

Deep Learning-Based Recommendation System for Breast Cancer Diagnosis

Soona Ahmed Abdullah

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
September 2020
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Hadi Işık Aybay
Chair, Department of Computer
Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Computer Engineering.

Assoc. Prof. Dr. Duygu Çelik Ertuğrul
Supervisor

Examining Committee

1. Assoc. Prof. Dr. Duygu Çelik Ertuğrul

2. Assoc. Prof. Dr. Önsen Toygar

3. Asst. Prof. Dr. Mehtap Köse Ulukök

ABSTRACT

Breast cancer is considered one of the deadliest cancers among females. Despite the advanced achievement in the field of medical imaging analysis, a few early research works proposed semi-automatic machine learning algorithms that were complex and computationally expensive.

Recently, developing a system based on deep learning concepts was the center of the attention to analyze mammograms, which is the golden standard imaging technique to diagnose the existence of an abnormality in breast tissues. Systems based on deep learning are still considered to be limited due to insufficient datasets.

In this study, numerous experiments were made on small size ROI samples of mammogram images in the CBIS-DDSM dataset to find the best configuration of a pre-trained Convolutional Neural Network with ImageNet dataset.

The main training concept is about extracting standard features automatically by striding filters over the input matrix (mammogram). Thus, a larger number of inputs lead to recognize a useful pattern to classify the abnormality.

The pre-trained models along with data augmentation algorithms are applied to minimize the dataset challenge in the breast cancer diagnosing field. On the other hand, image preprocessing techniques helped to enhance the image inputs.

The study addresses two pre-trained models VGG16 and ResNet50, both were fine-tuned in different depths and hyper-parameters. The experimental results obtained with good performance used a VGG16 pre-trained model after fine-tuning the last fully connected layer.

The proposed VGG16 model outperformed other deep learning algorithms with an F1 score and AUC of 82% when classifying the abnormality type into calcification/ mass. The model has also a high score with a mean AUC equal to 0.80 when classifying the mammograms into four classes: benign calcification, malignant calcification, benign mass, and malignant mass.

The final application in this study tries to assist radiologists to accomplish more precise decision on the abnormality pathology of breast lesions present in full mammogram images. The application also helps in reducing diagnosing time hence increasing the early detection time.

Keywords: CNN, Transfer Learning, Breast Abnormality, Classification, Mammograms, Calcification, Mass.

ÖZ

Göğüs kanseri, kadınlar arasında ölümcül kanserlerden biri olarak kabul edilir. Tıbbi görüntüleme analizi alanındaki gelişmiş başarıya rağmen, birkaç araştırma çalışması, karmaşık ve hesaplama açısından pahalı olan, yarı otomatik makine öğrenim algoritmaları önermişlerdir. Son zamanlarda, derin öğrenme kavramlarına dayalı bir sistem geliştirmek ilgi odağı olmuştur. Bununla birlikte, bu sistemler, çoğunlukla mamografi görüntülerini analiz eder ve meme anormalliğinin varlığını teşhis etmek için altın standart, mevcut veri setlerinin yetersiz olması nedeniyle hala sınırlı olduğu düşünülmektedir.

Bu çalışmada, ImageNet veri kümesiyle önceden eğitilmiş evrişimli sinir ağına dayalı bir transfer öğrenme modelinin en iyi modelini bulmak için, CBIS-DDSM veri kümesindeki mamogram görüntülerinin küçük boyutlu ROI örnekleri üzerinde çok sayıda deney yapılmıştır.

Eğitim veri seti, giriş matrisi (bir mamogram görüntüsü) üzerinden filtreler ileletilerek standart özelliklerin otomatik olarak çıkarılmasında kullanılır. Bu nedenle, daha fazla sayıda girdi, anormalliği sınıflandırmak için kullanışlı bir modelin ortaya çıkmasına yol açar.

Göğüs kanseri teşhisi alanında veri seti elde etme zorluğunu en aza indirmek için uygulanan veri büyütme algoritmalarıyla birlikte önceden eğitilmiş modellerin kullanımı sağlandı. Öte yandan, görüntü ön işleme teknikleri görüntü girdilerini istenen forma getirmek için uygulandı.

Bu çalışmada, önceden eğitilmiş iki ana model senaryosu ele alınmıştır; VGG16 ve ResNet50. Her ikisi de farklı derinliklerde ve hiper parametrelerinde ince ayarlamalar yapılarak uygulanmıştır. Önceden eğitilmiş VGG16 modeli ile, deneysel çalışmalarımız sonucunda, son

tam bağı katmanı üzerinde ince ayarlar yapıldıktan sonra, iyi performans değerleri elde edilmiştir.

Önerilen VGG16 modeli, anormallik türünü kalsifikasyon / kütle olarak sınıflandırmasını yaparken, F1 skoru ve AUC değeri %82 olarak elde edilmiş, diğer derin öğrenme algoritmalarından daha iyi performans göstermiştir. Model ayrıca mamogramları dört grupta sınıflandırırken, ortalama AUC değeri %80 olarak hesaplanmış ve yüksek bir performans elde edilmiştir. Önerilen model ile, “iyi huylu kalsifikasyon”, “kötü huylu kalsifikasyon”, “iyi huylu kitle”, ve “kötü huylu kitle” sınıflandırması yapılmıştır.

Radyologların meme lezyonlarının anormal patolojisi hakkında daha kesin karar vermelerine yardımcı olmak için nihai bir uygulama yapılmış ve tam mamografi görüntüsü göstermektedir. Uygulama aynı zamanda teşhis süresinin azaltılmasına ve dolayısıyla erken tespit süresinin artmasına da yardımcı olur.

Anahtar Kelimeler: CNN, Transfer Öğrenme, Meme Anormalliği, Sınıflandırma, Mamografi, Kalsifikasyon, Kütle.

DEDICATION

*“I dedicate my dissertation work with special feeling of gratitude to
almighty God*

*Special thanks to my beloved parents whose affection, love, words of
encouragement and, nonstop prayers push me forward to such success.*

My Sister Lava, being my best cheerleader.

*My sister and wonderful niece Avy, for there continuous wishes of
success.*

*I appreciate the motivation and support that my supervisor gave me
whenever it was hard to keep my spirits high.”*

ACKNOWLEDGMENT

First and foremost, I want to thank my supervisor, Asst. Prof. Dr. Duygu Çelik Ertuğrul for allowing me to work under her guidance and provide continual support throughout my thesis. Each of the members of my dissertation committee has provided me extensive professional guidance and taught me a great deal about scientific research. I would especially like to thank my teachers and mentors through my master's program, who taught me more than I could ever give them credit for here. Their patience, motivation, and enthusiasm deserve special gratitude. I am indebted to them for providing me with a conducive atmosphere to carry out my thesis.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZ	v
DEDICATION	vii
ACKNOWLEDGMENT	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xiv
1 INTRODUCTION	1
1.1 Motivation	1
1.2 Thesis Objective	2
1.3 Thesis Organization.....	2
2 BACKGROUND	4
2.1 Breast Cancer	4
2.2 Breast Anatomy	4
2.3 Breast Abnormality	5
2.3.1 Calcification.....	5
2.3.2 Mass	6
2.4 Breast Imaging Techniques	6
2.5 Computer-Aided Detection	7
2.6 Deep Neural Network.....	8
2.7 Convolutional Neural Networks (CNN).....	10
2.7.1 CNN Architecture	12
2.7.2 CNN Parameters	14

2.8 Transfer Learning	14
2.8.1 VGG16.....	15
2.8.2 ResNet50.....	15
3 LITERATURE REVIEW.....	16
3.1 Literature Review	16
4 DATA.....	21
4.1 Data Sources	21
4.1.1 MIAS	21
4.1.2 INbreast.....	21
4.1.3 CBIS-DDSM.....	22
4.2 Data Partitions	23
4.3 Data Organization.....	24
5 METHODOLOGY.....	25
5.1 Toolkit and Platform	25
5.2 System Workflow	26
5.3 Pre-processing Data.....	26
5.3.1 Data Conversion	27
5.3.2 Image Processing	27
5.4 Data Augmentation.....	29
5.5 Model Configuration	29
6 RESULTS OF EXPERIMENTAL STUDY	38
6.1 Evaluation Metrics	38
6.2 System Result Analysis	39
7 CONCLUSION.....	47
7.1 Summary	47

7.2 Future work	48
7.3 Recommendations	48
REFERENCES.....	50

LIST OF TABLES

Table 1: Summarization of mammogram databases' pros and cons	22
Table 2: CBIS-DDSM dataset taxonomy.....	23
Table 3: DL toolkits summarization.	26
Table 4: Comparison results between classifiers.	42

LIST OF FIGURES

Figure 1: Breast anatomy.	5
Figure 2: Calcification and mass abnormalities in mammograms.	6
Figure 3: Left and right breasts in MLO and CC views.....	7
Figure 4: Deep neural network structure.....	9
Figure 5: Convolutional Neural Network concept.	11
Figure 6: Kernel mathematically.....	12
Figure 7: Classification procedure stages.	17
Figure 8: Deep convolutional neural network in breast cancer statistics.....	19
Figure 9: Breast cancer classification system workflow.....	27
Figure 10: Image processing	28
Figure 11: VGG16 Architecture.....	30
Figure 12: The original Resnet50 Architecture [104].....	31
Figure 13: ReLU activation function.	33
Figure 14: Flowchart of our system's methodology.....	36
Figure 15: Fine-tuned VGG16 structure.	37
Figure 16: Confusion matrix for binary classification.	38
Figure 17: a) VGG16 ROC before training. b) VGG16 ROC after training.....	41
Figure 18: a) ResNet50 ROC before training. b) ResNet50 ROC after training.	41
Figure 19: ROC curves for each class with VGG16.	43
Figure 20: Screenshots of our system's results for random cases.	45

LIST OF ABBREVIATIONS

AI	Artificial Intelligence.
API	Application Programming Interface.
BIRADS	Breast Imaging Reporting and Data System.
CAD	Computer-Aided Diagnosis.
CBIS-DDSM	Curated Breast Imaging Subset of DDSM.
CC	Cranio Caudal.
CLAHE	Contrast Limited Adaptive Histogram Equalization.
CNN	Convolutional Neural Network.
DDSM	Digital Database for Screening Mammography.
DL	Deep Learning.
DNN	Deep Neural Network.
MIAS	Mammographic Image Analysis Society.
MLO	Medio Lateral Oblique.
ROI	Region Of Interest.
VGG	Visual Geometry Group from oxford.
WHO	World Health Organization.

Chapter 1

INTRODUCTION

1.1 Motivation

Breast cancer is a fatal disease caused by uncertain external and internal factors; breast cancer can affect people not only physically but also emotionally as it alters one's life along with their beloved ones.

Breast cancer affects women and men in a lower ratio, it has the second most death ratio after respiratory cancer among all the cancer deaths in the world.

Early diagnosis could reverse case severity and prevent deaths. Therefore, scientists and researchers concentrated their work on developing and enhancing imaging techniques.

The most frequently used imaging technique for breast cancer diagnosis is mammography helped to diagnose cancer conveniently. However, mammography could only reduce the ratio of death slightly.

Computer-aided diagnosing systems were introduced in the last decades using image processing techniques to enhance the image quality obtained which made mammogram interpretation quite better by the radiologists.

The manual interpretation by specialists is a slow and monotonous procedure that inspired researchers into creating systems that can predicate the possible cancer cases using the machine

learning method. Machine learning is not the most efficient way to process images due to the high computations required to extract distinctive features. On the other hand the deep learning method CNN is universally attractive because it's computationally efficient and detects features from images automatically without human supervision.

1.2 Thesis Objective

This thesis aims to improve the detection of breast cancers in mammography images by:

- Implementing the deep learning algorithm, a convolutional neural network specifically transfers learning with the use of deeplearning4j API.
- Combined image processing methods for preprocessing the mammogram images with deep learning algorithms for automated feature extraction.
- Finding the best model configuration through various experimental trials that can predicate the mammograms better than the specialists.
- Gain the trust of specialists and radiologists by expanding the research area in the medical field diagnosing with its practical success.

1.3 Thesis Organization

This thesis is organized into seven chapters comprising the introduction chapter, the other six chapters are as follows:

Chapter 2 gives background information explaining breast anatomy and abnormality, the imaging techniques used in the medical area, what is computer-aided diagnosis systems and what are the fundamentals of the main methodologies used in the modern computer-aided diagnosis systems.

Chapter 3 presents a detailed survey on breast cancer diagnosis systems made by other researchers describing the strength and weaknesses of the related works.

Chapter 4 presents the popular available dataset in this field, the chosen dataset in this study manifesting its properties and strength.

Chapter 5 is the most important section which elaborates on the platform, methods, and, algorithms used in all the stages of designing the classification system.

Chapter 6 presents a discussion and analysis of the experimental trials and results obtained until reaching the satisfying model accuracy.

Chapter 7 is the summarization of the outcomes of this dissertation including the challenges it faces. It also provides possible future works to build a comprehensive diagnosing system.

Chapter 2

BACKGROUND

This initial chapter presents an introductory explanation about breast cancer, abnormalities, and medical screening techniques.

2.1 Breast Cancer

Breast cancer is a disease that occurs whenever an abnormal growth of cells or uncontrollable division takes place in breast tissues [1].

According to the World Health Organization (WHO), the highest occurrence of cancer cases among women is breast cancer affecting 2.1 million women annually which consequently results in the greatest cancer-related deaths. However, breast cancer is not attached to women only perhaps it impacts men also but with approximate 1% of all breast cancer cases [1]. Breast Cancer is also ranked the highest death rate among all cancers among women after lung cancers [3].

2.2 Breast Anatomy

Generally, breast tissues consist of skin, fat, and glands. The tree-alike architecture of breast tissues forms the glands in which wrapped with fat under the skin. The fat/non-dense tissues and the glandular/dense tissues are not distributed evenly among women like the breast transition from predominantly glandular to predominant fatty tissue causing the reduction of support in breast structure by age [4]. Figure 1 presents normal breast anatomy.

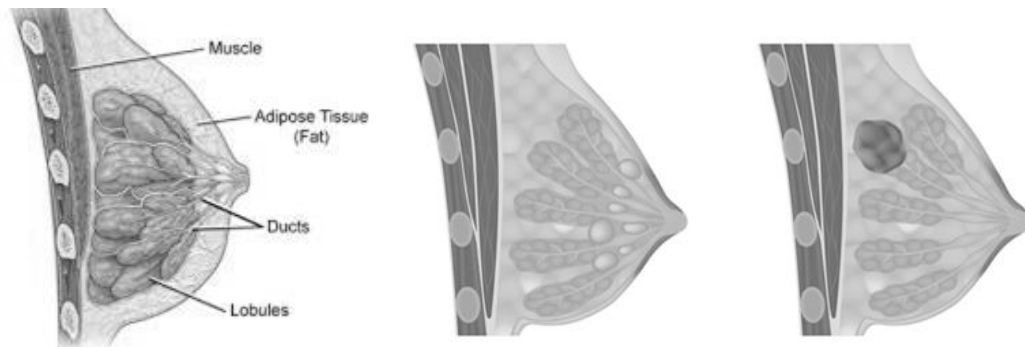


Figure 1: Breast anatomy for normal, mass, and calcification cases from left to right.

2.3 Breast Abnormality

Breast tissues may experience different changes in tissue structure or shape, such as physical distortions or the presence of unusual lumps, these lumps can mainly be categorized to:

2.3.1 Calcification

Calcification formed from calcium deposits in the dense tissue, this calcium is not related to individuals' diet perhaps it indicates an underlying process in the tissue. More often this happens with aging or when milk glands secrete milk in the ducts. As shown in Figure 1.

Calcification does not consider to be cancerous all the time. It can be one of either:

- Non-cancerous/ benign.
- Cancerous/ malignant.

Broadly, calcification is a small white clustered spot scattered in the breast tissue. There is a little shape variation when observing calcification from spheres to a popcorn-like shape or lines with sizes range between 0.1 – 1.0 mm, the larger the less suspicious to be malignant [5-7].

2.3.2 Mass

Mass or lump is an area of the dense tissue in the breast with edges and certain shapes that is odd in comparison to the normal tissues, as illustrated in Figure 1 previously. Masses as calcification can be malignant or benign but luckily the malignant cases are much less in masses [6].

2.4 Breast Imaging Techniques

The premium survival key to breast cancer is in the early detection that showed a drastic improvement in expanding patients' life span of [1]. For that reason, imaging modalities are required to assist radiologists to diagnose possible breast abnormalities. Several screening techniques are available in the medical field for that purpose, such as:

- X-ray.
- MRI
- Mammography

Mammography is the most preferable procedure to examine human breasts to identify the existence of any abnormalities by the use of low-dose x-rays to visualize the internal structure of the breast tissues [8, 9]. Figure 2 shows the mammogram samples of calcification and mass.

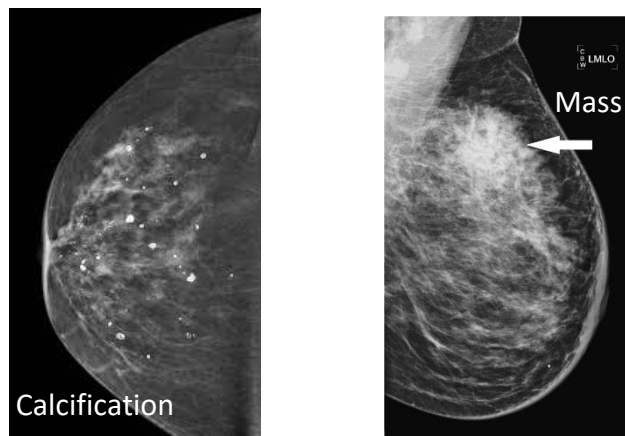


Figure 2: Calcification and mass abnormalities in mammograms.

Mammography-based screening normally acquires two viewpoints for each breast; top view Cranio Caudal (CC) or side view Mediolateral Oblique (MLO), seen in Figure 3. In the regular mammography examination, the breast is compressed and MLO, and CC radiographs are taken. Though further image projections may be requested to ensure suspicious regions.

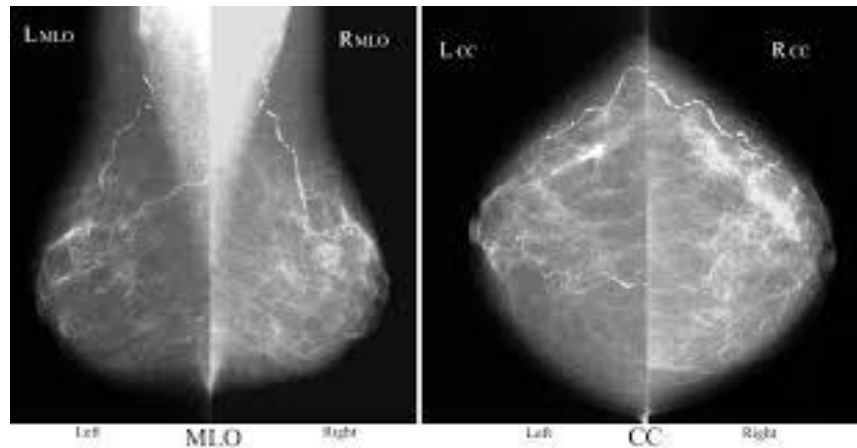


Figure 3: Left and right breasts in MLO and CC views.

2.5 Computer-Aided Detection

In 2018, breast cancer's death ratio was 15 percent of the death in women out of the total number of cancer-related death [1].

That indicated that the process of detection is not at the expected pace since interpreting the massive number of mammograms is a tedious and exhausting work of human beings generally and radiologists specifically. Even with modern medical screening technologies, radiologists are prone to misdiagnoses.

Studies stated that the error rate in diagnosing is estimated to be around 30% [7], 10-11]. Some other survey studies reported that most of the legal cases against radiologists were due to their failure in interpreting the medical images [12].

All the above issues motivated researchers to develop a Computer-Aided Diagnosis (CAD) system to assist medical specialists as a decision support tool.

The accuracy of breast cancer detection raised slightly because the majority of the CAD systems were using traditional machine learning techniques and manual crafting of features using image processing and segmentation methods for feature extraction in an extremely time-consuming manner.

In the last decade, deep learning concepts attracted the most attention of researchers to put in practice those techniques into the enhancement of the diagnosing of medical images, hence developing a higher quality of CAD systems [13]-16]. Deep learning is a special part of the artificial neural network that simulates the cognition system in humans [17]. In general, it performs classification directly from images, texts, and, sounds [18].

2.6 Deep Neural Network

Humanity always dreamed of intelligent machines that can assist in simplifying human tasks in multiple ways starting from the monotonous daily actions to the incredible mathematical calculations. Traditional computers proved this capability but always required human interactions for solving the problems.

On the other hand, artificial intelligence aims to give machines the ability to understand, recognize, detect, or even solve problems from raw inputs without the necessity of analyzing the discriminative features of the input data. In another word, treating the machine as a child in their early stage of learning when they learn to distinguish things by examples, not a formula.

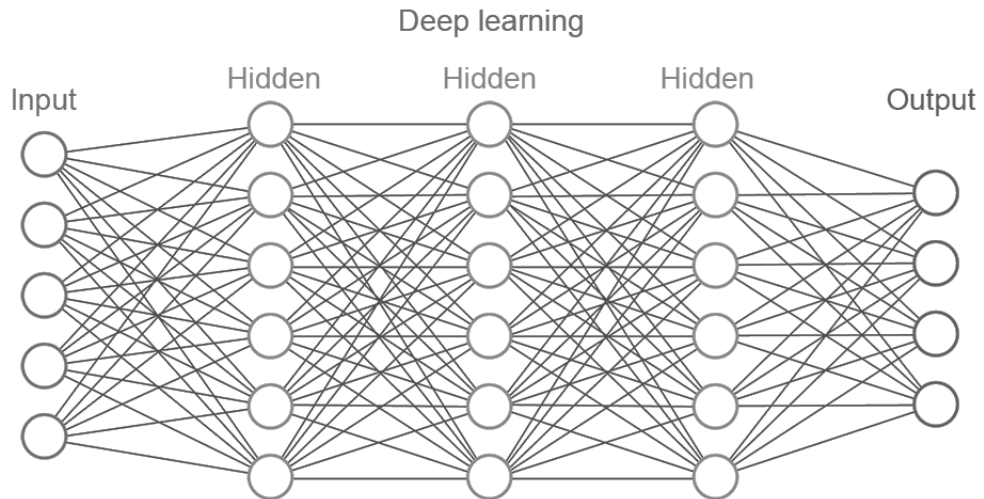


Figure 4: Deep neural network structure

A deep neural network (DNN) is simply a minor category of AI that allows machines to discover the right representation of a raw input for classification and detection. It is multiple layers of non-linear processing start from one level into a more abstract level transforming input to a slightly summarized data, after applying enough transformations a complex function can be learned. An illustration of the DNN structure is shown in Figure 4.

In classifications, the layers of DNN imply the characteristics of the input data that discriminate a class for another. For instance, in image inputs, the input is represented as an array of pixel values. In the first layer, the NN examines the presence of edges, the orientation and, image location. Then the second layer spots specific edges arrangement regardless of variations and later in the third layer NN assemble these arrangements into larger combinations to detect familiar objects and so on.

The key concept of DNN is that these transformations and filters are designed automatically without human feature engineering. DNN has many methods and algorithms that work with the

same concept but with some differentiation in architecture, weights, and hyper-parameters but this study tackles CNN and Transfer learning.

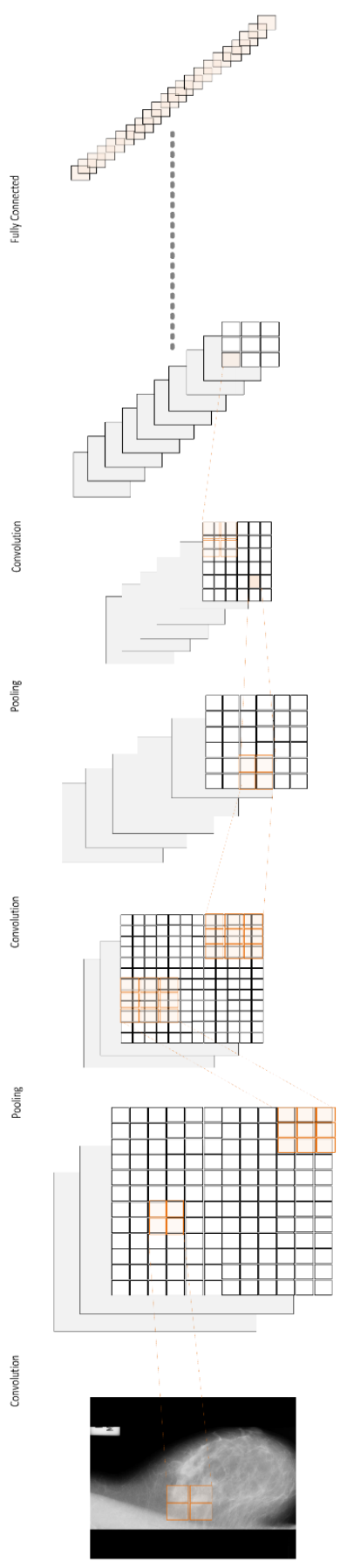
2.7 Convolutional Neural Networks (CNN)

In recent years, convolutional neural networks widely used and gained great success in image-focused tasks due to the allowance of encoding image-specific features to the architecture while reducing the setup parameters of a model.

Since medical images are characterized by local structures like texture, edges, curves, and corners and feature engineering is hard to be described by exact mathematical descriptors. Therefore, CNN applications in the medical domain are increased rapidly in the last decade. CNN is similar to the traditional neural networks consisting of neurons that self-optimize through learning. Neurons receive inputs and operate some processes (for example, a scalar followed by a non-linear function). The entire network will express a single score function (weights) from the vector image to the final output of the class. In Convolution networks, the last layer contains loss functions for classes, in addition to all tips for developing a traditional neural network. Figure 5 presents the concept of the CNN structure.

CNN can detect distinctive features in mammograms by striding kernels (i.e. matrix of weights of size $h*w$) on the array of pixel values to multiple it with the input matrix to extract features. See Figure 6. Kernels are learned by CNN automatically (not manually).

A combination of these filters gets more complex by each layer. As the complexity increase, CNN gets closer to identifying the specific feature.



Convolution Neural Network Concept

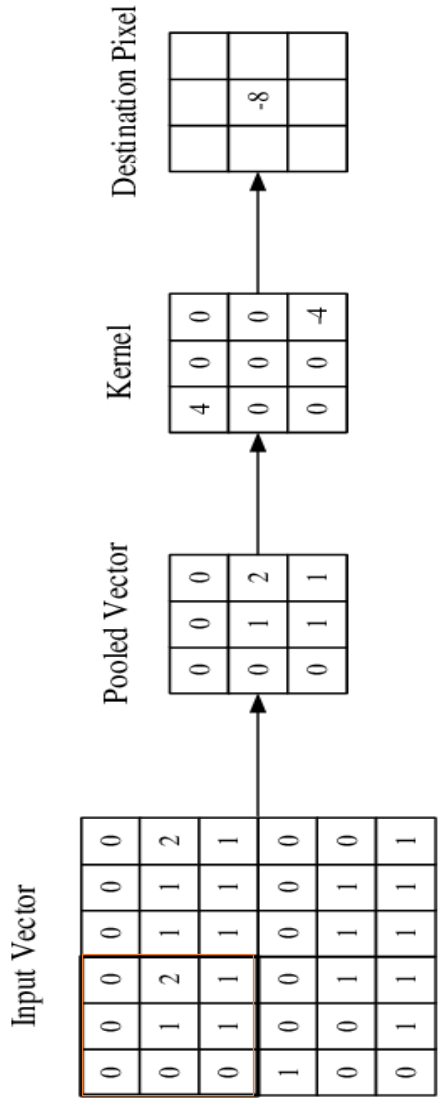


Figure 5: Convolutional Neural Network concept.

2.7.1 CNN Architecture

CNN's overall architecture is composed of multiple layers stacked on each other forming the CNN model. Initially, data are presented as the input layer holding the pixel value of the image.

Later, the CNN layers which categorized into three main layers: convolutional layers, pooling layers and, fully connected layers that are organized in a way to build the neural network.

- **Convolution Layer**

A generalized linear model for the underlying image patch, which determines the output of neurons connected to the local region using scalar production between the local region and the weights.

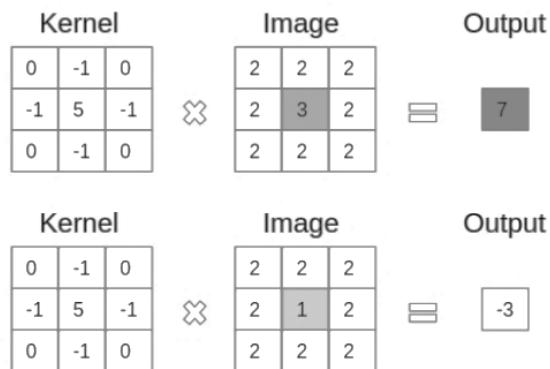


Figure 6: Kernel mathematically

The convolution layer focuses on using kernels which small spatial dimensions matrix that spread along with the depth of the input. The output of striding these kernels with the input produces activation maps stacked along the depth dimension and activated by an activation function.

Activation functions improve the performance of CNN in certain tasks by introducing non-linearity that is desirable for non-linear functions. Some of the typical activation functions are Sigmoid, SoftMax, and, ReLU.

As a result of the convolution layer, a huge number of parameters (weights) are produced. For example, if the input is an image of size $32 \times 32 \times 3$ (RGB image with the dimensionality of 32×32) and we set the receptive field size as 6×6 , we would have a total of 108 weights on each neuron only within the convolutional layer.

- **Pooling Layer**

This layer achieves shift-invariance with the reduction of feature map resolution (i.e. down-sampling along spatial dimension). The pooling layer is usually placed between two convolution layers lowering the computational-burden reducing the connection between the layers, see Figure 5.

Generally, the pooling layer reduces dimension, parameters, and computational complexity. For example, if the pooling filter was with size 2×2 and striding by 2 this result in a 25% reduction. Typically, there are two main types of pooling operations: max pooling and, average pooling.

- **Fully Connected Layer**

It takes the neurons of a previous layer and assembles every single neuron in the current layer generating semantically global information like a traditional neural network produces a class score from activations functions, see Figure 5. The fully connected layer is not essential as it can be replaced with a 1×1 convolution layer.

2.7.2 CNN Parameters

The understanding of the general architecture of CNN will not create an efficient model without understanding some hyper-parameters that are used to adjust the training and learning according to the problem given.

Overcoming the over-fitting problem in a model requires assigning regularization parameter to penalize the weight matrices in each neuron, L2 regularization forces the weights to decay to zero.

Another important parameter is defining the optimization algorithm for the learning process. Stochastic Gradient Descent is one of the popular algorithms that simply update the parameter for every training take steps to the bottom until finding the best parameter.

Finally, specifying the loss function which calculates the difference between the target output and the output from the model and aiming to minimize it.

2.8 Transfer Learning

Transfer learning is the process of re-using a solution in a specific domain which produces a specific target in another domain to produce another target. This concept is applied in the convolutional neural network by researchers with different classification tasks. Several pre-trained CNN models are created and trained on a massive dataset ImageNet which is composed of 1.2 million of image for 1000 categories of daily life object like a cat, dogs, flowers, vehicle and much more, each category have 100,000 images.

The architectural structure of pre-trained CNN models has the same concept of the complex CNN models. Transfer learning main parts of a model consist of input, feature extractor, classifier, and lastly the output.

Transfer learning can be repurposed using three main strategies to fine-tune the original model according to the problems hence achieving an accurate classification model:

- Train the entire model.
- Train some layers and freeze others.
- Freeze the convolutional base.

2.8.1 VGG16

The architecture of this model was designed by (Simonyan & Zisserman 2014) [88], the original model fed with RGB images of dimension 224x224 which then passed to the 16-convolution layer for feature extraction to end up classifying 1000 class images. The original VGG16 model architecture consists of 13 convolutional layers with 2 dense layers and lastly an output layer. The network has also 5 max-pooling layers.

2.8.2 ResNet50

The Residual Networks (ResNet50) that was designed by He et al. [20] is a very complex deep CNN model trained on the huge dataset ImageNet and was a winner of the ImageNet challenge in 2015.

Square images of size 224 fed to ResNet50 architecture that consists of convolutional layers, pooling layers, residual layers (a stack of three convolution layer followed by a batch normalizer) dense layers, and an output layer.

Resnet uses identity mapping which helps in avoid over-fitting in the training set by passing CNN weights if the current layer was not necessary.

Chapter 3

LITERATURE REVIEW

This chapter presents a survey on the related works done and the methods implemented throughout the stages of breast cancer detection.

3.1 Literature Review

CAD systems provide an objective view to radiologist diagnosing the mammograms and proved the effectiveness of CAD systems. However, the accuracy of detection is still challenging [17]. Recent advances in computational technologies, image processing, machine learning, and, the acceptance of mammogram images have opened up opportunities in tackling the challenging issues with the use of deep learning [18, 21, 22].

Although deep learning has been studied for decades [23-25], recently researchers focused on DL to improve the CAD systems performance not only in breast cancer but in many other medical fields. Authors in [25] developed a model classifying the skin lesions images using 129450 clinical labeled images as input achieving competitive performance concerning dermatologists. Similarly, a deep learning method developed in [26] showed high sensitivity and specificity level in detecting diabetic retinopathy in retinal photographs using 128175 annotated retinal images, and many other models in the medical fields.

Different studies applied different methods to execute the stages of developing a model for breast classification which is prescribed in Figure 7 below.

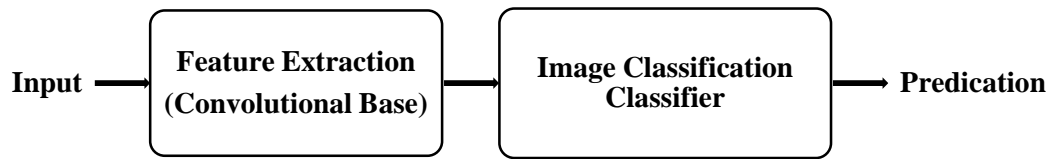


Figure 7: Classification procedure stages.

Firstly, in relevance to developing models; in 1995 [27, 28] investigated shallow CNN architecture, and recently both [29, 30] implemented the same technique. However, the authors in [31-37] presented more complicated architectures; [38, 39] both stated that deep CNN performed better than the former shallow layers but [38] concluded that neither a shallow architecture nor an exaggerated deep architecture is useful in mammograms classification; perhaps a model with moderated depth can distinguish features better. Some researchers [39-48] showed that combining hand-crafted features with features extracted by a CNN model score higher results for mass/ calcification lesion classifications depending on the size of the datasets.

On the other hand, other researchers [32, 49-55] tried to improve the accuracy of classification using the transfer learning concept. The majority used models pre-trained with general images in the “ImageNet” dataset [56]. The commonly CNN architectures used in transfer learning studies for mammography are Alex-Net [32, 39, 42, 49, 51, 53, 57-60], VGG16 [33, 50, 58, 61-67], ResNet50 [58, 61-66, 68].

Worth mentioning, transfer learning models have been fine-tuned in [29, 31-33, 49, 53, 55, 64, 69] while the works in [32, 49, 50, 53] preferred to combine the transferred model with one or several CNN layers.

Secondly, a comparison among the studies regarding the input data of the models varied from using the full image of mammograms like in [37, [65], 67, 70 -72] or using the ROI images as

in the studies [32, 60, 62, 64, 73]. In either input types ROI or Full mammogram images, the abnormality detection is usually done with two different mammogram views. The authors [40, 41, [57]7, 70, [74]4-[76] who used multi-view mammograms assisted to significantly improve performance and aid in the reduction of the false negative and false positive scores when compared to the other studies using a single mammogram view such as the works in [31, 77]. Although most of the models use low-resolution images feasible to the existing hardware specification, we can notice that the authors in [70] used high-resolution images in a CNN model clarifying that the performance raises when using the actual resolutions with a downside of increasing hardware requirements.

Furthermore, the main aim of the classifiers in the related works was either localization or classification of breast lesions such as the studies [29]29, 35, 42, 47, 60, 64, 67, and 78] or lesion classification tasks only. The researches [39, 40, 75, 78, 79, 80-82] were interested in lesion classifications applying CNN with probabilities for mass, calcification, and normal classes. While the studies in [31-33, 41, 43, 48- 51, 53, 73, 85] were interested in classifications of three or two classes of benign, malignant and normal in masses, the studies [75, 83] were about the classification of calcification.

Researchers achieved the above aims using multi-stage or E2E methods, while multi-stage methods consist of multiple stages of preprocessing, segmentation, feature extraction, and classification which have been implemented by the authors in [39, 74, 75, 79, 80]. End-to-end (E2E) takes all the stages and replace it with one single neural network [61, 86, 87]. The key to capture distinctive features using the E2E method is having sufficient data when training the model. While the multi-stage method found effective in reducing false positive detection.

Developing a model from scratch requires a lot of work beyond the skills of medical imaging researchers for that reason lots of library toolkits can be found in different programming languages. Platforms widely used are PyTorch [19], TensorFlow [88], Keras, Caffe [89], deeplearning4j, and many others depending on researchers' preferences from the quality of toolkit documentation, computational speed, GUP support, and the ease of the language.

Finally, Figure 8 shows the ratio of studies employing some of the best practices of CNN discussed previously according to a survey study done in 2019 [90].

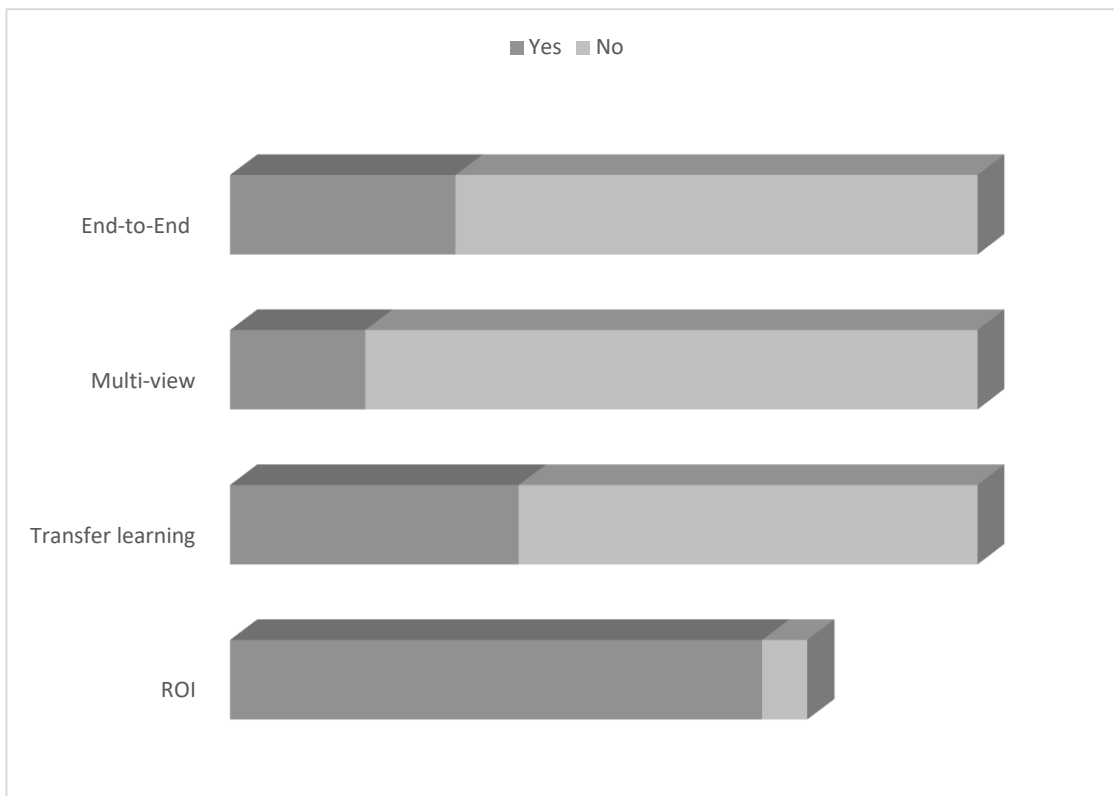


Figure 8: Deep convolutional neural network in breast cancer statistics.

This survey summarizes the recommendations for a significantly improved CNN model in classifications of mammogram images as:

- Input: larger dataset with higher image resolution if computationally feasible, multi-view, and ROIs are more competent in feature extraction.
- Model: going deeper in layers with a huge dataset or pre-trained transfer learning for smaller datasets.

Chapter 4

DATA

This chapter explains various available mammogram datasets stating the main characteristics of each option, followed by the detailed information and partitions of the chosen dataset.

4.1 Data Sources

There are various well-known mammogram datasets in this area that are freely available and easily accessible for studies and researches. Some of the most used datasets are:

4.1.1 MIAS

The Mammographic Image Analysis Society (MIAS) database available at Essex University since 1994 contains 322 digitized film tapes along with radiologists' annotation for the abnormality it presents. All images were padded or clipped to have equal size of square images 1024x1024 with PGM file formats [91]. Despite the drawback shown in Table 1 like being a limited size dataset and being old, it has been widely used [50, 74, 86].

4.1.2 INbreast

The INbreast database was made of images collected at a Breast Centre of the University Hospital (Hospital de São João, Breast Centre, and Porto, Portugal) in 2008 - 2010. The database was composed of 410 images for 115 participants. The digitalized mammograms are in DICOM file formats with BIRADS annotations done by specialists in the mammography fields [92]. The dataset gained attention [71, 75, 78, 80, 84, 87] due to its segmentation accuracy and high resolution but it's still in the development phase and has other drawbacks presented in Table 1.

Table 1: Summarization of mammogram databases' pros and cons

Database	Pros	Cons
MIAS	<ul style="list-style-type: none"> • Widely used. 	<ul style="list-style-type: none"> • Low resolution. • MLO view only. • Limited size.
INbreast	<ul style="list-style-type: none"> • Standard file format. • Precise lesion position. 	<ul style="list-style-type: none"> • Limited size. • No longer supported. • Old.
DDSM	<ul style="list-style-type: none"> • Widely used. • Different lesion shapes 	<ul style="list-style-type: none"> • Compressed files. • No accurate lesion position. • Contain falsely annotated cases.
CBIS-DDSM	<ul style="list-style-type: none"> • Most recent. • ROI provided. 	<ul style="list-style-type: none"> • Contain no normal cases.

4.1.3 CBIS-DDSM

The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is a curated subset of the Digital Database of Screening Mammography (DDSM) database of 2,620 scanned mammograms [93].

The mammogram films in both database CBIS-DDSM and DDSM represent two different image-views the Cranio-caudal/top view (CC) and the Medio-Lateral Oblique/side view (MLO) as shown previously in Chapter 2 in Figure 3.

While the DDSM database contains both normal and abnormal breast mammograms published in 1997[94], the CBIS-DDSM dataset was organized by trained mammography technologists in 2010.

Each participant had given 10 different patient IDs in the database, therefore, the database has only 1566 actual patients out of 6671 decompressed mammograms.

The mammogram films were scanned, decompressed, and converted to DICOM file formats. Additionally, a series of processes on DDSM implemented to attain an enhanced dataset, including:

- Removing the uncertain mass cases.
- Eliminating library compile restrictions when decompressing the lossless JPEG (LJPEG) by converting the images into DICOM file formats.
- Cropping ROIs to be at the centroid of the square crops.
- Standardizing test and train dataset.

DDSM is commonly picked by researchers [35, 41, 67, 73, 77, 83] likewise CBIS-DDSM dataset chosen in [20, 58, 64, 65, 68].

On the ground of the above details and the fact that deep learning requires a reasonably big data to achieve an accurate model, this study chooses CBIS-DDSM as the data source.

4.2 Data Partitions

CBIS-DDSM dataset originally splits into a train set and test set with a ratio of 80:20 percentage of the whole dataset. Table 2 shows the number of samples in each partition and category.

Table 2: CBIS-DDSM dataset taxonomy

Dataset 6671							
Train set				Test set			
Full Images 2458		ROI Images 2864		Full Images 645		ROI Images 704	
Calc.	Mass	Calc.	Mass	Calc.	Mass	Calc.	Mass
1227	1231	1546	1318	284	361	326	378

This study is using hold-out validation for the validation partition to keep the training dataset as big as possible. The study also applies balanced filtration on the training dataset to assign equal weights on the importance of each sub-class.

The compressed data provided by The Cancer Imaging Archive (TICA) can be downloaded with the use of NBIA data retriever software creating folders with name patterns (abnormality-dataset partition_P_patient ID_breast side_view type) such as (Calc-Training_P_00011_LEFT_MLO)and all the mammogram files have unified name of (000000.dcm).

4.3 Data Organization

The original corps are arranged in separate folders for training and testing with two subfolders for each abnormality. While the files have been renamed with a pattern constructed from the first letter of abnormality type followed by the first letter of the pathology type_sampleNumber_patientID (e.g. cB_3_7 represents a calcification abnormality with benign pathology third sample and patient ID 7).

Chapter 5

METHODOLOGY

This chapter starts with an illustration of the steps taken until a mammogram predicated. It also covers the prevalent toolkits and platforms in deep learning getting cross the study's choice of platform. Additionally, it includes the core stage of developing the classifier for mammogram abnormalities elaborating on its architecture and the hyper-parameters.

5.1 Toolkit and Platform

Libraries in different languages have been provided for deep learning purposes with a variation of pros and cons. Developers' choice depends strongly on the problem in hand with consideration of the capabilities of the framework and the available hardware specifications. Table 3 shows a summary of some of these toolkits. Table 3 shows a comparison of a range of these libraries.

However, this study developed the system utilizing deeplearning4j with a hardware specification 8GB NVIDIA GTX 1660i. Whereas, Matlab used in applying the preprocessing methods due to its simplicity in processing images.

The main challenge with deeplearning4j is the limitation in documentation as it is not considered to be stable yet. Throughout the study period, the API experienced lots of changes leading to deprecating some methods and add others.

Table 3: DL toolkits summarization.

API	Language	Advantages	Disadvantages
PyTorch	Python	Multi GPU support. Custom data loader. Memory management. Widely used. Pre-trained models available.	Spotty documentations. Less plugin.
MatConvNet	Matlab	Multi-core processor and GPU. Parallelism. Create, train, and visualize the NN toolbox. Pre-trained models available.	Focused on CNN only.
Deeplearning4j	Java	Multiple GPU and CPU. Parallel computations. Available pre-trained models Widely used. Fast execution	Not stable. Limited documentation.
TensorFlow	Python	CPU, GPU support. Many documentations are available.	No external libraries. No pre-trained models. Slow execution time.

5.2 System Workflow

Developing a system to predicate the abnormality pass through many stages. Each stage applies one or more methods to achieve an accurate prediction. Figure 9 illustrates the workflow of the breast cancer classification system.

5.3 Pre-processing Data

Image processing algorithms have been applied by many researchers [31, 32, 43] in the aim of better feature extractions, some of the mainly performed methods are:

Breast Cancer Classification Workflow

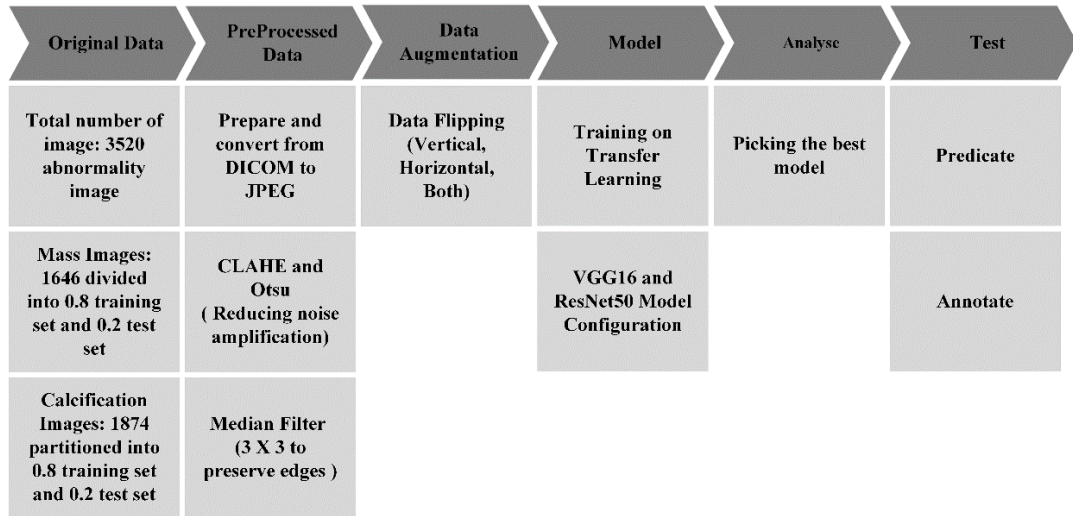


Figure 9: Breast cancer classification system workflow.

5.3.1 Data Conversion

The file formats of the digitalized mammograms in the CBIS-DDSM dataset demand conversion from DICOM type to a more convenient file type (JPEG) that can be accessed later by deeplearning4j easily.

Moreover, medical images mainly use 8-bits per grayscale pixel representing 256 shades but some other use 12-bit or 16-bit per pixel. Therefore, the entire dataset is unified to be in Unit 8 format.

5.3.2 Image Processing

One of the crucial issues which are observable solely with the full mammogram images is the artifact objects (i.e. numbers, characters, and symbols) located in the background that impacts the CNN development.

For that reason, a mask image was created with the largest blob from a binary image using the Otsu segmentation method [95] to distinguish the breast area from the background area. Figure 10 shows the real results of a sample of the dataset.

Furthermore, Filtering methods such as Contrast limited Adaptive Histogram Equalization (CLAHE) [96] and median filter applied on both full images and ROI images to improve image quality, reducing noise and preserving edges to further extend.

While [37, 85, 97-101] used the CLAHE technique, Sukassini and Velmurugan [102] measured the performance of median and mean filters and the result of experiments stated that the median filter scored better in mammograms' noise removal. Figure 10 demonstrates the effects of the above methods applied to a sample of the mammogram image from the CBIS-DDSM dataset.

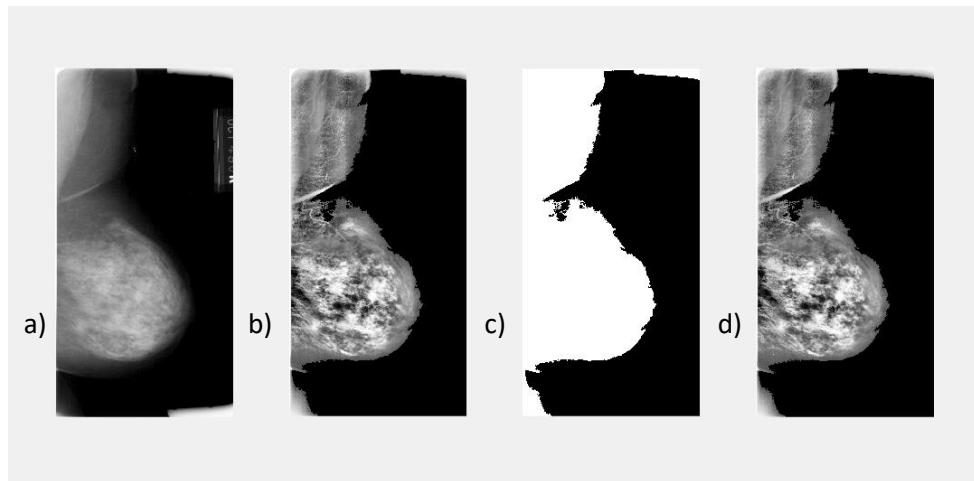


Figure 10: Image processing: a) Original image. b) Artifact removal. c) Mask. d) Image enhancement.

Finally, since images of the dataset were of various dimensions the entire images of the dataset resized to square images of 224x224 and converted to RGB to match the models' input shape.

5.4 Data Augmentation

Data augmentation is one of the popular ways to increase dataset size to double or more of the original size along with assisting in preventing over-fitting problem [103] when training deep learning models with comparatively small dataset size. This technique been used in many studies [10, 12, 29-33, 37, 42-44, 51, 52, 54, 72, 80, 81, 87].

This study applied transformations by flipping images both horizontally and vertically multiplied the initial dataset size five times and presented the lesions in different orientations.

5.5 Model Configuration

Transfer learning concepts implemented keeping the original structures of the models and only updating the hyper-parameters of the models through various trials of experiments until the best configuration for classification is found.

The different experiments fine-tuned (adjusted) the original architectural composition of VGG16 illustrated in Figure 11 and ResNet50 in Figure 12 in different depth. Either by fine-tuning the last fully connected layer or the two last connected layers.

The researchers used ResNet50 in the studies [33, 58, 61, 62, 65, 66, 68] for classification task purpose. Whereas the studies [31, 56, 59, 61-65] show the potential applications of VGG based CNN models in breast cancer.

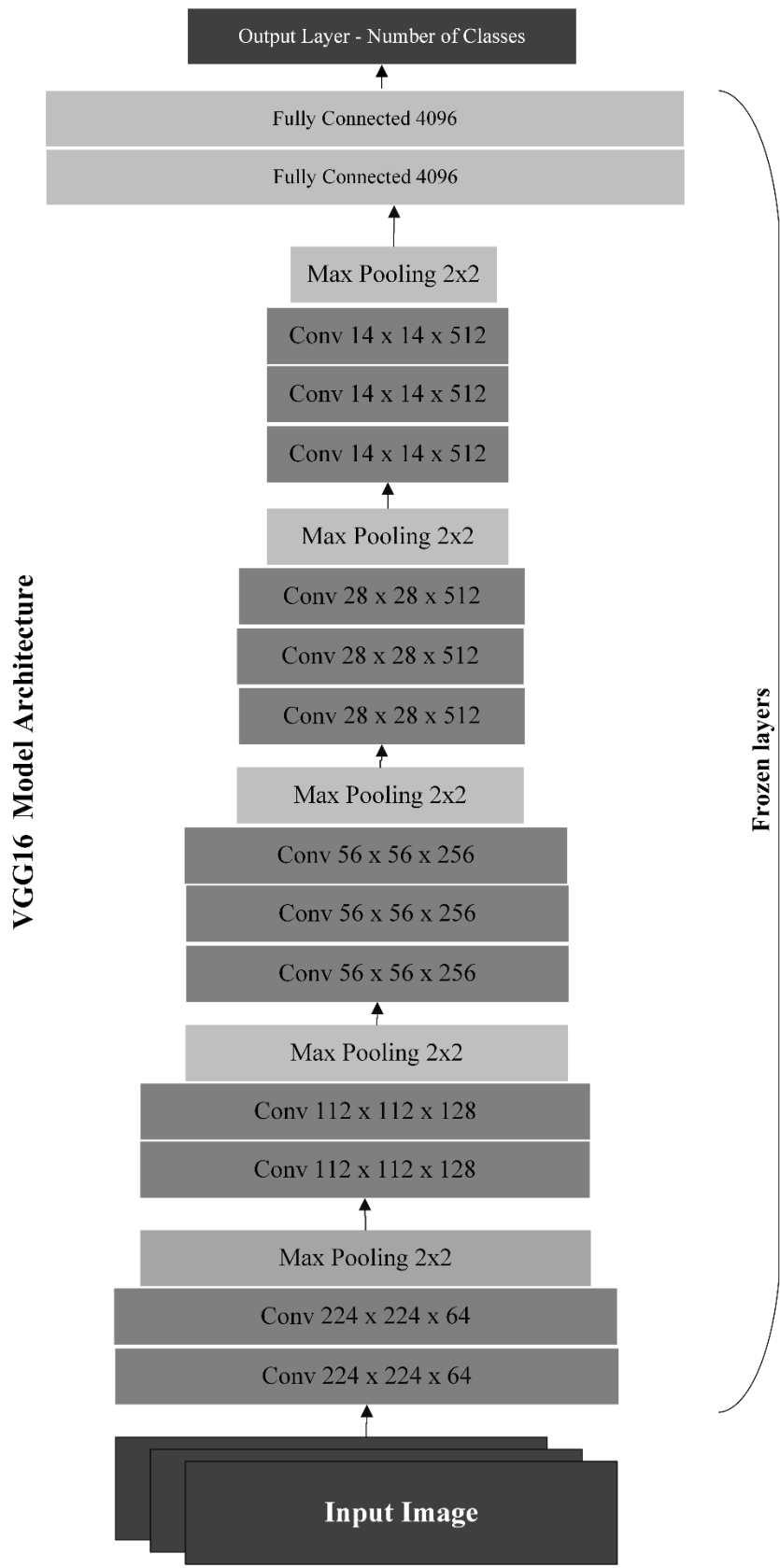


Figure 11: VGG16 Architecture.

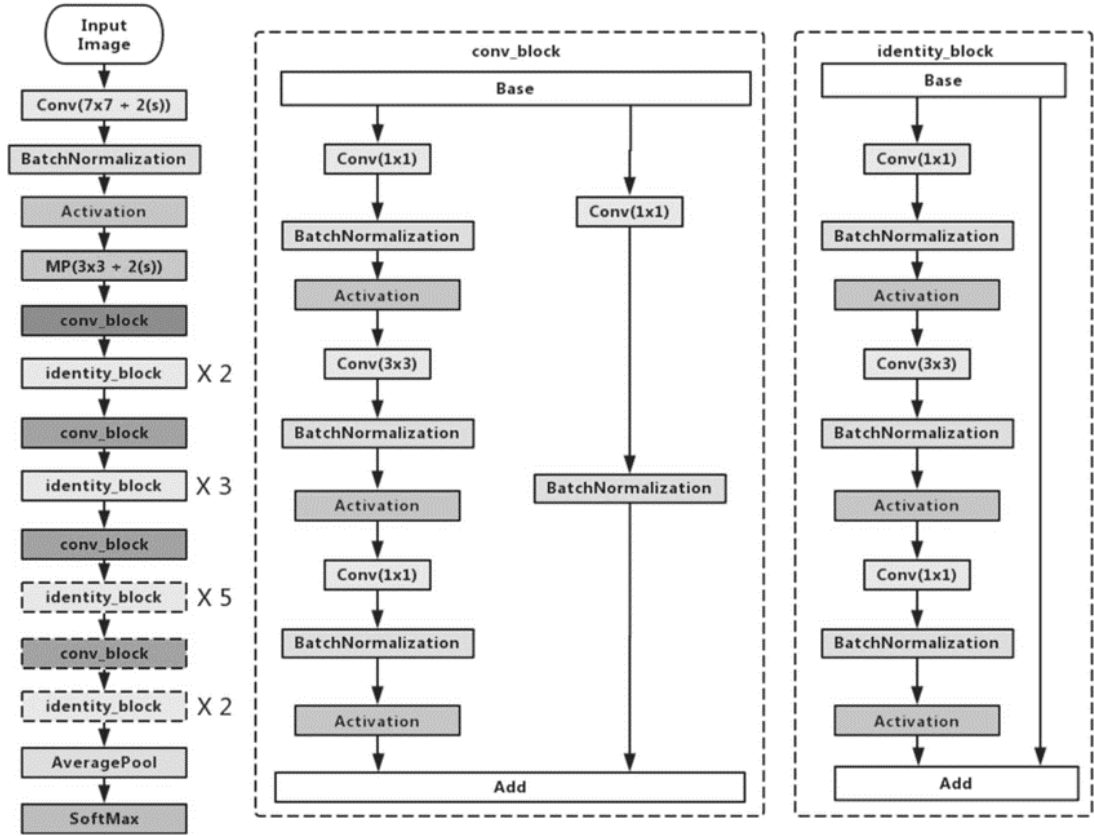


Figure 12: The original Resnet50 Architecture [104].

The parameters of the output layer for both architecture VGG16 and ResNet50 set to use Adam optimizer [105] as an optimization function instead of the Stochastic Gradient Descent (SGD) optimizer that is used in the original model.

Adam optimizer [105] is used to compute individual learning rates for different parameters to maximize model efficiency following the below equation:

$$w_t = w_{t-1} - \eta \frac{m_t}{\sqrt{v_t + \epsilon}} \quad (1)$$

where:

w : current model weight.

η : step size, depending on the iteration.

v_t, m_t : calculated estimators.

ϵ : Epsilon.

The optimizer is trying to find the optimal point of the cost function evaluated on a random mini-batch of data. In each iteration, the weights of a feature will be updated with the help of step size and computing the gradient to calculate the biased corrected mean estimator and variance estimator for a certain moment. moment coefficient which represents the gradient percentage possessed in every iteration is equal to (0.9) because smaller values turns out to flacuate a lot which means the value get closer to noisy data while bigger values such as (0.9) get showed smoother gradient curve. [106].

The step size in Adam optimizer is invariant to the computed gradient and can be updated better as it navigates quickly through tiny gradients (saddles and ravines), unlike the SGD.

The cross-entropy loss function [107] is applied as the validation loss function to measure the classification performance by minimizing the distance between two distributions of probabilities.

The output of cross-entropy is a probability between 0 and 1 whereas the actual predication is either 0 or 1. Therefore, the output probability is tried to be as close as possible to the actual predictions, it can be calculated for binary classification as:

$$L = -y(\log(p) + (1 - y) \log(1 - p)) \quad (2)$$

where,

L: the validation loss, y: the actual labels, p: the predicted labels.

The training is done with batch size 64 at most for a maximum of 100 epochs using a learning rate that ranges from 10^{-4} to 10^{-7} . When the CNN model initialized using the pre-trained weights, the feature learned during the initial training are more sensitive, so a smaller learning rate is used with the default momentum value of 0.9.

The activation functions [108] for the whole convolutional layers is ReLU which is a piecewise linear function that prunes the negative parts to zero and returns the positive parts only as applied in [51, 52, 84] See Figure 13, ReLU can be calculated mathematically:

$$A_{i,j,k} = \max(Z_{i,j,k}, 0) \quad (3)$$

where, A: the activation function in i, j, k position, Z: the input image at position i, j, k.

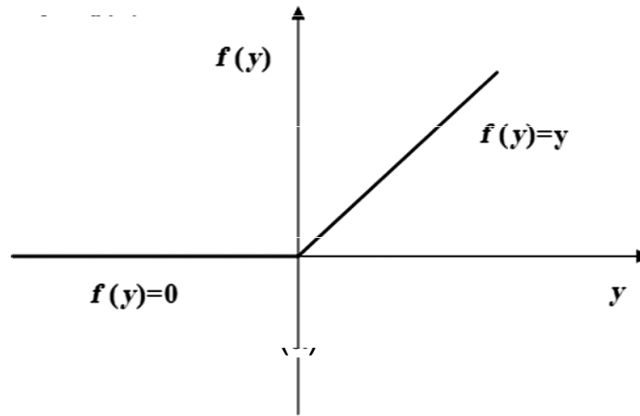


Figure 13: ReLU activation function.

Whereas the activation function for the output layer is dependable on the number of classes. Thus, in the models that classify the abnormalities between mass and calcification, the Sigmoid function used to predict the probability based on the output, as calculated below:

$$f(x_i) = \frac{1}{1 + e^{-x}} \quad (4)$$

Models that classify the ROIs depending on the abnormality as well as the pathology (benign calcification, malignant calcification, benign mass, and malignant calcification) of it used the activation function SoftMax that produces an output ranging from 0 to 1 by computing the probability distribution from a vector, using the formula below:

$$f(x_i) = \frac{e^{(x_i)}}{\sum_j e^{(x_i)}} \quad (5)$$

The weight initializations are set to Xavier [109] which is similar to random initialization but turns out to perform much better.

Xavier initialization prevents layer activation output from vanishing through training by setting layers' weights to a random value chosen in between the uniform distribution boundary. The relationship between the input and output variance may vary dramatically according to the initialized weight where variance with values greater than 1 lead to exploding, variance with values less than 1 vanish the forward signal. Therefore, the chosen value of weights can be calculated by the below equation:

$$\text{Uniform Distribution Boundary} = \pm \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}} \quad (6)$$

where,

n_i : the number of input connections to the network.

n_{i+1} , the number of output connections of the network.

The over-fitting problem [103] where the model starts to memorize the dataset and learns the noises presented with the dataset achieving high performance with the same dataset and poor performance with new datasets.

Over-fitting can be detected when the trained image fed to the model achieving a high accuracy rate whereas, feeding the validation dataset (unseen data) shows a decreased accuracy proving that the model learned attributes of the training set well but couldn't perform the same with the new unseen data.

Over-fitting has been suppressed in our system defining the L2 [110] regularization (weight decay). L2 can be defined as the square summation of all feature weights as shown in the equation below.

$$R(w) = \sum_i \sum_j w_{i,j}^2 \quad (7)$$

where w is the weight matrix.

The loss function is updated to control the amount of the regularization by assigning in the below equation to 0.1 penalizing the large weights and preferring the small one.

$$L = \frac{1}{n} \sum_{i=1}^n L_i + \lambda R(w) \quad (8)$$

The configuration of the previous hyper-parameter assisted in developing a classifier for breast abnormalities and pathology of breast abnormalities. Figure 14 illustrates a flowchart for the methodology used in our system during the training and prediction phase.

VGG16 abnormality classifier classifies the mammograms into two classes (calcification and mass) according to the abnormality presented in the mammogram whereas, the VGG16 pathology classifier classifies mammograms into one of four classes (benign calcification, malignant calcification, benign mass, and, malignant mass) showing pathologies' probabilities of the mammogram.

Although our system shows the detailed probability results in both classifiers, it also compares the result of both classifiers to predicate the class label if both classifiers agreed on the same abnormality

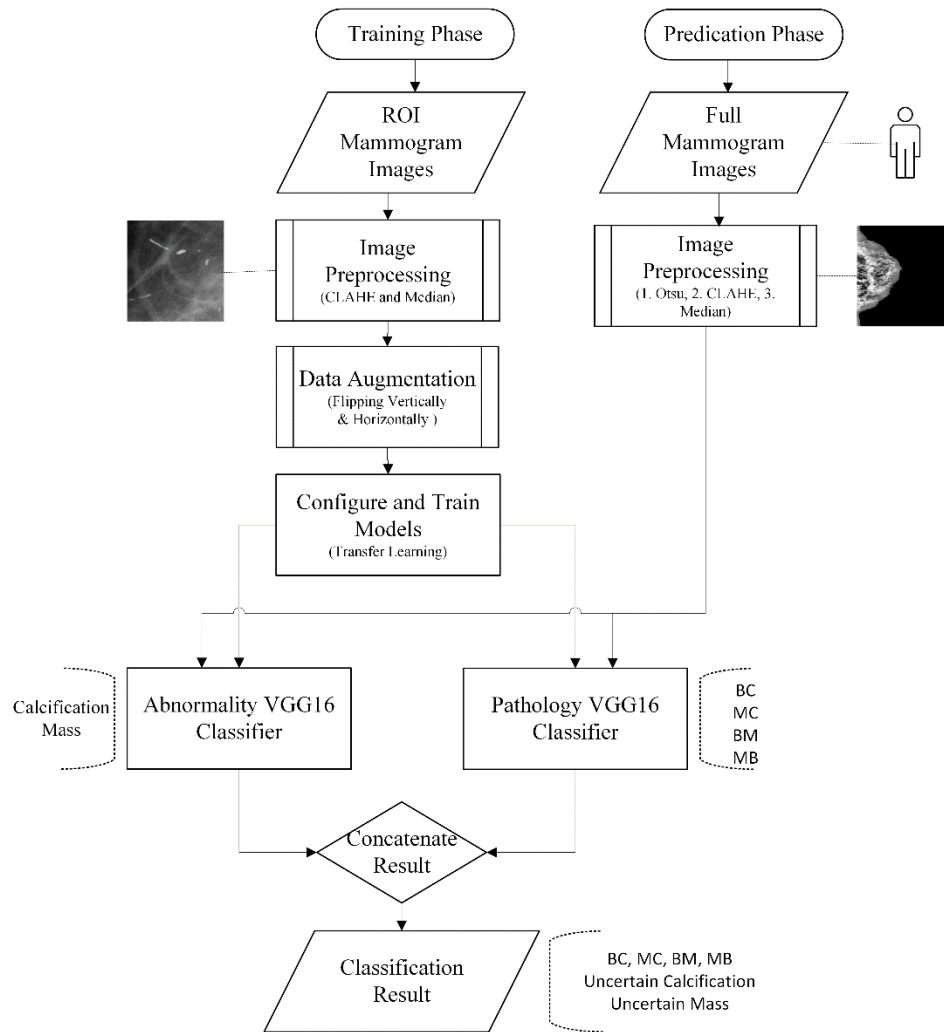


Figure 14: Flowchart of our system's methodology.

The flowchart clarifies that the training stage uses the ROI images, unlike the prediction stage which uses full mammogram images. Then, the input images of both phases being preprocessed separately. Later, the preprocessed images in the training phase convey into batches of size 64 which are fed into the classifiers (abnormality classifier and pathology classifier) to extract the distinctive features after training the unfrozen layers of the pre-trained VGG16 model.

Finally, the flowchart shows how both phases predicate the results for both classifiers giving the possible probability for each class label then compares these outcomes. The comparison step

```

=====
VertexName (VertexType)          nIn,nOut    TotalParams  ParamsShape          Vertex Inputs
=====
input_1 (InputVertex)           -, -        -            -                    -
block1_conv1 (Frozen ConvolutionLayer)  3,64       1,792       W:{64,3,3,3}, b:{1,64}  [input_1]
block1_conv2 (Frozen ConvolutionLayer)  64,64      36,928      W:{64,64,3,3}, b:{1,64} [block1_conv1]
block1_pool (Frozen SubsamplingLayer)   -, -        0            -                    [block1_conv2]
block2_conv1 (Frozen ConvolutionLayer)  64,128     73,856     W:{128,64,3,3}, b:{1,128} [block1_pool]
block2_conv2 (Frozen ConvolutionLayer)  128,128    147,584    W:{128,128,3,3}, b:{1,128} [block2_conv1]
block2_pool (Frozen SubsamplingLayer)   -, -        0            -                    [block2_conv2]
block3_conv1 (Frozen ConvolutionLayer)  128,256    295,168    W:{256,128,3,3}, b:{1,256} [block2_pool]
block3_conv2 (Frozen ConvolutionLayer)  256,256    590,080    W:{256,256,3,3}, b:{1,256} [block3_conv1]
block3_conv3 (Frozen ConvolutionLayer)  256,256    590,080    W:{256,256,3,3}, b:{1,256} [block3_conv2]
block3_pool (Frozen SubsamplingLayer)   -, -        0            -                    [block3_conv3]
block4_conv1 (Frozen ConvolutionLayer)  256,512    1,180,160  W:{512,256,3,3}, b:{1,512} [block3_pool]
block4_conv2 (Frozen ConvolutionLayer)  512,512    2,359,808  W:{512,512,3,3}, b:{1,512} [block4_conv1]
block4_conv3 (Frozen ConvolutionLayer)  512,512    2,359,808  W:{512,512,3,3}, b:{1,512} [block4_conv2]
block4_pool (Frozen SubsamplingLayer)   -, -        0            -                    [block4_conv3]
block5_conv1 (Frozen ConvolutionLayer)  512,512    2,359,808  W:{512,512,3,3}, b:{1,512} [block4_pool]
block5_conv2 (Frozen ConvolutionLayer)  512,512    2,359,808  W:{512,512,3,3}, b:{1,512} [block5_conv1]
block5_conv3 (Frozen ConvolutionLayer)  512,512    2,359,808  W:{512,512,3,3}, b:{1,512} [block5_conv2]
block5_pool (Frozen SubsamplingLayer)   -, -        0            -                    [block5_conv3]
flatten (FrozenVertex)            -, -        -            -                    [block5_pool]
fc1 (Frozen DenseLayer)           25088,4096  102,764,544 W:{25088,4096}, b:{1,4096} [flatten]
fc2 (Frozen DenseLayer)           4096,4096   16,781,312  W:{4096,4096}, b:{1,4096} [fc1]
predictions (OutputLayer)         4096,2      8,194       W:{4096,2}, b:{1,2}     [fc2]
=====
Total Parameters:  134,268,738
Trainable Parameters:  8,194
Frozen Parameters:  134,260,544
=====

```

Figure 15: Fine-tuned VGG16 structure.

occurs in away if both classifiers didn't state the same abnormality class then our system will indicate uncertain results that may require further medical investigations.

The trainable parameters in the VGG16 model of our system are only 8194 while the original VGG16 model has 134268738 parameters. On the other hand, the ResNet50 model has 25636712 total parameters, 4475906 parameters of them are trainable. Figure 15 shows the detailed structure of the fine-tuned VGG16 model in our system.

Training the above-mentioned parameters achieved accuracy that is higher in the VGG16 model than the accuracy of the ResNet50 model, the first scored %80 even though the number of the trainable parameters were much less than the ResNet50 model which scored only %60 accuracy.

Chapter 6

RESULTS OF EXPERIMENTAL STUDY

The previous chapter discusses the system architecture and its hyper-parameters and how to decide the best classifier for the task. This chapter analyzes the outcome results of different experiments mentioning the effects of specific parameters on the model performance. It also includes a visual demonstration of the system GUI and results.

6.1 Evaluation Metrics

Evaluation is essential for any project to measure the quality of a learning model or an algorithm, typically to quantify the performance of a predictive model a combination of individual evaluation matrices. The below confusion matrix in Figure 16 is required for clearly interpreting and understanding the metrics.

	Actual Class 1	Actual Class 2
Predicted Class 1	True Positive TP	False Positive FP
Predicted Class 2	False Negative FN	True Negative TN

Figure 16: Confusion matrix for binary classification.

- **Accuracy:** the easiest evaluation metric suitable for binary and multiclass classifications, it simply calculates the portion of true values among the examined cases.
- **Recall (TP / TP+FN):** is used when the goal is capturing as many possible positive cases.
- **Precision (TP / TP+FP):** It's a valid metric used when the certainty of the prediction is important.
- **F1 score:** harmonic mean of precision and recall. Calculate as below

$$F1 = 2 \times \frac{\text{recall} * \text{Precision}}{\text{recall} + \text{Precision}} \quad (9)$$

- **AUROC (Area Under Receiver Operating Curve):** measures how well a prediction is ranked rather than its actual value (i.e. scale-invariant) by calculating the area under the ROC curve.

The training set was used to update the weights of the network of both VGG16 and ResNet50 classifiers, while the validation set was used to measure how well the trained set is performing after each epoch.

6.2 System Result Analysis

The decision on the best model is made after trials of experiments on the training set and evaluations on the validation set to update the hyper-parameters until choosing the most suitable values for the system. The validation dataset is a hold-out partition of the training set.

In the initial experiments, firstly when using a pre-trained CNN model with the full breast image (background cropped) but without applying data augmentation the accuracy 34% obtained. This unsatisfactory accuracy is due to the full breast image where the exact abnormality region is not

observable, therefore changing the inputs to ROI images under the same circumstances up-raised the accuracy slightly to 48%.

The previous outcome manifests that the amount of data is considered insufficient and data augmentation is required (see Chapter 5). After increasing the training dataset applying the above mention technique the accuracy of 54% has been achieved.

As the first stage of trails decided the type of input patches for the training CNN models, the next stage of experiments investigated the proper model parameters.

These investigations confirmed that the larger the batch size the fastest the computations. It also verified that batch size and learning rate impact on generalization gap between train and validation data. The combination of large batch size with small learning rates or vice versa is found to be a balanced trade-off between them according to the computation speed and the hardware specifications. Furthermore, when fixing the learning rate at a certain value and doubling the batch size the performance improved %1 only.

In the aspect of epoch numbers, increasing it 5 times bigger observed to raise the performance of the CNN model lightly by %2 only and extended the training time.

The performance of the training using pre-trained weights obtained from ImageNet versus the performance of initializing the network randomly and training the samples from scratch achieved better in terms of accuracy and speed. The later classified the samples randomly with accuracy equal to 0.5 only and that's due to the small dataset size in comparison to ImageNet which contains hundreds of thousands of images for each class. Although, data augmentation has been applied, yet the augmented data is still correlated to the original data.

On the other hand, the computational speed for training only two epochs using small batch size (16) with the environment of (8 GB NVIDIA GTX 1660i) lasted for 24 hours while training in the same environment for a pre-trained model required merely 0.5 hours to train 100 epochs.

The nature of CBIS_DDSM is very different in content compared to the ImageNet dataset allowed to additional experiments of fine-tuning different deepness levels of the CNN model searching for performance improvement, either by training only the last fully connected layer or several layers.

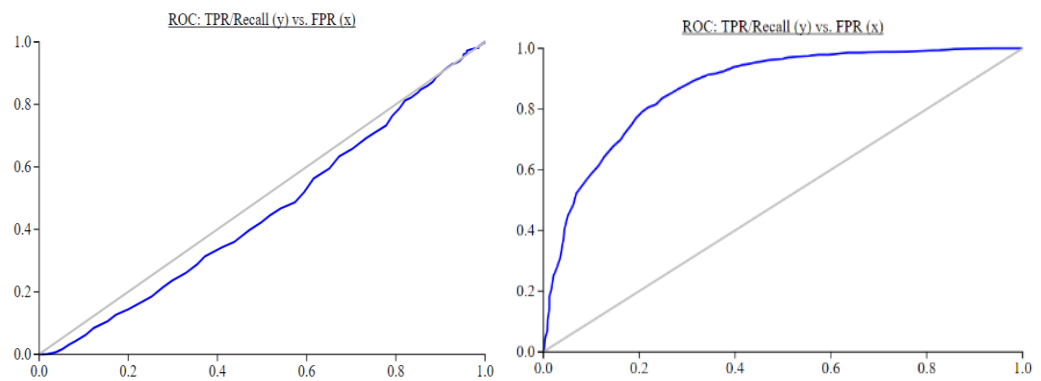


Figure 17: a) VGG16 ROC before training. b) VGG16 ROC after training.

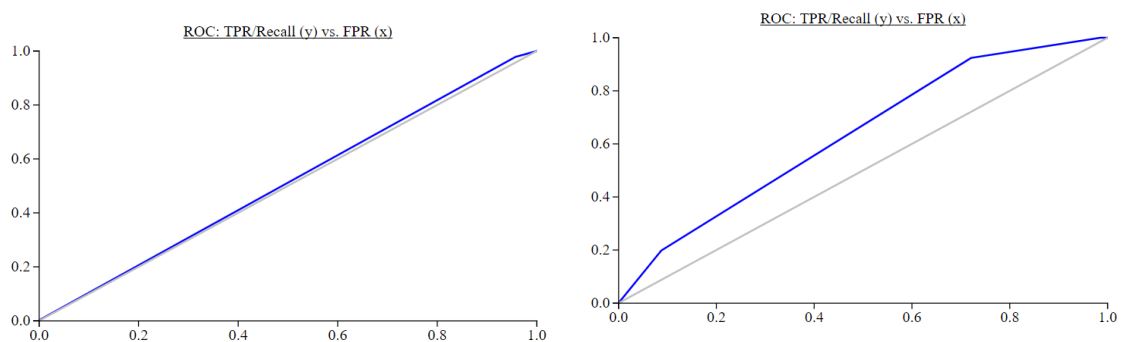


Figure 18: a) ResNet50 ROC before training. b) ResNet50 ROC after training.

Fine-tuning the last output layer by changing the activation function and the number of outputs in both models acquired a superlative score over fine-tuning the last three layers in VGG16 0

adding 3 more fully connected layers in the ResNet50. The former scored an accuracy of 0.8 while the later classified all the samples under the mass abnormality and failed to detect the calcification abnormality. In another trial, it could only classify the calcification abnormality.

Despite everything, all experiments scored more effectively after training rather than before the fitting process, Figure 17 shows the differences in the ROC curve of the VGG16 model whereas, Figure 18 shows the differences in the ResNet50 model.

Table 4: Comparison results between the system classifier and previous classifiers.

Researcher	Classification	Dataset	Image Type	Model	Acc.	AUC
Xavier [68]	Mass vs. Normal	CBIS-DDSM + INbreast	Patches*	VGG16	NA	NA
				ResNet50	0.84	0.92
Perez [63]	Mass Malignant vs. Mass Benign	CBIS-DDSM	ROI	VGG16	0.64	0.64
				ResNet50	0.84	0.84
Lazaros [62]	Mass Malignant vs. Mass Benign and Normal	CBIS-DDSM	ROI	VGG16	0.71	0.78
				ResNet50	0.74	0.80
LiShen [65]	Mass Malignant, Mass Benign, Calcification Malignant, Calcification Benign and Background	CBIS-DDSM+ INbreast	Patches*	VGG16	NA	0.84
				ResNet50	NA	0.63
Our system	Mass vs. Calcification	CBIS-DDSM	Training ROI to classify full images	VGG16	0.80	0.87
				ResNet50	0.60	0.66
Our system	Mass Malignant, Mass Benign, Calcification Malignant, Calcification Benign	CBIS-DDSM	Training ROI to classify full images	VGG16	0.72	0.73
				ResNet50	0.60	0.64

*Patches: cropping the full image into a certain number of smaller images.

The results obtained from the VGG16 classifier were the best having (F1 score=0.82) compared to what has been achieved by other researchers using the same dataset CBIS-DDSM but with a different classification goal. Contrary to the ResNet50 classifier which had poor performance (F1 score= 0.66) but not the least compared to related works. See Table 4.

While the systems in [62, 63, 68] were merely interested in predicting mass abnormality which considers being insufficient for reducing breast cancer rate, the work in this study along with [65] classified both abnormalities calcification and mass.

The model in [65] used 10 patches classifiers leading to a higher accuracy than the system's accuracy since the background patches had high probabilities of nearly 0.97.

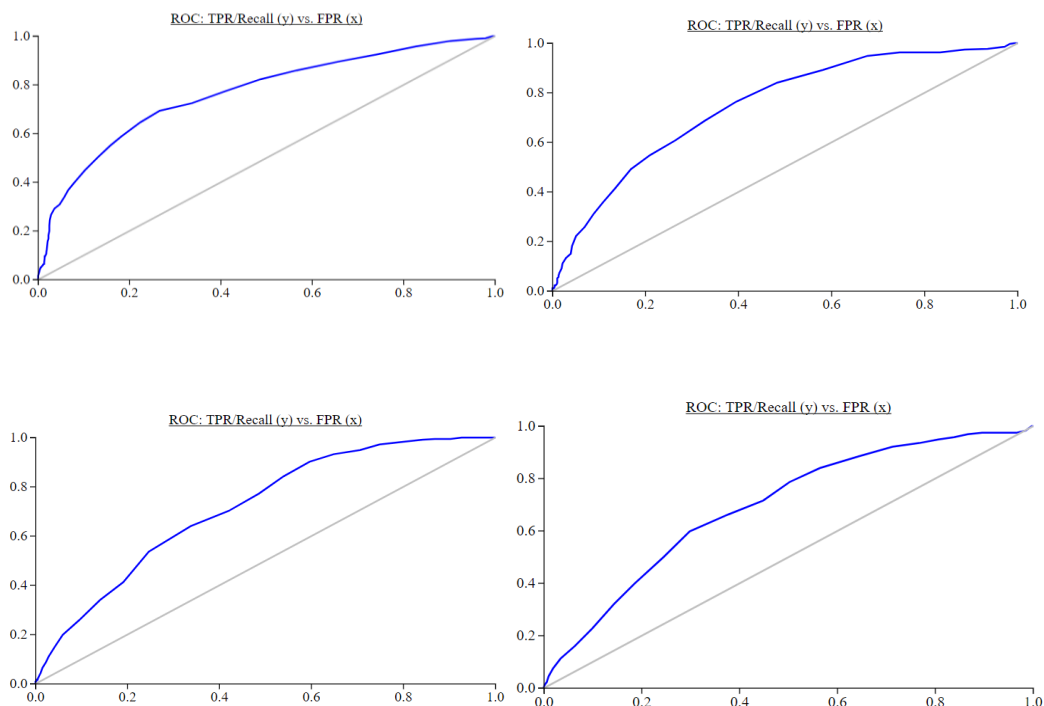


Figure 19: ROC curves for each class with VGG16.
a) Benign calcification. b) Malignant calcification. c) Benign mass. d) Malignant mass.

The classification of abnormality between mass and calcification is the easiest, malignant and benign is more challenging, especially in mass abnormality. Figure 19 shows the ROC curve for the four different classes. The AUC score for each was (benign calcification= 0.76, malignant calcification =0.74, benign mass = 0.71 and lastly, malignant mass=0.70).

The total number of parameters trained in the VGG16 classifier is 8194 whereas the trainable parameter in the ResNet50 classifier increased.

The motivation for building an easily accessible decision supporter tool led to a simple application. Screenshots of different cases of the mammogram classification application are presented in Figure 20.

Users of the application simply can choose the mammogram file using the “Browse Image” button. As soon as the image is successfully loaded the classification button will be activated automatically. The classification button classifies the image to one of the four classes (benign calcification, malignant calcification, benign mass and, malignant mass). The annotation button shows the detailed probabilities of the classification results.

The final prediction of the classification is made using most two accurate models achieved; the VGG16 model which classifies the mammograms to mass/calcification and the VGG16 model which classifies the mammograms to benign calcification, malignant calcification, benign mass and, malignant masses.

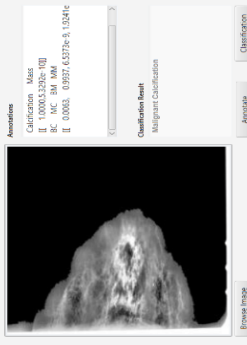
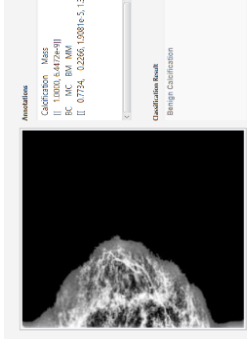
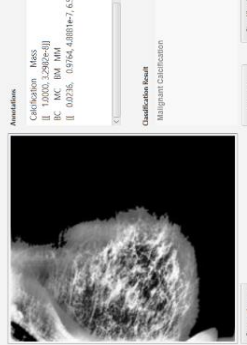
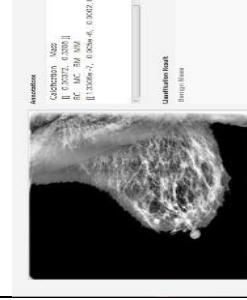
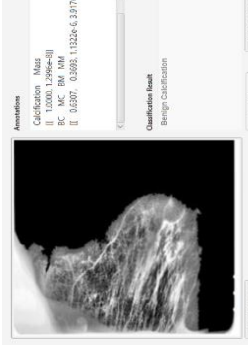
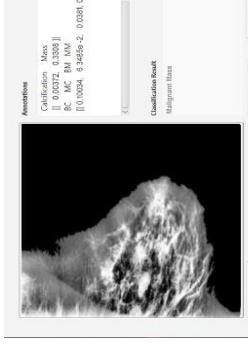
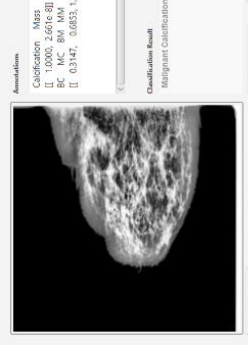
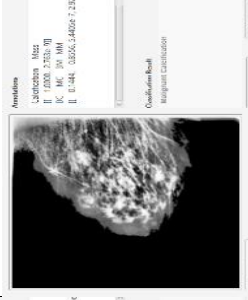
Actual Label	a) Malignant Calcification	b) Benign Calcification	c) Benign Calcification	d) Benign Mass
Our System Results				
Actual Label	e) Malignant Calcification	f) Malignant Mass	g) Malignant Calcification	h) Benign Mass
Our System Results				

Figure 20: Screenshot of our system's result for some random cases. The cases in (a, b, d, and, f) have been correctly classified. The cases (c) and (e) failed to classify the pathology correctly and case (h) falsely classified both the abnormality and the pathology of the mammogram.

Thus, if both classifiers agree on predicting the abnormality type then the pathology is determined. Otherwise, the application will indicate uncertain classification results so that further investigation will be decided by the specialists. The network is designed to be relatively fast since the system can infer a full mammogram at most in 25 seconds only considering the computational environment of (NVIDIA GTX 1660i 8GB).

Chapter 7

CONCLUSION

This chapter expresses in brief the results obtained as well as the contribution of this thesis and recommendations for future work.

7.1 Summary

The thesis focused on the problem of breast abnormality classifications in the mammogram images. A classification framework has been proposed for assisting the radiologist in diagnosing breast cancer cases with the least effort and time. The result showed higher accuracy than human's accuracy. The robustness of the methodology used showed vividly when comparing the result with previous efforts of researchers' in this field in terms of detailed classification and slightly higher accuracy.

Initially, the essential role of pre-processing the data was admitted after observing the amplification it adds to the region of interest where an abnormality exists providing a visual separation among breast tissues. The system deduces the complete decision benefitting from the automated CNN methodology in classifying mammograms. Later, the methodology took the smallest region of interest as patches and trained a VGG16 architecture. Training results on the CBIS_DDSM dataset demonstrated that the feature domain can be well adapted from natural images for breast abnormality classifications. The usage of the transfer learning technique enhanced the performance to a significant extent overtraining a CNN model from scratch. Thereafter, the system takes advantage of two different classifiers to conduct the last decision ensuring the certainty of it to the specialist.

The main accomplishment of this study can be summarized in rejuvenating the use of the inhibited CNN models by data limitation in breast cancer imaging fields with the use of efficient transfer learning models achieving an accuracy %80 using the VGG16 model and %60 using the ResNet50 model in classifying breast abnormality whereas the pathology classifier scored %72 using VGG16 model and %60 with the use of the ResNet50 model.

In conclusion, this study achieves an efficient classification accuracy for predicting both abnormality and pathology of the breast image. Our system requires further studying to achieve the predication of the normal cases along with the detection of the abnormal area precisely.

7.2 Future work

In this thesis, transfer learning has been proposed to develop an advanced decision supporter to assist radiologists in breast screening. The framework presented is considered to be limited in images. As part of future work, increasing the dataset size will be interesting to work with since each class requires at least 1000 samples to extract useful features more precisely. Additionally, the patch classifier in the system used the ROI of abnormal cases only. For the future, adding the normal patches will help not only to classify the abnormality type perhaps it will provide comprehensive classifications. Another subsequent improvement on the decision supporter tool is integrating segmentation function to detect the exact location of the abnormality in the breast images.

7.3 Recommendations

Deep learning revelation is dependable on the availability of big data, model algorithms, and preprocessing power. However, future works demand resolving issues like:

- Extensive collaboration between hospital stakeholders and deep learning scientists require to widen up to improve the quality of health services.

- Deep learning applications rely on massive annotated datasets, therefore sharing these data and making it available in different healthcare service providers may help to overcome the lack of annotated data.
- Changing neural network supervised learning into unsupervised learning is another way to overcome limited data availability annotated by experts.
- Boosting health professionals' trust in deep learning by unlocking the black-box, as the acceptance of the black-box could be a barrier as hospital provider may not be comfortable and experts have unanswered questions as who will be responsible for the wrong diagnosis.

REFERENCES

- [1] American Cancer Society. Breast Cancer Facts & Figures 2019-2020. Atlanta: American Cancer Society, Inc. 2019.
- [2] K. J. Ruddy & E. P. Winer, "Male breast cancer: risk factors, biology, diagnosis, treatment, and survivorship", Medical Oncology, Dana-Farber Cancer Institute, Boston, USA, 2012.
- [3] American Cancer Society. Cancer Facts & Figures 2020. Atlanta: American Cancer Society; 2020.
- [4] T. Nelson, L. Cerviño, J. Boone, and K. Lindfors, "Classification of breast computed tomography data", Medical Physics, vol. 35, no. 3, pp. 1078-1086, 2008.
- [5] "Understanding Breast Calcifications", Breastcancer.org, 2020. [Online]. Available: https://www.breastcancer.org/symptoms/testing/types/mammograms/mamm_show/calcifications.
- [6] "What Does the Doctor Look for on a Mammogram", Cancer.org, 2020. [Online]. Available: <https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/mammograms/what-does-the-doctor-look-for-on-a-mammogram.html>.

- [7] I. Schreer, "Dense Breast Tissue as an Important Risk Factor for Breast Cancer and Implications for Early Detection", *Breast Care*, vol. 4, no. 2, pp. 89-92, 2009.
- [8] J. S. Suri and R. M. Rangayyan, *Recent advances in breast imaging, mammography, and computer-aided diagnosis of breast cancer*. Bellingham, WA: SPIE Press, 2006.
- [9] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G Hislop, A. Chiarelli, S. Minkin, et al., "Mammographic density and the risk and detection of breast cancer", *New England Journal of Medicine*, vol. 356, no. 3, pp. 227–236, 2007. DOI: 10.1056/NEJMoa062790.
- [10] K. Kerlikowske, "Performance of Screening Mammography among Women with and without a First-Degree Relative with Breast Cancer", *Annals of Internal Medicine*, vol. 133, no. 11, p. 855, 2000. Available: 10.7326/0003-4819-133-11-200012050-00009.
- [11] L. Berlin, "Radiologic errors, past, present and future", *Diagnosis*, vol. 1, no. 1, pp. 79-84, 2014. Available: 10.1515/dx-2013-0012.
- [12] J. Whang and S. Baker, "The Causes of Medical Malpractice Suits against Radiologists in the United States", *Radiology*, vol. 266, no. 2, pp. 548-554, 2013. Available: 10.1148/radiol.12111119.

- [13] E. Kozegar, M. Soryani, B. Minaei, and I. Domingues, "Assessment of a novel mass detection algorithm in mammograms", *Journal of Cancer Research and Therapeutics*, vol. 9, no. 4, p. 592, 2013.
- [14] N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms", *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 611-619, 1996.
- [15] T. Kooi, B. van Ginneken, N. Karssemeijer, and A. den Heeten, "Discriminating solitary cysts from soft tissue lesions in mammography using a pre-trained deep convolutional neural network", *Medical Physics*, vol. 44, no. 3, pp. 1017-1027, 2017.
- [16] J. Desautels, R. Rangayyan and N. Mudigonda, "Gradient and texture analysis for the classification of mammographic masses", *IEEE Transactions on Medical Imaging*, vol. 19, no. 10, pp. 1032-1043, 2000.
- [17] Y. Li and H. Chen, "A Survey of Computer-aided Detection of Breast Cancer with Mammography", *Journal of Health & Medical Informatics*, vol. 7, no. 4, 2016.
- [18] G. Litjens, T. Kooi, B. Bejnordi, and A. Setio, "A survey on deep learning in medical image analysis", *Medical Image Analysis*, vol. 42, pp. 60-88, 2017.
- [19] A. Lerer et al., "Automatic differentiation in PyTorch", 2017.

- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. Available: 10.1109/cvpr.2016.
- [21] R. Zhao, R. Yan, and Z. Chen, "Deep learning and its applications to machine health monitoring", *Mechanical Systems and Signal Processing*, vol. 115, pp. 213-237, 2019.
- [22] J. Lee, S. Jun, and Y. Cho, "Deep Learning in Medical Imaging: General Overview", *Korean Journal of Radiology*, vol. 18, no. 4, p. 570, 2017.
- [23] Y. Lecun and Y. Bengio, "Convolutional Networks for Images, Speech, and Time-Series", *The Handbook of Brain Theory and Neural Networks*, 1995.
- [24] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, pp. 85-117, 2015. Available: 10.1016/j.neunet.2014.09.003.
- [25] A. Esteva, B. Kuprel, and R. Novoa, "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.
- [26] V. Gulshan, L. Peng, and M. Coram, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs", *JAMA*, vol. 316, no. 22, p. 2402, 2016.

- [27] H. Chan, S. Lo and B. Sahiner, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network", *Medical Physics*, vol. 22, no. 10, pp. 1555-1567, 1995.
- [28] Datong Wei, B. Sahiner, Heang-Ping Chan and N. Petrick, "Detection of masses on mammograms using a convolution neural network", 1995 International Conference on Acoustics, Speech, and Signal Processing.
- [29] M. Ertosun and D. Rubin, "Probabilistic visual search for masses within mammography images using deep learning", 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015.
- [30] T. Kooi, A. Gubern-Mérida, and C. Sánchez, "Large scale deep learning for computer-aided detection of mammographic lesions", *Medical Image Analysis*, vol. 35, pp. 303-312, 2017.
- [31] E. Sert, S. Ertekin, and U. Halici, "Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms", 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017.
- [32] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning", *Proceedings of the 5th International Conference on Bioinformatics and Computational Biology - ICBCB '17*, 2017.

- [33] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep Convolutional Neural Networks for breast cancer screening", *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 19-30, 2018.
- [34] C. Szegedy, S. Reed and D. Anguelov, "Going deeper with convolutions", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. Available: 10.1109/cvpr.2015.
- [35] R. Platania, S. Shams, S. Yang, and S. Park, "Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID)", *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2017
- [36] S. Charan, M. Khan, and K. Khurshid, "Breast cancer detection in mammograms using convolutional neural network", 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018.
- [37] H. Lu, E. Loh and S. Huang, "The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection", 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019.
- [38] N. Tajbakhsh, M. Gotway, and J. Liang, "On the Necessity of Fine-Tuned Convolutional Neural Networks for Medical Imaging", *Deep Learning and*

Convolutional Neural Networks for Medical Image Computing, pp. 181-193, 2017.

- [39] S. Suzuki, X. Zhang, and M. Yoshizawa, "Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis", 2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2016.
- [40] T. Kooi, B. van Ginneken, N. Karssemeijer, and A. den Heeten, "Discriminating solitary cysts from soft tissue lesions in mammography using a pre-trained deep convolutional neural network", *Medical Physics*, vol. 44, no. 3, pp. 1017-1027, 2017.
- [41] Yi D, Sawyer RL, Cohn III D, Dunnmon J, Lam C, Xiao X, et al. "Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors", arXiv preprint arXiv:170506362, 2017.
- [42] N. Dhungel, G. Carneiro, and A. Bradley, "Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests", 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2015.
- [43] J. Arevalo, F. González, R. Ramos-Pollán, J. Oliveira, and M. Guevara Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks", *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248-257, 2016.

- [44] J. Sharma, J. Rai, and R. Tewari, "Identification of pre-processing technique for enhancement of mammogram images", 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014.
- [45] N. Antropova, B. Huynh, and M. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets", *Medical Physics*, vol. 44, no. 10, pp. 5162-5171, 2017.
- [46] N. Dhungel, G. Carneiro, and A. Bradley, "The Automated Learning of Deep Features for Breast Mass Classification from Mammograms", *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 106-114, 2016.
- [47] N. Dhungel, G. Carneiro, and A. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention", *Medical Image Analysis*, vol. 37, pp. 114-128, 2017.
- [48] Y. Qiu, S. Yan, and M. Tan, "Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study", *Medical Imaging 2016: Computer-Aided Diagnosis*, 2016.
- [49] B. Huynh, H. Li and M. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks", *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016.

- [50] Gallego-Posada J, Montoya-Zapata D, Quintero-Montoya O. "Detection and Diagnosis of Breast Tumors using Deep Convolutional Neural Networks".
- [51] Levy D, Jain A. "Breastmass classification from mammograms using deep convolutional neural networks", arXiv preprint, arXiv: 161200542, 2016.
- [52] Z. Jiao, X. Gao, Y. Wang and J. Li, "A parasitic metric learning net for breast mass classification based on mammography", *Pattern Recognition*, vol. 75, pp. 292-301, 2018.
- [53] R. Samala, H. Chan, L. Hadjiiski, M. Helvie, K. Cha, and C. Richter, "Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms", *Physics in Medicine & Biology*, vol. 62, no. 23, pp. 8894-8908, 2017.
- [54] Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K., "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography", *Med Phys*,43(12):6654–66, 2016.
- [55] S. Suzuki et al., "WE-DE-207B-02: Detection of Masses On Mammograms Using Deep Convolutional Neural Network: A Feasibility Study", *Medical Physics*, vol. 43, no. 640, pp. 3817-3817, 2016.
- [56] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al., "Imagenet large scale visual recognition challenge", *Int J Comput Vis*, 3(115):211–52, 2015.

- [57] G. Carneiro, J. Nascimento, and A. Bradley, "Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning", *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2355-2365, 2017.
- [58] C. Shu and R. Goubran, "Abnormality Detection in Mammography using Deep Convolutional Neural Networks", arXiv preprint. ArXiv: 180301906, 2018.
- [59] X. Zhang, N. Jacobs, Q. Han and J. Liu, "Whole mammogram image classification with convolutional neural networks", 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017.
- [60] H. Cai, Q. Huang, W. Rong and J. Wang, "Breast Microcalcification Diagnosis Using Deep Convolutional Neural Network from Digital Mammograms", *Computational and Mathematical Methods in Medicine*, vol. 2019, pp. 1-10, 2019.
- [61] L. Shen, "End-to-end Training for Whole Image Breast Cancer Diagnosis using An All Convolutional Design", arXiv preprint. ArXiv: 170809427. 2017.
- [62] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, "Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study", *Journal of Imaging*, vol. 5, no. 3, p. 37, 2019.
- [63] L. Falconi, M. Perez, W. Aguila, and A. Conci, "Transfer Learning and Fine Tuning in Breast Mammogram Abnormalities Classification on CBIS-DDSM

Database", *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 2, pp. 154-165, 2020.

- [64] R. Agarwal, O. Diaz, X. Lladó, M. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks", *Journal of Medical Imaging*, vol. 6, no. 03, p. 1, 2019.
- [65] L. Shen, L. Margolies, J. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography", *Scientific Reports*, vol. 9, no. 1, 2019.
- [66] J. Park, K. Geras, and J. Phang, "Screening Mammogram Classification with Prior Exams", vol. 190713057, 2019.
- [67] H. Chougrad, H. Zouaki, and O. Alheyane, "Multi-label transfer learning for the early diagnosis of breast cancer", *Neurocomputing*, vol. 392, pp. 168-180, 2019.
- [68] O. Diaz, R. Marti, X. llado and R. Agarwal, "Mass detection in mammograms using pre-trained deep learning models", *14th International Workshop on Breast Imaging (IWBI 2018)*, 2018.
- [69] A. Mohamed, W. Berg, H. Peng, Y. Luo, R. Jankowitz, and S. Wu, "A deep learning method for classifying mammographic breast density categories", *Medical Physics*, vol. 45, no. 1, pp. 314-321, 2017.

- [70] Geras KJ, Wolfson S, Kim S, Moy L, Cho K., "High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks", arXiv preprint arXiv:170307047, 2017.
- [71] G. Carneiro, J. Nascimento, and A. Bradley, "Unregistered Multiview Mammogram Analysis with Pre-trained Deep Learning Models", Lecture Notes in Computer Science, pp. 652-660, 2015.
- [72] Hwang S, Kim HE. "Self-transfer learning for fully weakly supervised object localization", arXiv preprint arXiv: 160201625, 2016.
- [73] J. Arevalo, F. Gonzalez, R. Ramos-Pollan, J. Oliveira, and M. Guevara Lopez, "Convolutional neural networks for mammography mass lesion classification", 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.
- [74] J. Wang, H. Ding, F. Bidgoli, and B. Zhou, "Detecting Cardiovascular Disease from Mammograms With Deep Learning", IEEE Transactions on Medical Imaging, vol. 36, no. 5, pp. 1172-1181, 2017.
- [75] R. Ben-Ari, A. Akselrod-Ballin, L. Karlinsky, and S. Hashoul, "Domain-specific convolutional neural nets for detection of architectural distortion in mammograms", 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017.

- [76] N. Dhungel, G. Carneiro, and A. Bradley, "Fully automated classification of mammograms using deep residual neural networks", 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017.
- [77] G. Carneiro, J. Nascimento, and A. Bradley, "Deep Learning Models for Classifying Mammogram Exams Containing Unregistered Multi-View Images and Segmentation Maps of Lesions" This work is an extension of the paper published by the same authors at the Medical Image Computing and Computer-assisted Intervention (MICCAI 2015).", *Deep Learning for Medical Image Analysis*, pp. 321-339, 2017.
- [78] N. Dhungel, G. Carneiro and A. Bradley, "Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms", *Lecture Notes in Computer Science*, pp. 605-612, 2015.
- [79] Kooi T, Gubern-Merida A, Mordang JJ, Mann R, Pijnappel R, Schuur K, et al., "A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography", *International Workshop on Digital Mammography*, Springer, 2016.
- [80] Mordang JJ, Janssen T, Bria A, Kooi T, Gubern-Merida A, Karssemeijer N., "Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks.", *international Workshop on Digital Mammography*, Springer. p. 35-42, 2016.

- [81] Domingues I, Cardoso J. "Mass detection on mammogram images: a first assessment of deep learning techniques", 19th Portuguese Conference on Pattern Recognition (RECPAD), 2013.
- [82] I. Wichakam and P. Vateekul, "Combining deep convolutional networks and SVMs for mass detection on digital mammograms", 2016 8th International Conference on Knowledge and Smart Technology (KST), 2016.
- [83] A. Bria et al., "Spatial Enhancement by Dehazing for Detection of Microcalcifications with Convolutional Nets", Image Analysis and Processing - ICIAP 2017, pp. 288-298, 2017.
- [84] Q. Abbas, "DeepCAD: A Computer-Aided Diagnosis System for Mammographic Masses Using Deep Invariant Features", Computers, vol. 5, no. 4, p. 28, 2016.
- [85] M. Jadoon, Q. Zhang, I. Haq, S. Butt, and A. Jadoon, "Three-Class Mammogram Classification Based on Descriptive CNN Features", BioMed Research International, vol. 2017, pp. 1-11, 2017.
- [86] V. Raman, P. Sumari, H. Then and S. Al-Omari, "Review on Mammogram Mass Detection by Machine Learning Techniques", International Journal of Computer and Electrical Engineering, pp. 873-879, 2011.

- [87] N. El Atlas, M. El Aroussi and M. Wahbi, "Computer-aided breast cancer detection using mammograms: A review", 2014 Second World Conference on Complex Systems (WCCS), 2014.
- [88] J. Chen, Z. Chen, M. Abadi, and P. Barham, "TensorFlow: A system for large-scale machine learning", vol. 16, 2016.
- [89] Y. Jia, E. Shelhamer, "Caffe Zoo Model", 2015.
- [90] M. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future", Lecture Notes in Computational Vision and Biomechanics, pp. 323-350, 2017.
- [91] J. Suckling, D. Parker, S. Dance, and I. Astley, "The mammographic image analysis society digital mammogram database", *Excerpta Medica. International Congress Series*, pp. 375–8, 1994.
- [92] I. Moreira, I. Amaral, I. Domingues, M. Cardoso, A. Cardoso, and J. Cardoso, "INBreast: toward a full-field digital mammographic database", *Acad Radiol*, no. 19, pp. 236–48, 2012.
- [93] R. Lee, F. Gimenez, A. Hoogi, K. Miyake, M. Gorovoy, and D. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research", *Scientific Data*, vol. 4, no. 1, 2017. Available: [10.1038/sdata.2017.177](https://doi.org/10.1038/sdata.2017.177).

- [94] M. Heath, K. Bowyer, R. Moore, "The digital database for screening mammography", Proceedings of the 5th international workshop on digital mammography; Medical Physics, p. 212–8, 2000.
- [95] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.
- [96] N. Sasi and V. Jayasree, "Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images", Engineering, vol. 05, no. 10, pp. 326-331, 2013.
- [97] D. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines", PeerJ, vol. 7, p. e6201, 2019.
- [98] D. Abdelhafiz, S. Nabavi, R. Ammar, and C. Yang, "Survey on deep convolutional neural networks in mammography", 2017 IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), 2017.
- [99] H. Li, D. Chen, W. Nailon, M. Davies and D. Laurenson, "A Deep Dual-path Network for Improved Mammogram Image Processing", ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2019.

- [100] P. Kaur and A. Kaur, "Review of Different Approaches in Mammography", International Journal of Advance Research, Ideas and Innovations in Technology, vol. 2, 2016.
- [101] S. Bandyopadhyay, "Pre-processing of Mammogram Images", International Journal of Engineering Science and Technology, vol. 2, 2010.
- [102] T. Velmurugan and M. Sukassini, "Noise removal using Morphology and Median filter Methods in Mammogram Images ER", 2016.
- [103] X. Ying, "An Overview of Overfitting and its Solutions", Journal of Physics: Conference Series, vol. 1168, p. 022022, 2019.
- [104] Q. Ji, J. Huang, W. He and Y. Sun, "Optimized Deep Convolutional Neural Networks for Identification of Macular Diseases from Optical Coherence Tomography Images", Algorithms, vol. 12, no. 3, p. 51, 2019.
- [105] B. Jimmy and P. Diederik, "Adam: A Method for Stochastic Optimization", conference paper at the 3rd International Conference for Learning Representations, 2014.
- [106] Sashank J. Reddi, Satyen Kale, Sanjiv Kumar. On the Convergence of Adam and Beyond. 2018.
- [107] Y. Bengio, I. Goodfellow and A. Courville, Deep learning. Massachusetts: MIT Press, 2017.

- [108] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning", arXiv:1811.03378, vol. 181103378, 2018.
- [109] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", Journal of Machine Learning Research, vol. 9, 2010.
- [110] S. Raschka, Python Machine Learning - Third Edition. [Place of publication not identified]: PACKT Publishing Limited, 2019.