. Karasfi, (2018).Computeraided detection/diagnosis of breast cancer in mammography and ultrasound: a review," Clinical Imaging, vol. 37, no.3,pp.420

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R., (2019). A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction, Radiology 292(1), 60{66.URL: http: // pubs.org/ doi/10 .1148/20 19182716

Yu, X., Pang, W., Xu, Q., & Liang, M., (2020). Mammographic image Classification with deep fusion learning. Scientific Reports, 10 (1), 14361. https://doi. org/10.1038/s41598-020-71

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. URL: http: // www. mdpi. com/ 2411-9660 /2/ 2/ 13.

Zeiler, M. D., Ranzato, R. Monga, M. Mao, K. Yang, Q. V. Le, & G. E. Hinton, (2013). "On rectified linear units for speech processing," in International Conference on Acoustics,Speech and Signal Processing. IEEE, pp. 3517–3521, https://doi.org/10.1109/ICASSP.2013.6638312

APPENDIX A (Some Breast Cancer Classification Simulation Codes)

```
% data
augmentation clear
data_dir = 'C:\Users\Law\2019
Projects\Cancer_CNN\Prog_Code\canc_code\all-mias';
data_dir_new = 'C:\Users\Law\2019
Projects\Cancer_CNN\Prog_Code\canc_code\all
mias_aug'; n_im = 322;
for k = 1:n_im
ims = dir(fullfile(data_dir,'*.pgm'));
len_vid = length(ims);
st = 1;
stp = 8;
for j = 1:len_vid
%
the_file = strcat(data_dir,'\',ims(j).name);
the_im = imread(the_file);
im_nw = imresize(the_im, [64 64]);
% one (original im)
im_rf = fliplr(im_nw); % two (reflected im)
%imshow(im_nw)
im_rot90_org = imrotate(im_nw,90); % 3 original is 1 to 322,
im_rot90_rf = imrotate(im_rf,90); % 4 original is 1 to 322,
im_rot180_org = imrotate(im_nw,180); % 5 original is 1 to 322,
im_rot180_rf = imrotate(im_rf,180); % 6 original is 1 to 322,
im_rot270_org = imrotate(im_nw,270); % 7 original is 1 to 322,
im_rot270_rf = imrotate(im_rf,270); % 8 original is 1 to 322,
% 323 upward is for the augmented version
% data saving>>
xlswrite(strcat(data_dir_new,'\',num2str(st),'.csv'),im_nw) % 2576
xlswrite(strcat(data_dir_new,'\',num2str(st+1),'.csv'),im_rf) % 2576
xlswrite(strcat(data_dir_new,'\',num2str(st+2),'.csv'),im_rot90_org) %
2576
xlswrite(strcat(data_dir_new,'\',num2str(st+3),'.csv'),im_rot90_rf) %
2576
xlswrite(strcat(data_dir_new,'\',num2str(st+4),'.csv'),im_rot180_org) %
2576
xlswrite(strcat(data_dir_new,'\',num2str(st+5),'.csv'),im_rot180_rf) %
2576
xlswrite(strcat(data_dir_new,'\',num2str(st+6),'.csv'),im_rot270_org) %
2576
xlswrite(strcat(data_dir_new,'\',num2str(st+7),'.csv'),im_rot270_rf) %
2576
st = st+stp;
end
% data size of
2576; clear
rng('default') % For reproducibility
%
exts = '.png';
digitDatasetPath = fullfile('C:\Program
Files\Polyspace\R2019a\bin\Data_sample3');
digitData = imageDatastore(digitDatasetPath,...

'IncludeSubfolders',true,'LabelSource','foldernames','FileExtensions',ext s);
```

```
figure;
perm = randperm(322,20);
for i = 1:20
    subplot(4,5,i);
    imshow(digitData.Files{perm(i)});
end
CountLabel = digitData.countEachLabel;
img = readimage(digitData,1);
size(img)
trainingNumFiles = 100;
[trainDigitData,testDigitData] = splitEachLabel(digitData, ...
                trainingNumFiles,'randomize');
conv1 = convolution2dLayer(5,96,'Stride',2); % chang strid from 4 to
2 conv2 = convolution2dLayer(5,256,'Stride',1,'Padding',2);
conv3 = convolution2dLayer(3,384,'Stride',1,'Padding',1);% 384
3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
conv4   =   convolution2dLayer(3,384,'Stride',1,'Padding',1);
conv5   =   convolution2dLayer(3,256,'Stride',1,'Padding',1);
crsch = crossChannelNormalizationLayer(5);
% (3,'Stride',2); gav 80.33
mypool = maxPooling2dLayer(3,'Stride',2);
layers = [imageInputLayer([64 64 1]) % L1
          conv1   % filtsiz, numofFilt  L2
          %batchNormalizationLayer
          reluLayer  %L3
          crsch  % 4
          mypool  % 5
          % %
          conv2   % 6
          %batchNormalizationLayer
          reluLayer  % 7
          crsch % 8
          mypool % 9, pool 2
          %
          conv3  % 10
          %batchNormalizationLayer
          reluLayer % 11
          %mylayer
          %
          conv4  % 12
          %batchNormalizationLayer
          reluLayer  % 13
%           mylayer
%           % chk
          conv5 % 14
%           batchNormalizationLayer
          reluLayer % 15
          mypool  % 16, this is pool 5
%
          fullyConnectedLayer(4096)  % 17
          reluLayer % 18
          dropoutLayer % 19
          fullyConnectedLayer(4096) % 20
          reluLayer % 21
          dropoutLayer % 22
          fullyConnectedLayer(2) % 23
          softmaxLayer % 24
```

```
          classificationLayer()]; % 25
options = trainingOptions('sgdm','MaxEpochs',10, ...
                          'ValidationData',testDigitData, ...
                          'ValidationFrequency',10, ...
                          'InitialLearnRate',0.01,...
                          'Plots','training-progress');
convnet = trainNetwork(trainDigitData,layers,options);
YTest = classify(convnet,testDigitData); TTest =
testDigitData.Labels;
% Calculate the accuracy.
accuracy = sum(YTest == TTest)/numel(TTest);
accuracy = accuracy*100;
fprintf('The recognition rate is %1.3f\n',accuracy)
precision = (YTest == TTest)/(TTest);
fprintf('The recognition rate is %1.3f\n',precision)
%Tabulate the results using a confusion matrix.
confMat = confusionchart(TTest, YTest );

% % Convert confusion matrix into percentage form

 clear
rng('default') % For reproducibility
%
exts = '.png';
digitDatasetPath =
fullfile('C:\Users\Dlaw\Documents\Projects\data_sampl3');
digitData=imageDatastore(digitDatasetPath,...

'IncludeSubfolders',true,'LabelSource','foldernames','FileExtensions',ext s);

figure;
perm = randperm(2576,20);
for i = 1:20
    subplot(4,5,i);
    imshow(digitData.Files{perm(i)});
end
CountLabel = digitData.countEachLabel;
img = readimage(digitData,1);
size(img)
trainingNumFiles = 100;
[trainDigitData,testDigitData] = splitEachLabel(digitData, ...
                trainingNumFiles,'randomize');
%
conv1 = convolution2dLayer(5,96,'Stride',2); % chang strid from 4 to
2 conv2 = convolution2dLayer(5,256,'Stride',1,'Padding',2);
conv3 = convolution2dLayer(3,384,'Stride',1,'Padding',1);% 384
3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
conv4   =   convolution2dLayer(3,384,'Stride',1,'Padding',1);
conv5 = convolution2dLayer(3,256,'Stride',1,'Padding',1);
%
crsch = crossChannelNormalizationLayer(5); %
(3,'Stride',2); gav 80.33
mypool = maxPooling2dLayer(3,'Stride',2);
layers = [imageInputLayer([64 64 1]) % L1
          conv1   % filtsiz, numofFilt  L2
          %batchNormalizationLayer
          reluLayer  %L3
```

```
            crsch  % 4
            mypool  % 5
            % %
            conv2   % 6
            %batchNormalizationLayer
            reluLayer  % 7
            crsch % 8
            mypool % 9, pool 2
            %
            conv3  % 10
            %batchNormalizationLayer
            reluLayer % 11
            %mylayer
            %
            conv4  % 12
            %batchNormalizationLayer
            reluLayer  % 13
%             mylayer
%             % chk
            conv5 % 14
%             batchNormalizationLayer
            reluLayer % 15
            mypool  % 16, this is pool 5
%
            fullyConnectedLayer(4096)  % 17
            reluLayer % 18
            dropoutLayer % 19
            fullyConnectedLayer(4096) % 20
            reluLayer % 21
            dropoutLayer % 22
            fullyConnectedLayer(2) % 23
            softmaxLayer % 24
            classificationLayer()]; % 25
options = trainingOptions('sgdm','MaxEpochs',10, ...
                          'ValidationData',testDigitData, ...
                          'ValidationFrequency',10, ...
                          'InitialLearnRate', 0.01...
                          );
convnet = trainNetwork(trainDigitData,layers,options);
YTest = classify(convnet,testDigitData);
TTest = testDigitData.Labels;
% Calculate the accuracy.
accuracy = sum(YTest == TTest)/numel(TTest);
accuracy = accuracy*100;
printf('The recognition rate is %1.3f\n',accuracy);
%Tabulate the results using a confusion matrix.
confMat = confusionchart(TTest, YTest );
% % Convert confusion matrix into percentage form
```

APPENDIX B (Publication)

Emmanuel, L. O., Michael, D., Achonu, A., & Saliyu, A. (2020). Breast Cancer; Tumor Detection in Mammogram Images using Modified AlexNet Deep Convolution Neural Network, International Conference in Mathematics, Computer Engineering and Computer Science(ICMCECS)978-1-7281-3126-9/20/$31.00©2020IEEE 10.1109/ICMCECS47690.2020.240870