JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

SCHOOL OF BIOMEDICAL ENGINEERING

BIOMEDICAL IMAGING PROGRAM

M.Sc. Thesis

Automatic Diagnosis of Breast Cancer from Histopathological Images Using Deep Learning Technique

By

Elbetel Taye

A thesis submitted to the School of Graduate Studies of Jimma Institute of Technology in partial fulfillment of the requirements for the Degree of Master of Science in Biomedical Engineering (Biomedical Imaging)

Advisor: Gizeaddis L. Simegn (Ph.D.)

Co-advisor: Mr. Abel W. (M.Sc.)

January, 2020

Jimma, Ethiopia.

Declaration

I declare this thesis research entitled "Automatic diagnosis of breast cancer from histopathological images using deep learning technique" is my original work and has not been presented for a degree in any other university.

Done By:

Ms. Elbetel Taye   _____   _____

                   Signature   Date

Approved By:

Dr. Gizeaddis L. Simegn   _____   _____

                   Signature   Date

Mr. Abel Worku   _____   _____

                   Signature   Date

Examiners:

_____   _____   _____

External Examiner name   Signature   Date

_____   _____   _____

Internal Examiner name   Signature   Date

_____   _____   _____

Chair name   Signature   Date

Acknowledgments

First of all, I would like to thank God for his limitless help throughout my life and for blessing me much more than I deserve. Next, I would like to acknowledge my advisors Dr. Gizeaddis Lamesgin and Mr. Abel Worku for their essential guidance on every aspect of the work.

I would also like to thank my clinical collaborators Dr. Solomon (pathologist at JUMC) and Dr. Tewodros (resident student in pathology department at JUMC) for their unlimited assistance and support throughout the research period. I would also want to give my great gratitude to Mr. Solomon Gebru, Mr. Kokeb Desse and Mr. Fetulak for their advice and assistance during the research time.

Finally, I would like to express my deepest gratefulness to my families and friends who supported me in every direction of my life and for their encouragement.

Abstract

Breast cancer is an invasive tumor that develops in the breast tissue. It is the most common cancer and also the leading cause of cancer mortality among females worldwide. Survival from breast cancer can be increased through advances in screening methods and early diagnosis. Clinical examination, screening using imaging modalities and pathological assessment (biopsy test) are common methods of breast cancer diagnosis. Among these, pathological assessment can be taken as a gold standard due to its potential in identifying the cancer type, grade and stage. However, current diagnosis using biopsy test is commonly done through visual inspection. This manual diagnosis is time consuming, tedious and subjective which may lead to misdiagnosis.

Different machine learning and deep learning methods have been proposed in the literature to automate the manual breast cancer diagnosis mechanism. To our knowledge, an integrated system that can classify breast cancer to its subtypes and identification of the cancer grade from biopsy tests has not been previously done. Identification of breast subtypes and grade is essential for understanding the biological characteristics and clinical behavior as well as for developing personalized treatments. Moreover, the results claimed from previously proposed automation techniques are less accurate and unreliable.

In this thesis, an automatic detection of breast cancer type, subtype and grade was proposed based on deep learning neural network model. The system was developed using python software. It was trained and validated using histopathological images acquired from 'break-his dataset', the 'zendo online dataset' and local data acquired from Jimma University Medical Center (JUMC), by using optikam PRO5 digital camera and optika microscope. All images have been pre-processed using CLAHE algorithm and histogram matching techniques, prior to feeding to the ResNet 50 pre-trained model. The developed system is capable of classifying breast cancer into binary classes (benign and malignant) and multi-classes (sub-types). Identification of cancer grade is done for cancers that are classified as malignant ductal carcinomas. Our test results showed that, the proposed method is 96.75%, 96.7%, 95.78%, and 93.86 % accurate for binary classification, benign sub-type classification, malignant sub-type classification, and grade identification, respectively.

Key words: Breast Cancer, Cancer Sub-type, Classification, Grade, Transfer Learning, ResNet.

Preface

Breast cancer is the most common, frequently occurring and fatal disease which is affecting women worldwide. It is clinically diagnosed using triple assessment technique (physical examination, medical image analysis and biopsy test). The gold standard in the diagnosis procedure is the biopsy test.

Biopsy test is the analysis of tissue or cell structure using images acquired from a microscope. It is clear that accurate detection and assessment of breast cancer in its early stages is crucial when it comes to reducing the mortality rate. However, accurate diagnosis of breast cancer is merely based on accurate analysis of the structures. The manual biopsy test method requires expert pathologists and the decision is subjective or is based on the knowledge and experience of the pathologist that may sometimes require two or more experts for reliable decision making.

Now a days, machine learning and deep learning techniques are applied in different medical image and signal analysis works in order to automate diagnosis systems and help physicians in getting support for making accurate decision. This thesis work mainly focuses on automating the diagnosis mechanism of breast cancer from histopathological images (tissue) so as to ease the work load of pathologists. This helps to reduce the mortality rate of women, due to breast cancer, by providing accurate results at early stage. It contributes: (1) an end-to-end automatic classification of histopathological image of breast cancer to its type, sub-types and grade from histopathological images using pre-trained model; (2) the system is developed to be magnification factor invariant, achieved good result with different magnification factors (including 40×, 100×, 200×, and 400×); and (3) the proposed pre-trained model has been evaluated and has shown better classification accuracy than the existing benchmark studies.

The first chapter of this thesis discusses the background of the study, clinical significance, existing solutions and their gaps, problem statement, objective of the study and the scope. The second chapter focuses on related works (literature review). The third chapter discusses the method proposed including the procedures followed, techniques and materials used. Results and discussion are discussed in chapter four. The fifth chapter concludes the main findings of this thesis. A sample source code used is provided in the appendix.

Table of Contents

List of Figures

List of Tables

Acronyms

AGHMC- Adama General Hospital and Medical College

AHRI- Armaeur Hansen Research Institute

AI – Artificial Intelligence

AUSH- Ayder University Specialized Hospital

BC - Breast Cancer

CDF- Cumulative distribution function

CNN – Convolutional Neural Network

CT – Computed Tomography

DCIS –Ductal Carcinoma In situ

ER - Estrogen-Receptor

GGH- Gambi General Hospital

GUH- Gondar University Hospital

H & E – Hematoxiline and Eosin

IDC – Invasive Ductal Carcinoma

IDE - Integrated Development Environment

IHC-Immuno-Histo Chemistry

JUSH- Jimma University Specialized Hospital

MRI – Magnetic Resonance Image

ResNet - Residual Network

SNNPR- Southern Nations Nationalities People region

TASH- Tikur Anbessa Specialized Hospital

VGG - Visual Geometry Group

WHO- World Health Organization

Chapter One

Introduction

1.1.   Background of the study

According to the world health organization (WHO), global cancer burden was estimated to have risen to 18.1 million new cases and 9.6 million deaths in 2018 [1]. This is due to several factors, including population growth, ageing and the changing incidence of certain causes of cancer related to social and economic development.

Among all existing cancer types, breast cancer is the most prevalent and fatal disease which is ranked 2nd after lung disease worldwide [1]. According to the statistical data revealed by the Addis Ababa city cancer Registry Quadrennial report of 2012-2015 investigation, females accounted for two-third of the cancer burden, largely due to breast cancer, which is the leading type of cancer highly affecting the women population in the city followed by cervical cancer [2]. As shown in Figure 1-1, breast cancer accounts for more than 30% from the total incidence of cancer in females and above 4% in males. This makes it the number one cancer type among women in Addis Ababa and the statistics can be a representative of the rate of breast cancer presence in the country. Figure 1-1 shows the statistical data of most common cancer types among males and females in Ethiopia.
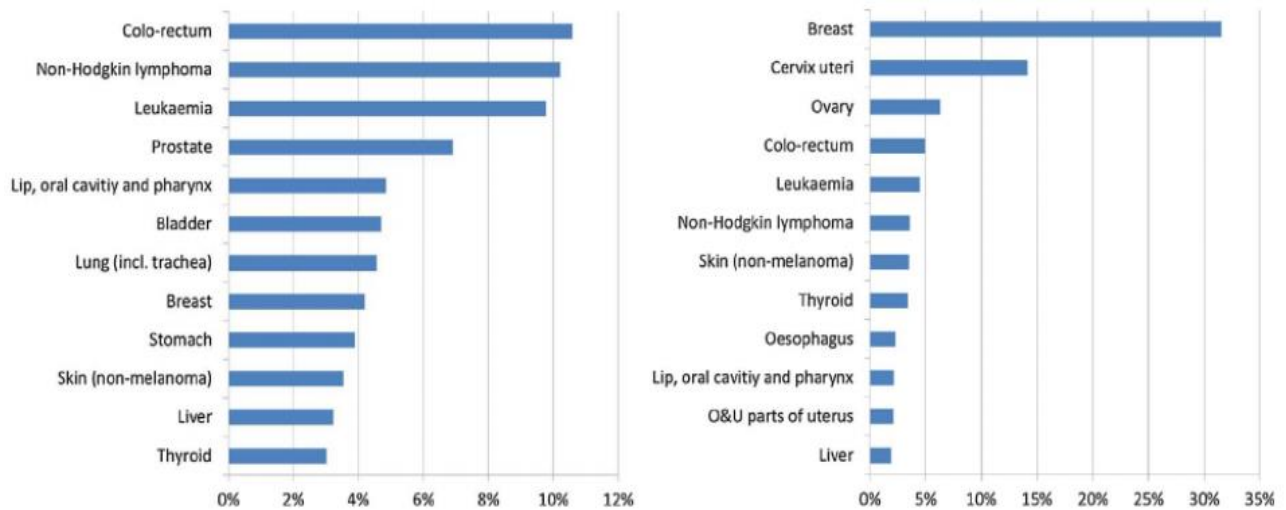
Figure 1-1: Statistical representation of most common cancer in Ethiopia a) males and b) females [2].

1.2. Breast Anatomy

The breast is very complex organ which undergoes many changes than any other part of the human body, starting from birth, puberty, pregnancy, breast feeding and menopause. It is an endocrine gland placed on the front of the chest, consisting of alveoli, coated by cells that have the property of safeguarding milk under the influence of hormones. It is made up of the glandular tissues, connective tissue, and adipose tissue which determine the size, shape and texture of the organ [3]. At the apex of the breast is the mammary areola, which is a pigmented skin having a modified sebaceous glands that secretes hormone and help the nipple to be soft and elastic.

In every female breast, the glandular tissue is imbedded with fat and divided into 15-20 glands called lobes, where milk is produced. And these lobes are connected to the nipple by 6-8 tubes called ducts which carry milk to the nipple [3]. The male breast comprises small ducts without alveoli and is supported by fibrous tissues and fat. Figure 1-2 shows the schematic diagram of the anatomy of the breast.
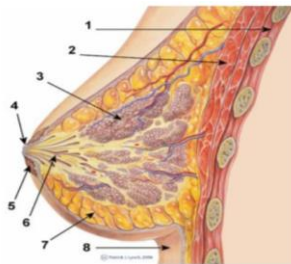
Figure 1-2: Schematic diagram of the anatomy of the breast [4]. (1- Chest wall, 2-Pectoralis major, 3-Lobules, 4-Nipple, 5-Areola, 6-Mammary ducts, 7-Fatty tissue, 8-Skin)

1.3. Breast cancer and its types

Breast cancer (BC) is an uncontrolled growth of breast cells. The term refers to a malignant tumor that has developed from cells in the breast. Mostly, it begins in the cells of the milk producing glands called the lobules, or the ducts which are the passages of milk from the lobules to the nipples [4].

Like any other organ in the body, the breast consists billions of microscopic cells. These cells usually multiply in an organized fashion, healthy cells continue to divide and replace the ones that have died. But sometimes these cells grow quite rapidly in uncontrolled and abnormal way.

BC is a public health problem as it is the second commonest cancer with increasing incidence in the world. It is not just one disease, but rather includes different types of cancer that occurs in the breast. Some are, benign cancer types such as: adenosis, fibroadenoma, Phyllodes tumor and tubularadenoma tumor and malignant cancer types like: invasive ductal carcinoma, invasive lobular carcinoma, mucinous carcinoma and papillary carcinoma [5].

1.3.1. Adenosis breast tumor

The term "adenosis" is used to describe a common benign proliferation of glandular structures that generally maintains normally arranged epithelial and myoepithelial components. Adenosis is a lobular-based lesion, with variably increased numbers of acinar structures [6]. It mostly occurs on women with 30-40 years of age. Adenosis is characterized by a lobulocentric, loosely-structured proliferation of acinar or tubular structures, composed of an epithelial and myoepithelial cell layer and surrounded by a basement membrane [6]. Figure 1-3 presents histopathological slide image for Adenosis tumor type.
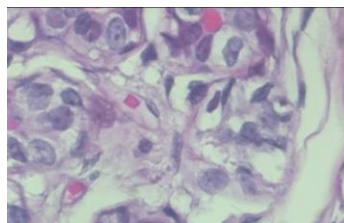
Figure 1-3 : Histopathological slide image for adenosis tumor

1.3.2. Fibroadenoma breast tumor

It is grouped under fibroepithelial tumor type and it is common benign biphasic tumor. It consists of a constrained breast neoplasm arising from the terminal duct lobular unit and containing a proliferation of both epithelial and stromal elements. This disease occurs most often in women of reproduction age, especially those aged < 30 years, although it may happen at any age [6]. Figure 1-4 shows histopathological slide image for fibroadenoma tumor type.
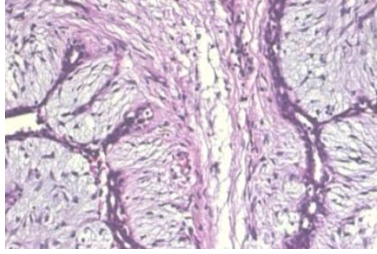
Figure 1-4: Histopathological image of fibroadenoma tumor

1.3.3. Phyllodes breast tumor

It is a group of circumscribed fibroepithelial neoplasms, histologically similar to intracanalicular fibroadenomas, characterized by a double-layered epithelial component arranged in clefts surrounded by a hypercellular stromal mesenchymal component which in combination elaborate leaf-like structures [6]. Figure 1-5 shows histopathological slide image for phyllodes tumor type.
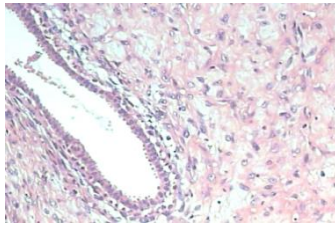
Figure 1-5: Histopathological image of phyllodes tumor

1.3.4. Tubular adenoma breast tumor

The lesions are well-circumscribed and composed entirely of small round tubules, lined by uniform epithelial cells with surrounding myoepithelial cells and separated by little intervening stroma. It mostly occurs in young females [6]. Figure 1-6 shows histopathological slide image for tubular Adenoma tumor type.
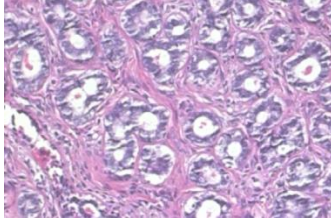
Figure 1-6: Histopathological image of tubularadenoma tumor

1.3.5. Papillary carcinoma

It is one of malignant breast cancer types. During histopathological tests, microscopic examination shows ducts and terminal-duct lobular units filled with slender, branching fibrovascular stalks covered by a single cell population of neoplastic epithelial cells. The neoplastic cells can be arranged in one or several layers of columnar cells overlying the stalks in an orderly manner and with a deceptively bland appearance, also known as the stratified spindle cell pattern [6]. Figure 1-7 displays histopathological slide image for papillary tumor type.
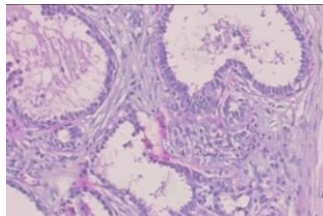
Figure 1-7: Histopathological slide image of papillary carcinoma

1.3.6. Ductal carcinoma

A neoplastic proliferation of epithelial cells confined to the mammary ductal-lobular system and characterized by subtle to marked cytological atypia and an inherent but not necessarily obligate tendency for progression to invasive breast cancer [6]. It accounts for 80% of the total occurrence of BC [7]. Figure 1-8 presents histopathological slide image for Ductal carcinoma type.
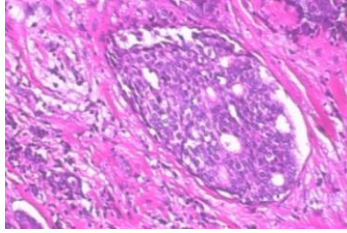
Figure 1-8: Histopathological image of ductal carcinoma

1.3.7. Mucinous carcinoma

Mucinous carcinoma is characterized by nests of cells floating in lakes of mucin partitioned by delicate fibrous septae containing capillary blood vessels. The cell clusters are variable in size and shape; sometimes with a tubular arrangement. It often occurs on patients aged 55 years and above [6]. Figure 1-9 indicates histopathological slide image for mucinous Carcinoma type.
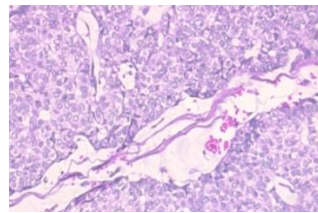
Figure 1-9: Histopathological image of mucinous carcinoma

1.3.8. Lobular carcinoma

The definitive pattern of invasive lobular carcinoma (ILC) is characterized by a proliferation of small cells, which lack cohesion and appear individually dispersed through a fibrous connective tissue or arranged in single file linear cords that invade the stroma. Lobular carcinoma represents 5-15% of invasive breast tumors [6]. Figure 1-10 presents histopathological slide image for lobular carcinoma type.
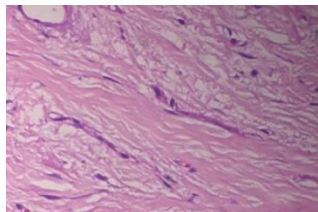
Figure 1-10: Histopathological image of lobular carcinoma

1.4.   Breast cancer diagnosis

Triple assessment test, that includes clinical examination, imaging and biopsy test, is the most common technique that is used to identify the cancer type, grade and stage of breast cancer before performing a treatment plan.

The histopathological image (biopsy test) is the gold standard for identifying breast cancer compared to other medical imaging techniques like: mammography, magnetic resonance imaging (MRI), ultrasound and computed tomography (CT) [8]. The imaging modalities help to screen the tissues, but for further diagnosis and confirmation, biopsy test is always required.

Biopsy test is the examination of disorder indicatives on tissue samples prepared and fixated on glass slides using light microscope. To diagnose breast cancer using a biopsy test, first samples of breast tissues are removed from the breast. Then from this sample, histology slides are produced by following seven slide preparation procedures indicated in Figure 1-11.
These are:

1. Fixation and Dehydration:
    - Fixation: apply a chemical solution such as formalin to preserve the proteins and inhibit the action of enzymes.
    - Dehydration: the tissue will be dehydrated using alcohol. And finally the alcohol will be removed.
2. Tissue imbedding:
    - The tissue will be placed in a small frame and liquid wax will be added on it.
    - Then it will be exposed to low temperature in order to harden the mass.
3. Sectioning:
    - By using a microtome the mass will be sectioned into extremely thin section (1 - 10μm[4] thickness).
4. Water base:
    - The sectioned mass will be put on warm water so as to prevent folding and then from the water it will be placed on to a glass slide.
5. Apply mild temperature (oven):
    - The slide will be exposed to mild temperature so as to remove the wax.

6. Dyes and Chemicals:
   - In order to differentiate the cellular structures and other tissue components, dye chemicals are applied to the sectioned mass.
7. Stained slides:
   - Finally, stained slides are acquired.

The slides are commonly stained with hematoxylin and eosin (H&E). Hematoxylin is a basic dye that helps to stain basophilic structures including the cell nucleus (that contains DNA and nucleoprotein), and organelles (that contain RNA), such as ribosomes and the rough endoplasmic reticulum. On the other hand, eosin is an acidic dye that typically stains acidophilic structures, which include the cytoplasm, cell walls, and extracellular fibers [5]. Finally the cytoplasm will be presented as pinkish and nuclei will be presented as either blue or purple. Once the slide is prepared, pathologists make visual examination through the microscope and then tell the biopsy result. Figure 1-11 illustrates a sample staining procedure.

Figure 1-11: Tissue sample staining procedure

A study in 2015 revealed that the available histopathology laboratories with full set up in Ethiopia at that time were 11 [9]. Among these, 5 of the histopathology laboratories namely (Tikur Anbessa Specialized Hospital, Armaeur Hansen Research Institute (AHRI), Armed Forces

General Hospital, Kadisco General Hospital and St. Paulos General Hospital) are found in Addis Ababa. The rest histopathological laboratories are found in Tigray region (Ayder University Specialized Hospital (AUSH)), in Southern Nations Nationalities People region (SNNPR) (Hawassa referral Hospital), in Oromia region (Jimma University Specialized Hospital (JUSH) and Adama General Hospital and Medical College (AGHMC)) and in Amhara region (Gondar University Hospital (GUH) and Gambi General Hospital (GGH) at Bahrdar). These hospitals are equipped with histopathology laboratory set up and have a potential to give full cancer diagnosis service.

1.5. Breast cancer grade identification method

Breast cancer grading is a way in which healthcare professionals describe how far the disease has spread, and how fast it may grow or how aggressive it is. Identifying BC's grade help doctors to have the best treatment options [10]. Breast tumors are graded between 1 and 3. Grade 1 indicates that the cancer cells look small and uniform like normal cells and are usually slow-growing compared to other grades of breast cancer (Figure 1-12a). Grade 2 is characterized by slightly bigger cells, varying in shape and growing faster than normal cells (Figure 1-12b). Likewise, Grade 3 cancer cells look different (Figure 1-12c), and are usually faster-growing than normal cells [10].
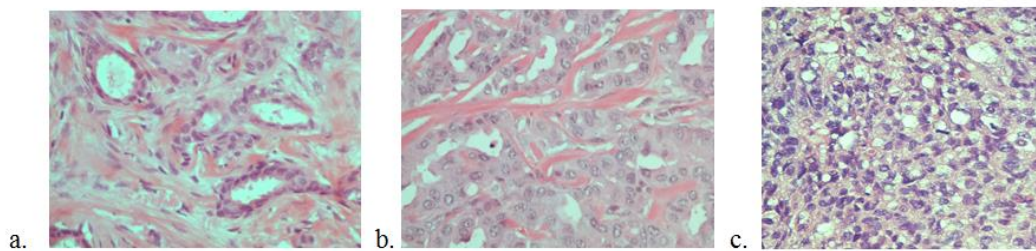
Figure 1-12: a) indicates grade1 IDC, b) indicates grade2 IDC, c) indicates grade3 IDC

Breast cancer grade is usually assessed after breast tumor is removed and tested in a laboratory using microscope. The grade of BC is mostly diagnosed using the Nottingham system. This system grades breast cancer based on the following features [11]:

- Tubule formation: how much of the tumor tissue has normal breast (milk) duct structure
- Nuclear grade: an evaluation of the size and shape of the nucleus in the tumor cells

- Mitotic rate: how many dividing cells are present, which is a measure of how fast the tumor cells are growing and dividing

1.6. Drawbacks of manual biopsy test

Because biopsy test is done through visual inspection, it is only as reliable as the physician's performance. Thus, multi-classification of breast cancer using histopathological images is a challenging task for the reason that: it is highly subjective, and the result's accuracy depends on the observer's knowledge and experience, it is a tedious task and very time consuming. Moreover, due to lack of experienced pathologists in most developing countries like Ethiopia, a pathologist is expected to examine multiple kinds of biopsy samples with different types of problems per day. The variety and large amount of data analyzed by the pathologist and complexity of the images make the task time consuming, difficult and tedious which may lead to misdiagnosis.

Biopsy test misdiagnosis can be either over interpreted or under interpreted. Over interpreted means that women without cancer are exposed to potentially harmful treatments and unnecessary expenses. On the other hand, inaccurate interpretation (under interpretation) of biopsy result could prevent women from getting the treatment early causing the cancer to grow more to invasive stage.

Every year millions of women undergo breast cancer biopsies. Pathologists take these tissue biopsies and, categorize it to either normal, atypia, ductal carcinoma in situ (DCIS) or invasive carcinoma types. Studies show that differentiating precancerous cells from DCIS is very challenging for pathologists [12, 13, 14]. Figure 1-13 shows the high variability in the interpretation of the breast tissue with resulting concern in patient harm. Each 100-slide grid shows the predicted number of cases over interpreted or under interpreted relative to the reference consensus panel diagnosis or verified [13].
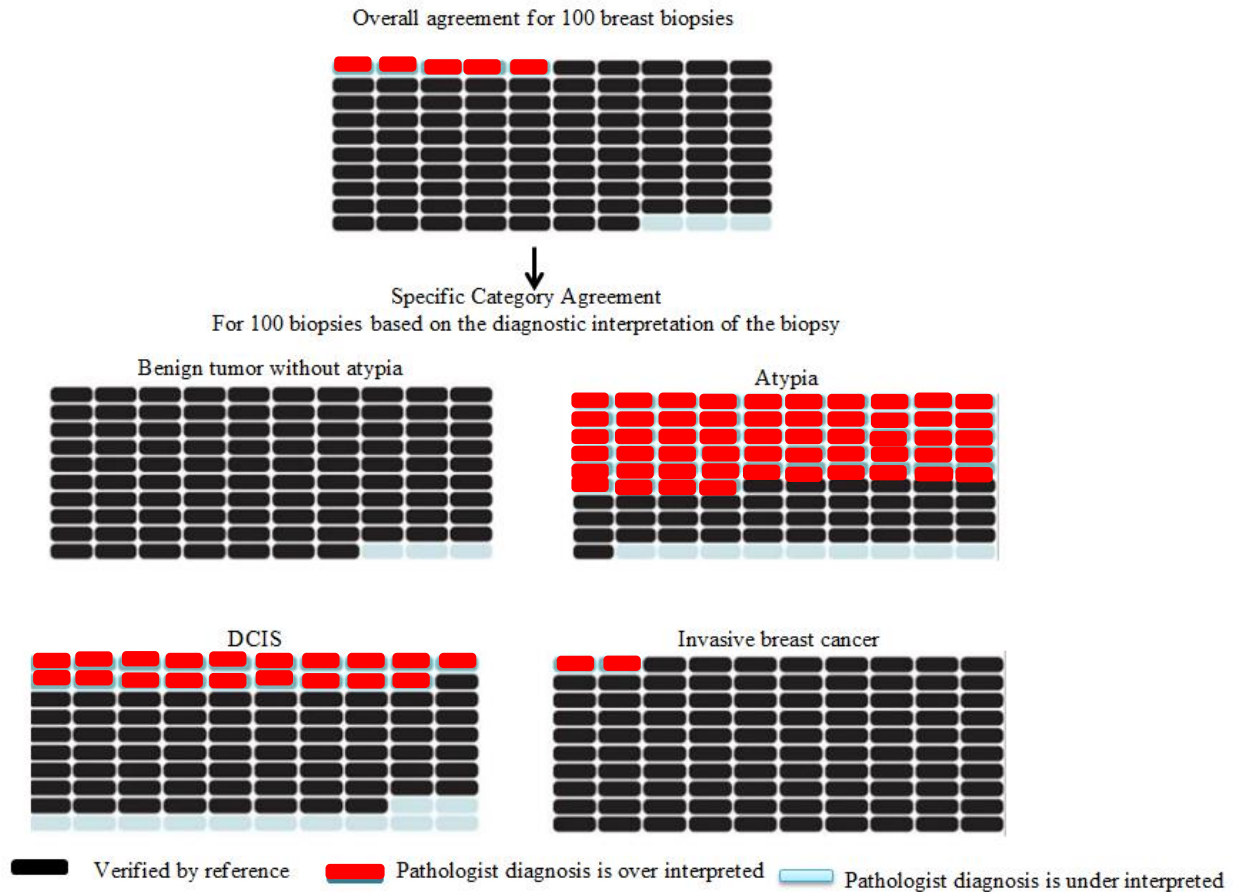
Figure 1-13: predicted outcomes per 100 breast biopsies, overall and by diagnostic category [13].

The decision for a best therapeutic strategy of breast cancer is based upon advanced multi-class classification of the disease type. That is because, doctors who know the subclasses of breast cancer accurately can control the metastasis of tumor cells early and make important therapeutic schedules [15]. However, multi-class classification using the manual biopsy test technique could be unreliable [14].

1.7. Computer aided diagnosis of breast cancer

Computer aided diagnosis (CAD) systems on breast cancer histology can help increase diagnosis accuracy and reduce subjective variability of breast cancer type classification and grading apart from reducing the work load of pathologists. Such automated systems are developed either using machine learning or deep learning techniques. These techniques are generally included under

artificial intelligence systems. Where, an artificial intelligence is an umbrella term consisting of both machine learning and deep learning techniques.

1.7.1. Machine learning

Machine learning is a technique that is used to automate systems in order to ease work load of humans. It is used to automatically solve classification problems by computing and encoding hand crafted features of an image (i.e., shape, texture, color, etc.). In this technique, the image pixels are used as an input to the feature extraction process in order to get the feature vectors through quantifying the image content [16]. The feature vectors that resulted from feature extraction serves as an input to the machine learning models.

1.7.2. Deep learning

Deep learning is a subset of machine learning that has networks capable of learning from data. It differs from machine learning in a way that an individual is not expected to select and tell the features for the system rather the algorithm learns on its own which features are best for classifying the data. Figure 1-14 shows the schematic representation of terms of AI, machine learning and deep learning technique.
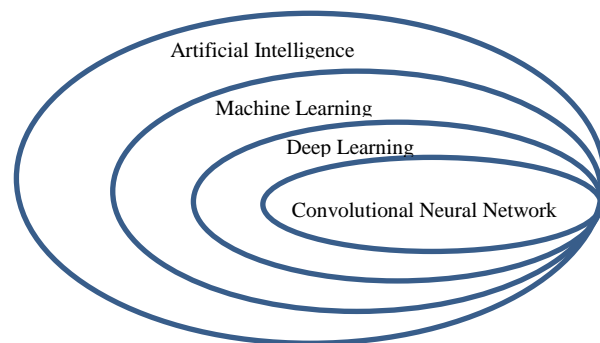
Figure 1-14: Schematic representation of the hierarchy of terms of artificial intelligence, machine learning and deep learning techniques [17].

Deep learning algorithms use a hierarchal learning method. That is, information from the first layer becomes input for the next layer. That is, primarily, only edge-like regions are detected in the lower level layers of the network. Then the edge regions are used to define corners and contours. Merging corners and contours can lead to identifying object parts in the next layer [16].

In deep learning models, a series of hidden layers are used to extract features from the input image. Finally, output layer is used to classify the image and obtain the output class label. Figure 1-15 shows the graphical representation of connection of layers in a neural network.

Input layer

Output layer

Hidden layer
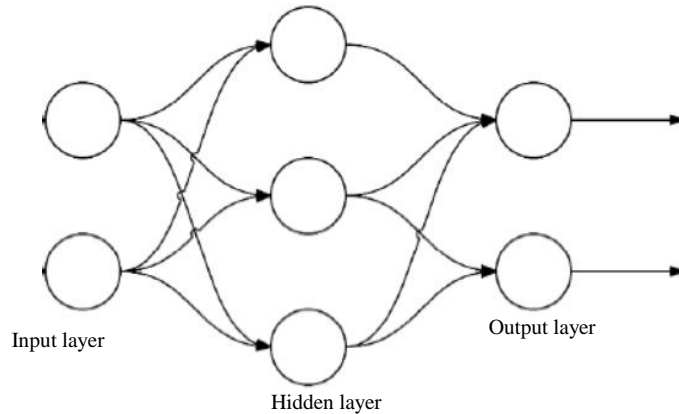
Figure 1-15: Schematic diagram of connection of layers in a neural network

Deep learning techniques include convolutional neural networks (CNN) which are highly applicable in medical image classification problems. Besides, there are also advanced network architectures that have an impact on medical image processing like Google inception network, ResNets and recurrent neural networks including transfer learning technique [18].

1.7.3. Transfer Learning

Transfer learning is a term that refers to passing weight values of a trained neural network to another new neural network, so that building and training a network from scratch will be avoided and thus the new network can know the learned features. This can be achieved by using pre-trained models with less computational power and time. Moreover, pre-trained networks are quite competitive against training from scratch [19, 20].

Pre-trained models can either be used as a feature extractor or as a classifier by directly freezing the whole convolutional base layer and optimizing the last fully connected layer or the model can be used as a classifier by fine-tuning it.

Currently there are different pre-trained models which can be used for image classification, prediction, feature extraction and fine-tuning (with weights trained on ImageNet). It includes

VGG16, VGG19, Xception, ResNet, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet, and NASNET[21].

1.7.4. Hyper-parameters and parameters of deep learning models

When training a neural network, there are two key things that need to be considered. These are Hyper-parameters and parameters. Hyper-parameters are parameters which are manually set before training a system to get a finest result and once they are set they cannot be changed during training. A typical set of hyper-parameters for neural network includes the number of hidden layers, initialization weight values, learning rate, its decay rate, dropouts and Activation functions.

Parameters are those which would be learned by the machine like weights and biases. These values differ for each experiment and epoch. Moreover, they are highly dependent on the type of data and the task at hand. Basically, parameters are the ones that the "model" uses to make predictions.

1.8. Statement of the problem

According to the WHO report, breast cancer is the most common cancer with high morbidity and mortality rate among women worldwide. It is also ranked as the first cancer disease affecting the women population in Ethiopia. Unfortunately, there is still a shortage of pathologists in most developing countries including Ethiopia, limiting many breast cancer patients from being diagnosed at the early stage. Due to this, pathologists are expected to examine a large number of biopsy samples per day making them exhausted. Moreover, the complexity of the cancer cells and their subjective decision, which is dependent on expert's performance, may lead to misdiagnosis.

Computer aided diagnosis systems on breast cancer histology can help increase diagnosis accuracy and reduce subjective variability of breast cancer type classification and grading apart from reducing the work load of pathologists. Many researches have been proposed to automate the diagnosis procedure of breast cancer from histopathological image using different machine learning and deep learning techniques. However, to our knowledge, an integrated system which classifies of breast cancer to its subtypes and identification of the cancer grade from biopsy tests,

which is essential to understand the biological characteristics and clinical behavior, has not been previously done in the literature. Moreover, the results claimed from previously proposed automation techniques are less accurate making them unreliable.

Hence, a sensitive, accurate and automatic system is required to determine the type and detail subtypes of breast cancers with the grade from histopathological data (biopsy sample). This help to reduce the heavy workloads of pathologists and avoid misdiagnosis.

1.9. Objectives of the study

1.9.1. General objective

The general objective of this study is to develop an automatic diagnosis system for breast cancer from histopathological image using deep learning technique.

1.9.2. Specific objectives

The specific objectives of this study are to:

- Determine the specific type of breast cancer.
- Determine the grade of invasive ductal carcinoma (IDC).
- Measure the performance of the system.

1.10. Significance of the study

Some of the importance of using automatic detection technique for breast cancer diagnosis are, but not limited to:

- To reduce diagnosis time.
- To determine the type of breast cancer.
- To determine the grade of the most common BC type (IDC)
- To enable on time and proper treatment schedule.
- To enable good decision regarding the treatment type.
- To prevent misdiagnosis due to over fatigue of pathologists.

1.11. Scope

This study is limited to classifying breast cancer into eight subtypes; which are: adenosis, fibroadenoma, phyllodes tumor, tubular adenoma tumor, invasive ductal carcinoma, invasive lobular carcinoma, mucinous carcinoma and papillary carcinoma, and determine the grade of invasive ductal carcinoma.

Chapter Two

Automatic Breast Cancer Diagnosis methods

2.1. Automatic breast cancer type classification

There has been a lot done to fight against breast cancer, though, judgment about the status of a disease from histopathological data is subjective and the result might vary from specialist to specialist. To improve this, computer aided diagnosis systems play an important role.

Recent state of the art deep neural network (DNN) models have been used in many works for breast cancer diagnosis [15, 22 – 38]. In order to classify breast tissues biopsy image as normal, benign, malignant and invasive carcinoma deep CNN like patch level voting model and merging model was used and an accuracy of 87.5 % was claimed [22]. Likewise, DCNN and gradient boosted tree method was used to classify breast cancer into the basic 4 types [23] and an accuracy of $93.8\pm2.3\%$ and $87.2 \pm 2.6\%$ was reported. CNN as a feature extractor and support vector machine as a classifier has been implemented in another study [24] by retrieving image information at different scales, including both nuclei and overall tissue organization an accuracy of 77.8% for the four classes and 83.3% for carcinoma (in situ and invasive) or non-carcinoma (normal and benign) was claimed. Inception-v3 convolutional neural network has also been adapted and fine-tuned to make patch classification and majority voting was considered [25] for the whole slide classification yielding an accuracy of 85% over the four classes and 93% for non-cancer (normal and benign) versus malignant (in situ or invasive carcinoma) was reported.

The above studies were successful (with reasonable accuracy) in classifying breast cancer into its four main classes. But, the need for classifying breast tissue histology images into further breast cancer sub-classes has led many researchers to do further studies using deep learning techniques. Classification of breast cancer into eight subtypes (adenosis tumor, fibroadenoma tumor, phyllodes tumor, tubularadenoma tumor, ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma) using deep neural network has been proposed in many litreatures [15, 26 - 31]. Among these, Convolutional Neural Network (CNN), a Long-Short-Term-Memory (LSTM), and a combination of CNN and LSTM [27] were used for extracting features from breast cancer histology images using the structural and statistical information of the images, Soft-max and Support Vector Machine (SVM) were used for doing the classification

task. The best accuracy value achieved from the experiment was 91.00% on the 200x magnified dataset, while the best precision value 96.00% was achieved on the 40x dataset, and the best *F-Measure* value is achieved on both the 40x and 100x datasets. Similarly, CaffeNet pre-trained model [28] was used and an accuracy of 81- 85 % was achieved. Furthermore, other pre-trained models such as VGG16, VGG19, and ResNet50 have been used to do the classification task [29]. Finally, pre-trained VGG16 with logistic regression classifier yielded the best performance with 92.60% accuracy, 95.65% area under ROC curve (AUC), and 95.95% accuracy precision score. Inception_V3 and Inception_ResNet_V2 architectures have also been adapted [30] and were used as a feature extractor to solve binary and multi-class breast cancer histopathological image classification task. An accuracy of 96 - 99 % was claimed depending on the magnification factor of the images. Additionally by using Inception recurrent residual convolutional neural network (IRRCNN) [31] an anticipated accuracy of 97.57±0.89% and 96.84±1.13% for multi- class BC classification at the image level and patient level respectively was reported. Another study [26] was done to classify different cancer types (breast cancer, bladder cancer, lung cancer and lymphoma) and breast cancer into the eight subtypes by applying pre-trained inception networks (inception v1 – inception v4) and ResNet models (ResNet 50, ResNet 101, ResNet 152). The models were trained for 3000 epochs and finally, their classification performance was measured. As a result, an accuracy ranging from 69.6% to 96.5 % was achieved for benign class classification and 56.5 % - 96.4% was achieved for malignant class classification by fine tuning all layers of the models. Specifically, using ResNet 50 model for benign and malignant subtype class classification, an accuracy of 94.8% and 94.6% was achieved respectively. Furthermore, 93.2% accuracy was achieved using Class Structure based Deep convolutional Neural Network (CSDNN) to classify breast cancer cells into sub classes [15].

Most of the aforementioned studies resulted better classification accuracies, however, they still lack further classification of biopsy images into their respective grade level, which is a very important characteristics in determining the treatment plans of breast cancer. Furthermore, the models used to determine the sub types of breast cancer are complex and computationally expensive [28, 30, 31], and their performance is limited to the magnification factor of the given images [15, 28, 31].

To overcome the problem of stage determination, DCNN was used by Cruz-Roa et al. [32]. However, some subtypes of invasive breast cancers, such as mucinous carcinoma were left undetected. In addition, some in situ breast cancer regions, which are different from invasive cancers, were incorrectly classified as invasive breast cancer.

2.2. Automatic breast cancer grading

The grade of BC is mostly diagnosed using the Nottingham system [11]. In this system, identifying the total number of mitotic cells present in a given histology slide image, the tubule formation or the tumor tissue structure and the nuclear grade or the size and shape of the nucleus in the tumor cell is mandatory for determining the grade of breast cancer. The number of mitotic cells per tissue area gives an important aggressiveness indication of invasive breast carcinoma. But, counting the mitotic cells and decision making on the grade of the disease always results in low agreement among different pathologists.

Application of DNN is not limited to classifying BC types. Deep learning and machine learning approaches have also been employed to automatically detect the grade of invasive BC [33 - 38]. A pleomorphic and tubular formation pattern from breast cell histopathology images were used to identify the grade of a histology image and an accuracy of 88%, sensitivity of 85% and 80% specificity have been claimed using k-means clustering method for image segmentation and Gray level Co-occurrence Matrix (GLCM) for feature extraction [33] with four features (i.e. angular second moment, contrast feature, entropy feature, and variance feature).

Even though computer-aided detection methods have been developed recently, they suffer from high FN (false negative) and FP (false positive) results in mitotic cell detection task. To solve this, a two stage detection method with multi-scale and similarity learning convents (MSSN) was used [34]. At first results, large amount of possible candidates were generated using the FNR-model to reduce FN. Then based on the obtained candidates, a similarity prediction model was performed on the obtained candidates for the final detection to reduce FP values. Finally, good recall result was achieved.

The total tumor cell population was determined by diagnosing the status of human epidermal growth factor receptor 2 (HER2), using DCNN [35]. Moreover, deep neural network has been used [36] to determine cancer status as low, intermediate or high tumor grade as well the

estrogen receptor (ER) status, Basal like or non-basal like and ductal or lobular cancer type were determined in order to tell the status of the disease. Finally, accuracy of 82%, 84%, 77%, 94%, and 75% respectively were achieved. Similarly, deep cascaded convolutional neural network was used to detect mitotic cells [37]. This method has achieved a result of 0.460 precision, 0.507 recall, and 0.482 f1-score. The encoding of histopathological images as Vector of Locally Aggregated Descriptors (VLAD) represented on the Grassmann manifold [38] was the other method used to perform automatic grading of invasive breast carcinoma. That is by considering each image as a set of multidimensional spatially-evolving signals that can be efficiently modeled through a higher-order linear dynamical systems analysis. Then, each H&E stained breast cancer histological images were represented as a cloud of points on the Grassmann manifold, while a vector representation approach was applied aiming to aggregate the Grassmannian points based on a locality criterion on the manifold. Finally, the performance of the method was evaluated using the zendo data set and 95.8 % accuracy was achieved.

Aforementioned methods have tried to solve the problem of automatically grading invasive breast carcinoma. However, the accuracy reported makes the techniques unreliable [15, 27, 28]. Moreover, in a study conducted by Chen et al. [37] the analyzed data was microarray data, which is less applicable. As a result, an automatic end to end (BC subtype and grade classifier) BC classifier having better accuracy and precision is required to be done using histopathological images.

Table 2-1 shows the summary of the most recent studies on automatic breast cancer classification.

Table 2-1: Summary of recent studies

Cite	Method	Class	Result	Gap
Han et al. (2017) [15]	Class Structure based deep convolutional neural network (CSDNN)	Eight subtypes	93.2%	- Model's result is magnification factor dependent. - Focuses only on BC subtype classification.
Nahid et al. (2018) [27]	*K*-Means and Mean-Shift clustering as a feature extract, CNN, LSTM structure and both CNN and LSTM structures together were used for data analysis.	Eight subtypes	Accuracy value of 91.00% was achieved on 200x dataset, Precision value 96.00% was achieved on 40x dataset, and *F*-Measure value was achieved on both the 40x and 100x datasets.	-Low accuracy -Result is magnification factor dependent -doesn't classify the grade level
Spanhol et al. (2017) [28]	CaffeNet model	Eight subtypes	81- 85 % accuracy was achieved	-Low accuracy -Result is magnification factor dependent -doesn't classify the grade level
Shallu et al.	VGG16,	Eight	VGG16 with logistic	-Low accuracy

(2018) [29]	VGG19, and ResNet50	subtypes	regression classifier yielded the best performance with 92.60% accuracy, 95.65% area under ROC curve (AUC), and 95.95% accuracy 2precision score.	- doesn't classify the grade level
Motlagh et el. (2018) [26]	ResNetV1- 50	Eight subtypes	94.8% for benign class and 94.6% for malignant	- Used high computational power. i.e. each classification task was done for 3000-6000 epochs. - doesn't classify the grade level
Dimitropoulos et al. (2017) [38]	Grassmannian analysis. VLAD encoding Image patch	Eight subtypes	95.8% for grade identification 91.38% with the Break-His dataset	- Result is magnification factor dependent - Less accuracy - The system is not integrated.

Chapter Three

Diagnosis of Breast Cancer from Histopathological Tumor Images Using Deep Learning Technique

3.1. Introduction

This chapter presents the proposed scheme for diagnosis of breast cancer from histopathological images using deep learning technique. It uses one of the "ImageNet" pre-trained models called the ResNet 50 model. The developed system is fine-tuned and mainly trained to classify the eight sub-types of breast cancer and identify the grade of malignant cancer types. Data has been collected from two different publicly available online datasets; the 'break-his' dataset [5] and the 'zenodo' dataset [38], for the breast cancer sub type classification and grade class classification respectively. The overall methodology of the system can be summarized as indicated in Figure 3-1.

Input Image → Preprocess Image → ResNet 50 model →

Benign class:
- Adenosis Tumor
- Fibroadenoma Tumor
- Phyllodus Tumor
- Tubularadenoma Tumor

Malignant class:
- Lobular Carcinoma
- Mucinous Carcinoma
- Papillary Carcinoma
- Ductal Carcinoma
  - Grade 1
  - Grade 2
  - Grade 3
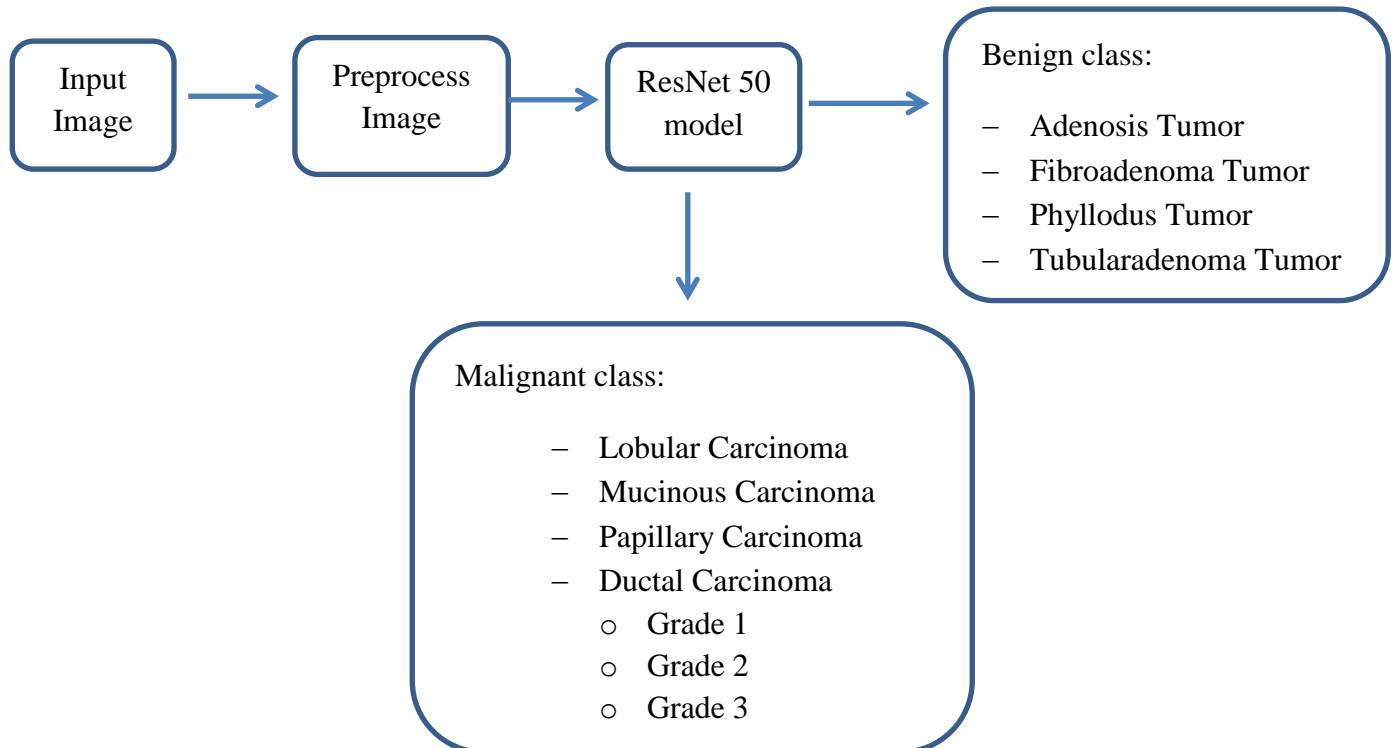
Figure 3-1: Block diagram of the system

3.2. Dataset

For this research, both local histopathological images and images from online datasets were used for training, validation and testing of the system. This helps the system to learn different features from variety of images.

There are two different online datasets which were used during the study. The first is the break-his dataset [5], which includes 7909 breast cancer histology raw images taken from 82 different patients. These images are labeled as benign tumors and malignant tumors. In this dataset the images were taken using different magnification powers (40X, 100X, 200X, and 400X) and are thus labeled accordingly. Moreover, the benign and malignant classes further include eight different subtypes. Adenosis, fibroadenoma, phyllodes and tubular adenoma are categorized under the benign class, whereas, ductal carcinoma, lobular carcinoma, papillary and mucinous carcinoma are categorized under the malignant class. All of the images were captured using Olympus BX-50 microscope and Samsung digital camera SCC-131A. All the images were saved in 3-channel Red-Green-Blue (RGB), with an 8-bit depth in each channel, using Portable Network Graphics (PNG) format, with a dimension of 700 x 460 pixels. Table 3-1 shows the type and number of data acquired from the break- his dataset.

Table 3-1: Types and number of histopathological images collected from Break-his dataset

Magnification	Benign				Malignant			
	Adenosis	Fibroadenoma	phyllodes	Tubular adenoma	Ductal	Lobular	Mucinous	Papillary
40 X	114	253	109	149	864	156	205	145
100 X	113	260	121	150	903	170	222	142
200 X	111	264	108	140	896	163	196	135
400 X	106	237	115	130	788	137	169	138
Total	**444**	**1014**	**453**	**569**	**3451**	**626**	**792**	**560**

The second dataset used in this study was from the zenodo repository [38]. The dataset consists of 300 images (Grade 1: 107, Grade 2: 102 and Grade 3: 91 images) of resolution 1280 x 960 acquired from 21 different patients with invasive ductal carcinoma. The image frames were from regions afflicted by tumor growth captured through a Nikon digital camera attached to a compound microscope with x40 magnification objective lens [38]. Table 3-2 shows the number and types of data acquired from the zendo dataset.

Table 3-2: Table representation of the Zendo dataset

Grade Type	Invasive ductal carcinoma	magnification
Grade 1	107	40 X
Grade 2	102	40 X
Grade 3	91	40 X

Additionally, a total of 110 local images were collected from Jimma University Medical Center (JUMC), Jimma, Ethiopia. These images are breast cancer histopathological images which are stained using H&E staining method. They were collected using the Optika vision camera mounted on a light microscope. The images were taken in different magnification powers (40X, 200X, and 400X) and they have a resolution of 2592 x 1936. Figure 3-2 shows the materials and procedure used to acquire the images. Table 3-3 shows the number of local data acquired from JUMC.
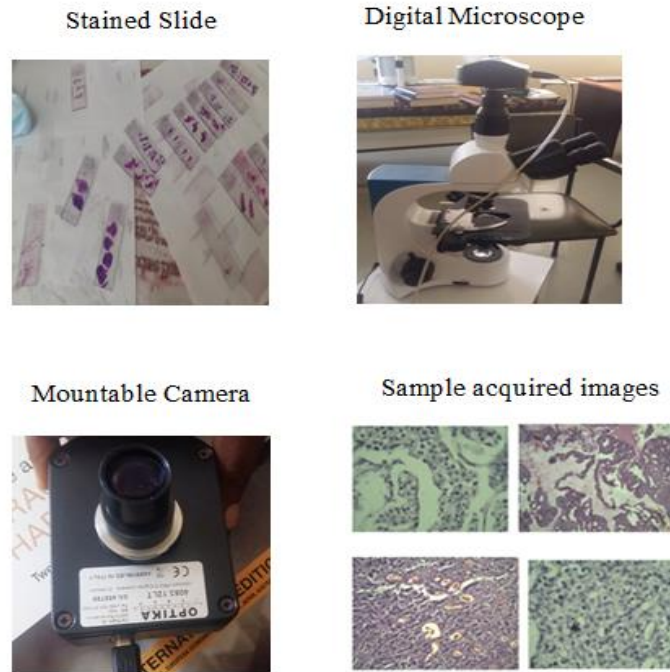
Figure 3-2: Local histopathological image acquisition procedure (Top-Right-Bottom) stained slides, Microscope with mounted camera, the camera used, sample acquired images.

Table 3-3: Table representation for local data

Name	Tumor Type	Total amount
Adenosis	Benign	10
Fibroadenoma	Benign	27
Ductal carcinoma	Malignant	42
Lobular carcinoma	Malignant	14
Mucinous carcinoma	Malignant	17

3.3. Image preprocessing

Image resizing and data augmentation are the key parts of model training procedure. In this study, all the images were resized to 224 x 224 pixel size and they were also rotated to 90^0, 180^0 and 270^0 in order to increase the number of data to be feed to the network by preserving tissue structures and different features of the histopathological images. As a result of this, meaningful samples were generated representing the same class as the original image to allow the network to

learn the features appropriately. When the images get augmented, the number of images in each class was increased four times the original.

After data augmentation, in order to have a uniform number of images in each class, a class with the smallest number of images (adenosis with 1776 images) was taken as a reference. Accordingly 1776 images were selected randomly from respective classes. Similarly for the grade dataset, data augmentation was done in order to increase the number of data. As a result, a total of four hundred twenty eight (428), four hundred and eight (408) and three hundred sixty four (364) images were generated for grade one, grade two and grade three, respectively. Since the difference of the number of images in each class is very small the data in each class is used as it is.

Finally, the whole data was split into training, validation and test sets and 80%, 10%, and 10 % of the images, respectively was used for each classification.

3.3.1. Image Enhancement

Images enhancement techniques are used in order to have more visual information than the original image. In this study contrast limited adaptive histogram equalization (CLAHE), which is one type of histogram equalization technique, was applied.

In the ordinary histogram equalization algorithm, once the cumulative distribution function (CDF) is computed, it is applied to all pixels of the given input image. These make the system to be susceptible to unnatural artifacts like intensity saturation, over-enhancement and noise amplification [41, 42, 43]. Due to this, a local contrast limited adaptive histogram equalization algorithm is necessary.

CLAHE is one type of adaptive histogram equalization which is used to enhance the local contrast of a given image. It emerged as an improvement for adaptive histogram equalization. It focuses on a small data region which is known as tile or block, thus every tile will be enhanced by disabling the amplification of noise problem and improving the contrast of the images [42].

For color images, same contrast limited adaptive histogram equalization algorithm was applied for each RGB channels separately to enhance the overall image contrast. The procedure of this research to transform an image using CLAHE is summarized in Figure 3-3.

```
┌──────────────────────────────────────────────────────────┐
│                     Get Input Image                      │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│              Change RGB to YCbCr color space             │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│     Separate the chrominance channels from luminance     │
│                         channel                          │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│        Set the number of regions and the clip limit      │
│              for both chrominance channels               │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│    Process each contextual region of the channels        │
│                       separately                         │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│      Combine the luminance channel with the processed    │
│                   chrominance channels                   │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│               Reverse the image to RGB space             │
└──────────────────────────────────────────────────────────┘
                            │
                            ▼
┌──────────────────────────────────────────────────────────┐
│               CLAHE result will be achieved              │
└──────────────────────────────────────────────────────────┘
```

Figure 3-3: Working principle of CLAHE

Color space models including RGB, YCbCr, YIQ, HSI and HSV are widely used for digital image processing and since humans eye perceive images in 3 measures: brightness, hue, and saturation, they can be divided into the part associated with the brightness and the part associated with the hue and saturation. YCbCr, YIQ, HSI, HSV color spaces are candidates to satisfy this condition. But, the YCbCr color space better preserves detailed information of luminance component than any other color spaces [41]. As a result, in this study, the first step for the transformation of images using the CLAHE algorithm is to convert the RGB images into YCbCr color space, to apply the algorithm in each color channel. The difference between the two color spaces is that the RGB represents color as red, green and blue. While, in the YCbCr, the Y component represents the brightness (luma), Cb is blue minus luma (B - Y) and Cr is red minus

luma (R - Y). That is the Y component is the luminance channel whereas, the Cb and Cr components are the chrominance channels.

CLAHE has two key parameters, the block size and the clip limit. The images quality is mainly controlled by the two parameters. The clip limit, limits the amplification by clipping the histogram at a predefined value. While the block size, is the size of the tile or block of an image. A given image will get brighter when the clip limit is increased. And as the block size is bigger, the dynamic range becomes larger and the contrast of image will increase [43]. For this study the default values were taken. That is, a clip limit of 2 and a block size of 8 x 8.

3.3.2. **Histogram matching**

Aforementioned, histogram equalization determines a transformation function which helps to get a uniform histogram of an image. But in some cases we need to have a specific shape of histogram for a specific image set rather than using a uniform histogram for all. When both the sample images and template images (reference image) are real color RGB images, using the histogram matching (HM) algorithm, each color channel of the given image will be matched independently with the corresponding color channel of the template image. Specifically for histopathological images, histogram matching helps to reduce the color difference in tissue samples due to variation in scanning condition and staining protocol [44].

In this study, a single image was chosen as a template image by considering the color of the tissue components. That is, as per the staining protocol the cytoplasm is presented as pinkish and nuclei are presented as blue. As a result considering the image as a properly stained slide image, it has been taken as a template image. Figure 3-4 shows the template for performing histogram matching.
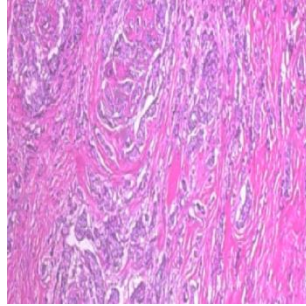
Figure 3-4: Template Image for histogram matching

Histogram matching/ histogram specification consists of equalization the original image's histogram, equalization of the desired probability mass function and applying inverse transformation function [39]:

Equalizing the original image's histogram using the CDF as a transformation function is done using the following formula:

$$s_k = T(r_k) = \sum_{j=0}^{k} \left(\frac{nj}{n} \right) \dots\dots\dots\dots\dots\dots (3.1)$$

where:

- $T(r_k)$: transformation function,
- s_k: new (mapped) pixel level
- r_k: original pixel level
- n_j: number of pixels having intensity r_k.
- n: total number of pixels in the image.

Equalizing the desired probability mass function (in other words, equalizing the desired histogram) is performed as indicated in equation (3.2) below.

$$v_k = G(z_k) \dots\dots\dots\dots\dots\dots\dots\dots (3.2)$$

where:

- $G(z_k)$: transformation function,
- v_k: new (mapped) pixel level of the reference image

- $z_{k:}$ original pixel level of the reference image

Finally inverse transformation function is done using the following equation.

$$z = G^{-1}(S) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.3)}$$

where:

- z : new pixel value after matching

In this step, the algorithm for obtaining the inverse of the transformation function for a given value of s_k would search through the values of v_k and store the index of the most recent value whose absolute difference $(|v_k - s_k|)$ is the smallest so far. This way the histogram of a template image can be mapped on to another so as to minimize color variation between images [39]. Finally data normalization was done using the built in Image-generator function in order to normalize all input image into the acceptable range of the model for having a better convergence.

3.4. Breast Cancer Classification

Image classification is a complex process in a computer vision which classifies an image according to its visual content. The accuracy of classification is mainly related to the characteristics of the dataset, the complexity of the problem under analysis, and the strength of the classification algorithm. Classification can be done using artificial intelligence system.

Several studies have been conducted to compare the performance of existing pre-trained models [41, 42]. Multiple performance indices were used to evaluate the models performance and to compare each model with the others. These are: recognition accuracy, model complexity (by counting the total amount of learnable parameters), computational complexity (considering the computational cost of each DNN model), memory usage (considering both the memory allocated for the network model and the memory required while processing the batch.), and inference time (considering the same batch sizes) are used. In most literatures ResNet 50 model is reported to outperform other state of the art models [22, 25]. In this study, ResNet-50 pre-trained model was chosen to solve breast cancer histopathology image classification problem.

3.4.1. Res-Net Model

It was first introduced by He et al. in 2015 in the paper entitled Deep Residual Learning for Image Recognition [46]. It is known that deep convolutional neural networks are highly useful in solving image classification problems. But, before ResNet, training a very deep neural network was difficult due to the problem of vanishing gradient, as the gradient is back propagated to earlier layers and having repeated multiplication, the gradient becomes extremely small. For this reason, as the network goes deeper, its performance gets saturated or it will start degrading rapidly. The major advantage in using ResNet is it allows training extremely deep neural networks successfully without being affected by vanishing gradients [46]. This is achieved by introducing 'identity shortcut connection" that skips one or more layers. And add the input of a layer with the output of the layer so as to feed the next layer of the network. Identity short- cut connections add neither extra parameter nor computational complexity [46].

There are many variants of ResNet architecture that have same concept but different number of layers. These are ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164, ResNet-1202 and the like. The numbers following the name ResNet implies the ResNet architecture with a certain number of neural network layers.

Like any other convolutional neural networks, the ResNet model has two parts; the first is the convolutional base which is composed of the convolutional and pooling layers, which is responsible for generating the features from an image, whereas, the second part of the model is the classifier that is composed of the fully connected layers [47]. The detail architecture of ResNet50 is shown in Figure 3-5.
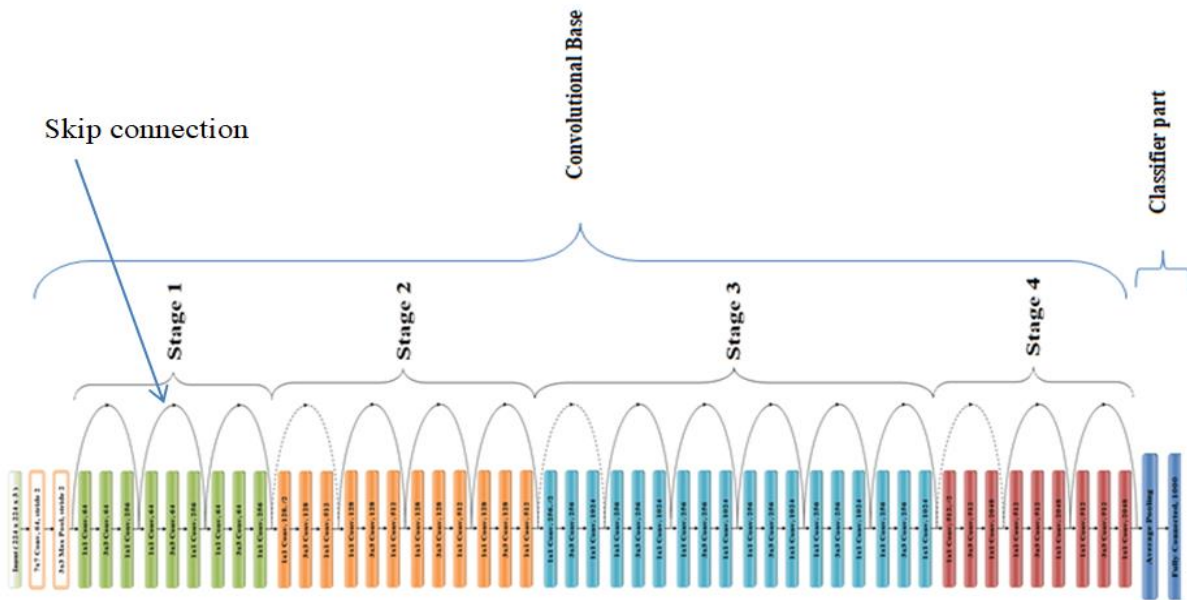
Figure 3-5: ResNet-50 Architecture

ResNet50 has four stages. And once an input is given to the network, it initially starts with performing the initial convolution and max-pooling using 7×7 and 3×3 kernel sizes respectively. After this, Stage one of the network will begin and it has three Residual blocks containing three layers each. The size of kernels used to perform the convolution operation in all three layers of the block of stage one are 64, 64 and 128 respectively. Moreover, in order to overcome vanishing gradient problem, the model makes use of Identity Connection (the process of taking the original input to the module and adding it to the output of a series of operations), which is referred by the curved arrows. The other key feature of ResNet50 is the use of bottleneck design. This feature is added in order to increase the performance of the network. As shown in the diagram, for each residual function, three layers are stacked together. These are 1×1, 3×3, 1×1 convolutions. The 1×1 convolution layers are used to reduce and restore the dimensions whereas the 3×3 layer is left as a bottleneck with smaller input/output dimensions. Lastly, the network has an Average Pooling layer followed by a fully connected layer having neurons equal to the number of classes to be classified [48, 49]. Further the detail network connection is shown in Figure 3-6.

In general the unique features of ResNet model are; it uses Batch normalization at its core. That is after each convolution and before activation layers, it uses bottleneck residual block design to

increase the performance of the network and it uses identity connection to avoid vanishing gradient problem.
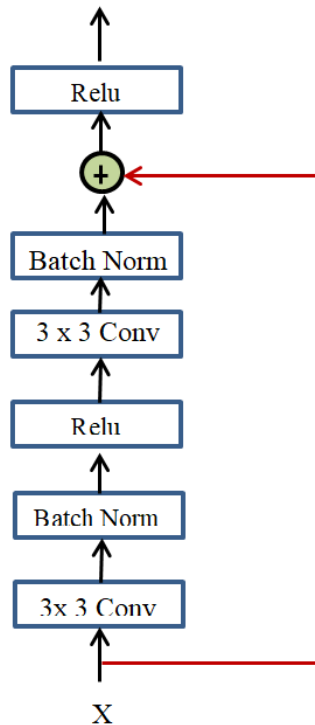

Figure 3-6: Detail visualization of the networks connection

It is worth to use ResNet model since it has fewer filters and lower complexity [46]. Moreover it is successfully trained on CIFAR-10 set with over 100 layers and explores models with over 1000 layers. Applying this model on the ImageNet dataset, it has won the 1st place in the ILSVRC 2015 classification competition. Moreover, this model has also been able to win and get the 1st place on: ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation in ILSVRC & COCO 2015 competitions [46]. From these results it can be generalized that the ResNet model is a reliable model to solve image classification problem.

There are different ways of repurposing pertained models based on the amount of dataset present and the computational power we have. The first is taking the architecture of the model and training the entire model. In this case the model will learn from scratch by using custom dataset. But, to achieve good result one needs to have a large number of dataset and good computational power. The other is again using the architecture of the model and leaving some of the layers frozen and some layers to be trained. In this case the frozen layers weight value will not change

during training. This method is usually used if the dataset is large and the number of parameters is small. The last method is freezing the whole convolutional base and using the model as a feature extractor and then feed its output to a classifier. This technique is used if there are fewer amounts of dataset and less computational power [47].

Likewise, in this study ResNet-50 model was adapted and fine-tuned to classify BC images into binary and multiclass classes. The model was used as a feature extractor. That is, both its architecture and weight value (from ImageNet) were adapted. Then the top layers were frozen, and some parameters like the learning rate, optimizer type, loss function and decay rate, which are used to optimize the model, were adjusted. Finally, by using soft-max classifier the probability distribution of each possible class was found and based on the most probable class the images were classified accordingly. Figure 3-7 demonstrates the overall methodology used in this study.
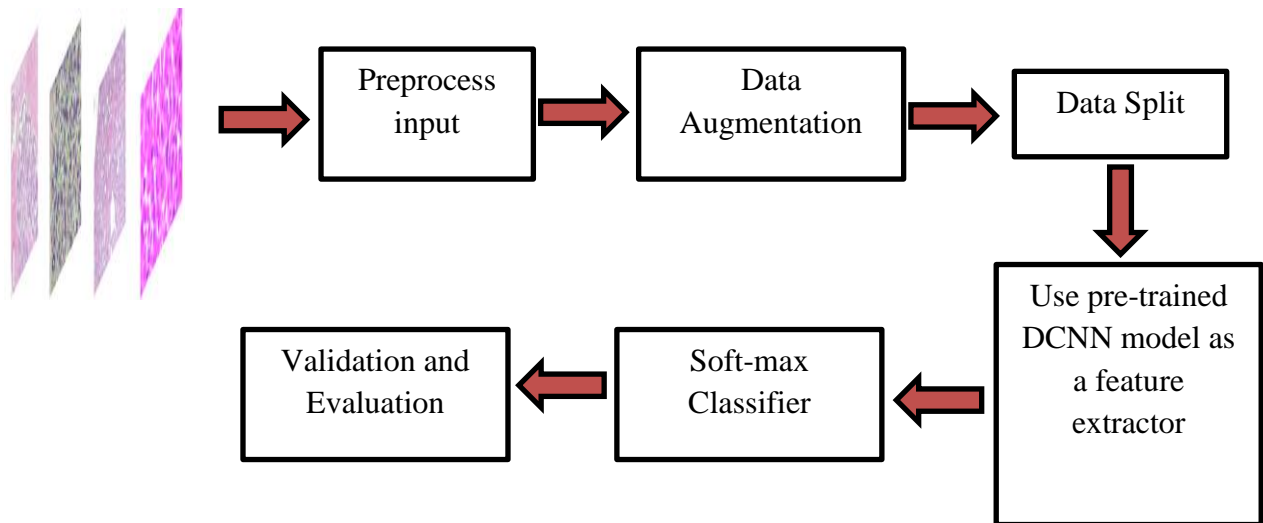

Figure 3-7: Overall methodology

3.4.1.1. Hyper-parameters and parameters of ResNet50

In this research, different hyper-parameters were fine-tuned to increase the performance of the system. These includes, choosing the right optimizer, adjusting the learning rate, choosing the appropriate activation function and choosing the proper loss function.

As an optimizer ADAM [49] optimizer was chosen for its best performance in terms of speed to converge faster and accuracy [19]. The number of epoch used was 50 while learning rate was set to 0.01- 0.001 and activation function used was ReLu. Loss function for binary class classification was binary_crosentropy, while for multi class classification it was categorical_crossentropy.

3.5. Performance evaluation metrics

After building a model and training the network, its performance must be evaluated so as to know the actual result. There are different ways of evaluating a model. The first one is by using a confusion matrix and getting the TP, TN, FP and FN rates of the predicted values. The second is by calculating and getting the precision, recall, F1-score and over all accuracy values of a model.

For example: Given a binary classifier of class A and class B, a model can be evaluated using confusion matrix. Table 3-4 indicates the TP, TN, FP and FN components for class A. From the table we can see that the green shaded part indicates the TP and TN part [50].

- A TP (True Positive) value indicates that what is predicted is true.
- A TN (True Negative) value indicates that the predicted class is truly negative.
- A FP (False Positive) value indicates that a thing is predicted as if it is part of the class while it is not.
- FN (False Negative) the prediction indicates that it is not part of the class while it is.

Table 3-4: Example of confusion matrix

Predicted class

		A	B
Actual class	A	TP	FN
	B	FP	TN

Once the confusion matrix is ready, the classification report containing the precision, recall and f1-score can be done. So, given a class prediction from the classifier, the precision is the one which answers the question "how likely is it to be correct?" It is calculated using equation (3.4), recall or sensitivity will indicate the answer for "will the classifier detect it?" It is calculated using equation (3.5); F1-score is the harmonic mean of precision and recall. It is calculated using equation (3.6). The model is said to be performing good if we have high F1-score and specificity is the one which determines the proportion of actual negatives that are correctly identified. To measure the proportion of actual negatives that are correctly identified, the specificity of the model was calculated using equation (3.7). Finally, the models performance was measured using the accuracy metrics using equation (3.8). The performance metrics were calculated using the equations given below [50]:

$$\text{Precision} = TP / (TP + FP) \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\quad (3.4)$$

$$\text{Recall} = TP / (TP + FN) \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\quad (3.5)$$

$$\text{F1-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad\dots\dots\dots\dots\quad (3.6)$$

$$\text{Specificity} = TN / (TN + FP) \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\quad (3.7)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FN + FP) \quad\dots\dots\dots\dots\dots\dots\dots\quad (3.8)$$

3.6. Graphical User Interface (GUI)

In this study, Pyqt5 designer software and a python programming language were used to develop the final interface of the system. It incorporates "Acquire image button" and "browse image button" in order to acquire or browse the digital histopathological image respectively, "preprocess button" to process the acquired image, "binary classification button" to operate the binary classifier model, "benign subtype analysis button" helps to classify a given image into detail benign subtype classes, "malignant subtype analysis button" to further classify malignant type of cancer into the basic four types, "grade level analysis button" to determine the grade of malignant ductal carcinoma disease. Furthermore, the "clear result button" helps to clear the result displayed previously and the "save result button" helps users to label the given image as per the result and save it. Figure 3-8 shows a snap shot of the general layout of the developed GUI.
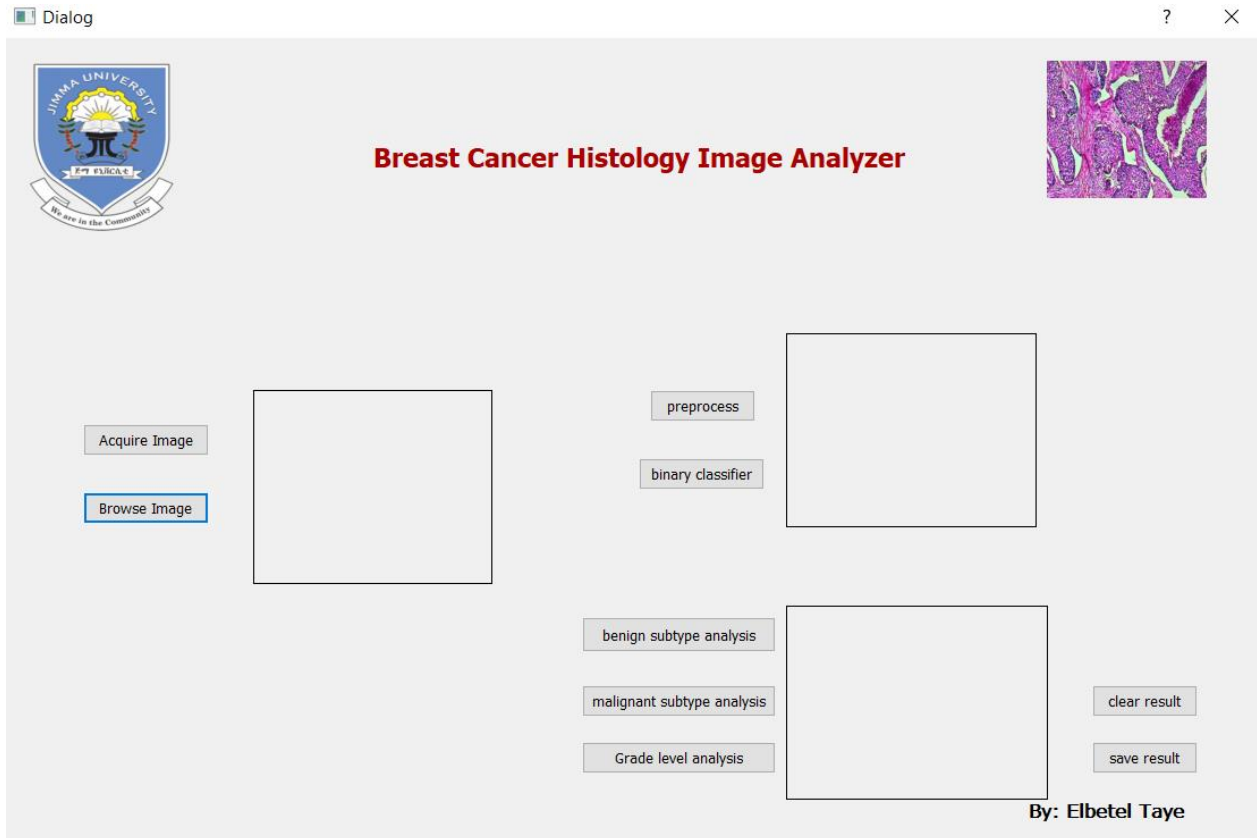

Figure 3-8: shows the layout of the developed GUI.

3.7. Material used in this research

The main materials used to conduct the study are discussed in table 3-5.

Table 3-5: Material used for the study

No	Soft Ware	Hard Ware
1.	Optika vision Lite 2.1 software: to read and save the image coming from the mounted camera.	Digital high-resolution (2592 x 1936) USB camera which is known as the Optikam PRO 5. Digital camera C-mount was used to capture a digital microscopic image.
2.	Python 3.7.3 software with different modules like keras, tensorflow, sklearn and others for writing and training the whole system.	Optika microscope having 40x, 100x, 200x & 400x magnification lenses. SN:455343
3.	Pycharm and Jupyter notebook python IDEs were used to write and execute the code.	ASUS Intel(R) Core(TM) i5-8250 CPU @ 1.60GHz 1.80GHz Laptop having 8GB RAM, 64 bit operating system, x64-based processor, window 10.
4.	PyQt5 designer to develop the graphical user interface of the system (GUI).	-

Chapter Four

Results and Discussion

4.1. Results

4.1.1. Preprocessing results

4.1.1.1. Image resize and data augmentation

Images were resized to 224 x 224 in order to get a uniform size of images and to make the data size equal with the input data size of the classifying network. Moreover, data augmentation was applied by rotating all the images to 90^0, 180^0 and 270^0 to increase the total number of data without affecting the images features. Figure 4-1 shows scale of images after being resized. And Figure 4-2 shows sample result from data augmentation.
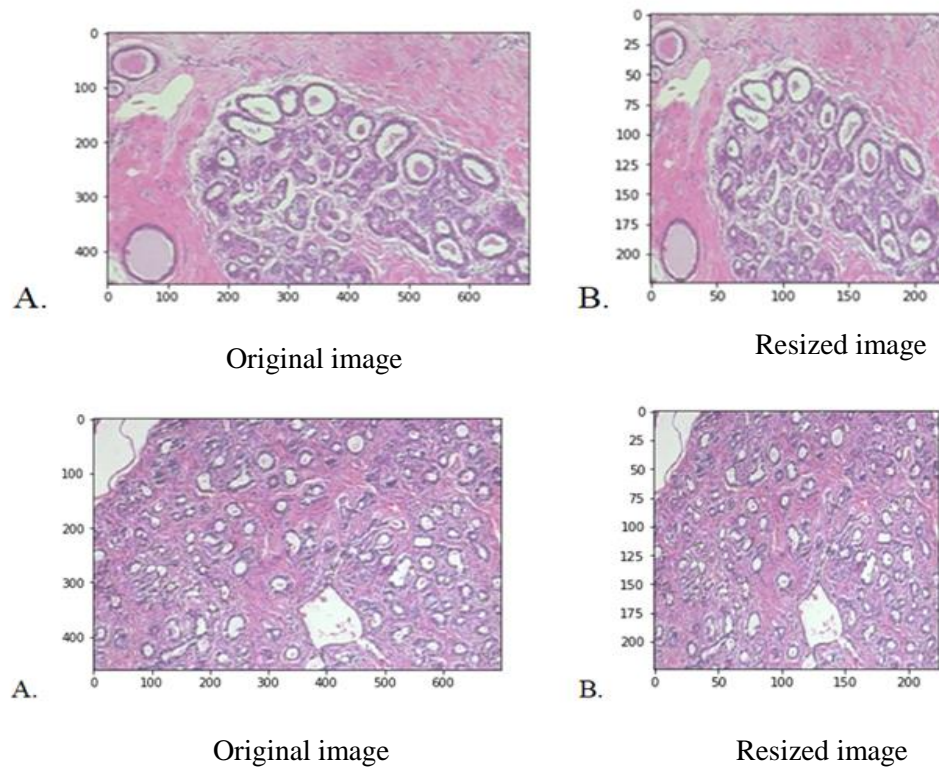

A.   Original image

B.   Resized image

A.   Original image

B.   Resized image

Figure 4-1: Sample resized images
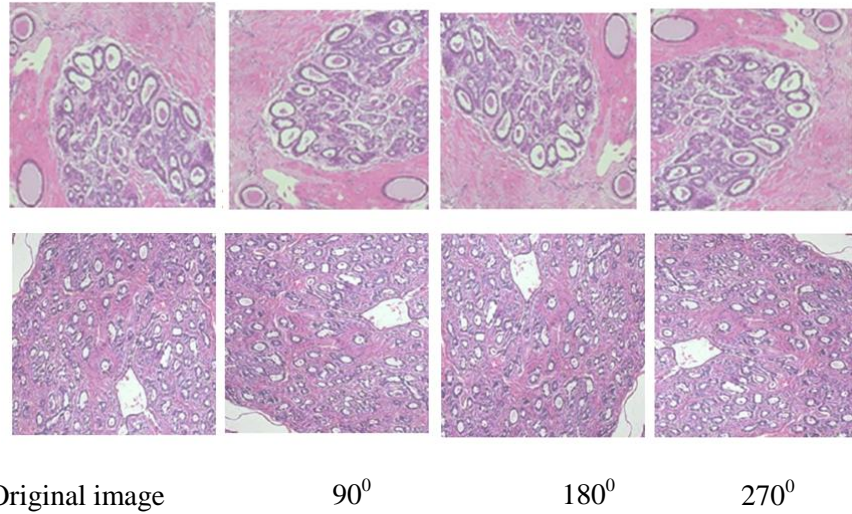
| Original image | 90^0 | 180^0 | 270^0 |

Figure 4-2: Sample images after Data augmentation using 90, 180 and 270 degrees respectively

After data augmentation the number of total image in the data repositories was increased by four times (4 x). Table 4-1 elaborates the number of images in each class after data augmentation was applied.

Table 4-1: Total number of images after data augmentation

	BC type	Total number of images
Benign	**Adenosis**	1,776
	Fibroadenoma	4,056
	Phyllodes	1,812
	Tubular adenoma	2,276
malignant	**Ductal**	3,451
	Lobular	2,504
	Mucinous	3,168
	Papillary	2,240

Moreover, to have uniform number of images in each class, the smallest number of image in a class after data augmentation was taken as baseline for limiting the number of images in all

individual classes. As a result, a total of one thousand seven hundred seventy six (1776) images were taken per BC type class and totally, fourteen thousand two hundred and eight (14,208) images were generated from all the eight subtypes of BC.

Finally, one thousand four hundred and twenty (1420) images, one hundred seventy nine (179) images and one hundred seventy seven (177) images were taken from each class for training, validating and testing the system respectively as indicated in Figure 4-3.
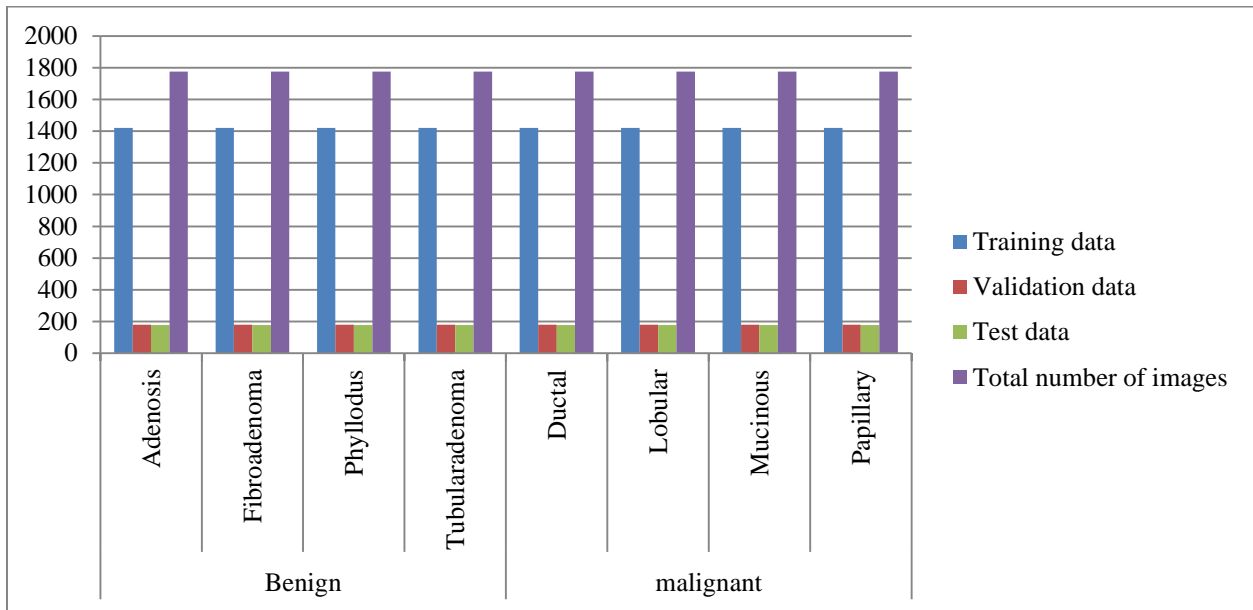

Figure 4-3: Distribution of data for training, validating and testing the system.

Similarly, for the grade classification task, all the images taken from the zendo data set [38] were resized from 1280 x 960 pixel to 224 x 224. The data augmentation was done in order to increase the total number of images. Figure 4-4 shows the result obtained after data augmentation. Finally, a total of one thousand two hundred (1200) images were generated. Then as indicated in Figure 4-5, the image data was split into training, validation and test sets with the ratio of 80 percent, 10 percent and 10 percent respectively.
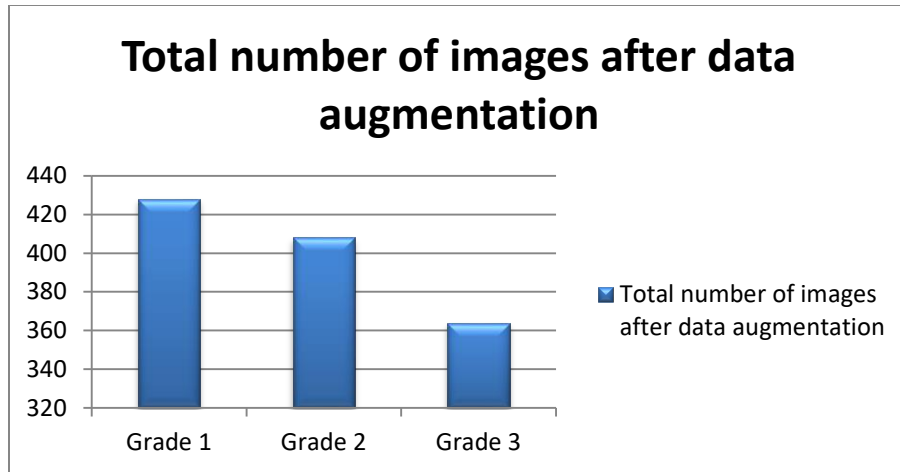
Figure 4-4: Total number of images after data augmentation

Finally, a total of four hundred twenty eight (428), four hundred and eight (408) and three hundred sixty four (364) images were generated for grade one, grade two and grade three respectively.
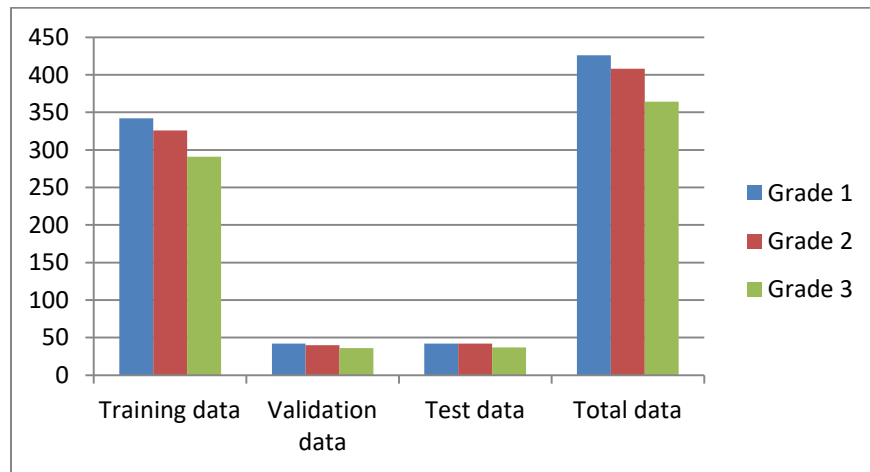

Figure 4-5: Data distribution for grade class classification

4.1.1.2.   Effect of Contrast limited adaptive histogram equalization

In this study, the images contrast has been increased using the CLAHE technique. Sample example is given in Figure 4-6 to show the effect of CLAHE on histopathological images.
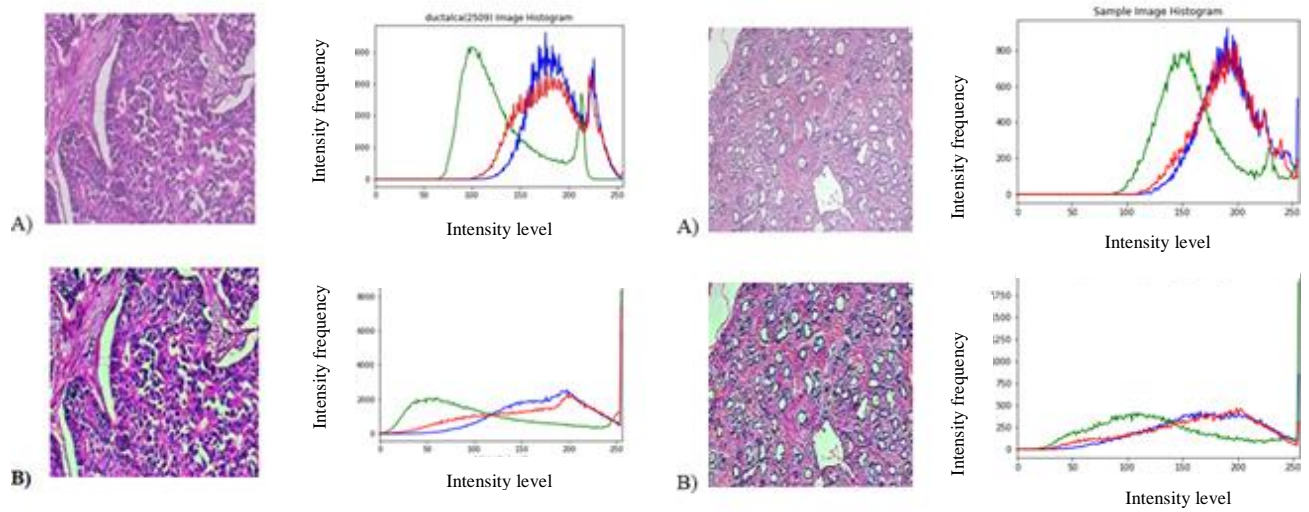
Figure 4-6: Effect of CLAHE on BC histopathological image. A) Original image B) CLAHE applied image

4.1.1.3.   Effect of histogram matching on original image

The color difference of the histology images was corrected using the histogram matching technique. Figure 4-7 shows sample result of the effect of histogram matching in this study.
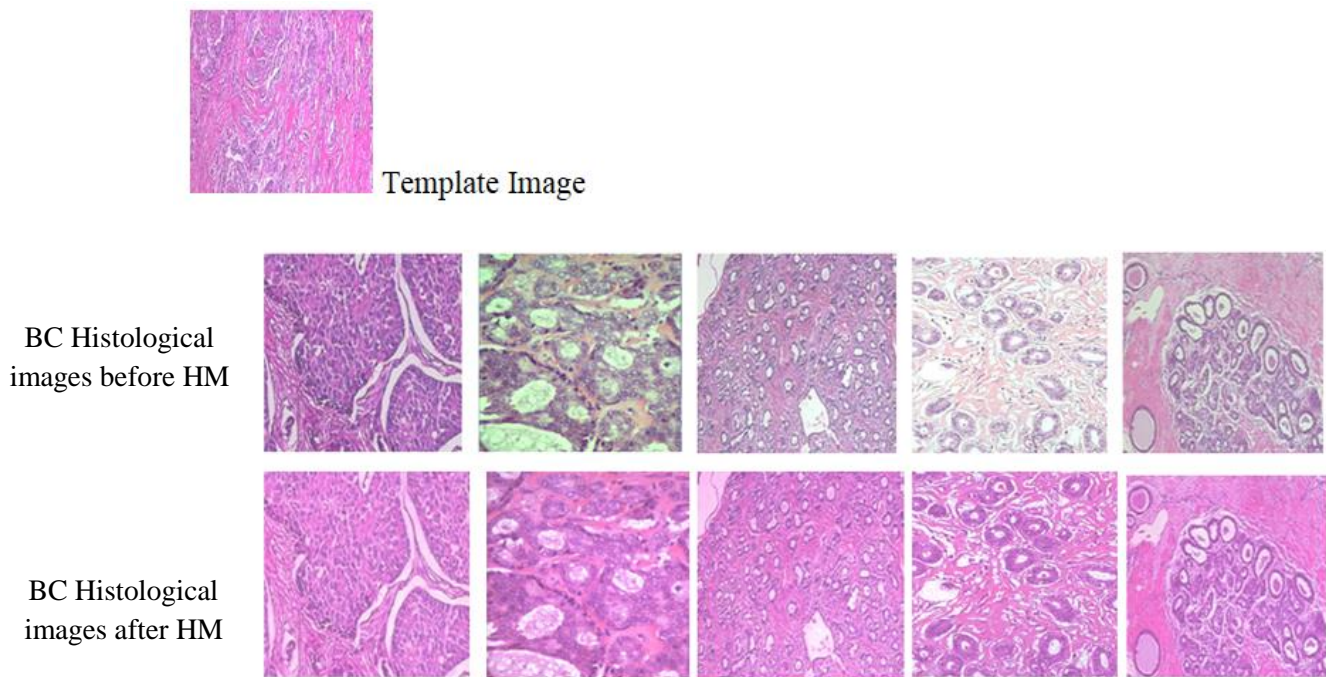

Figure 4-7: Effect of histogram matching

4.1.2. Model Training Results

4.1.2.1. Binary classification of breast cancer

In this study, binary classification indicates the classification of given input image into benign and malignant breast cancer type. As per the data split ratio used, the number of data used for training the system is five thousand six hundred eighty eight (5,688) images for each class. That is, a total of eleven thousand three hundred seventy six (11,376) image data were taken as a training set. And seven hundred and eight (708) images were taken as a validation data from each class. That is, a total of one thousand four hundred sixteen (1416) images. Likewise, for the test set a total of seven hundred and eight (708) images were taken for each class as a test set, which is a total of one thousand four hundred sixteen (1416) images.

Once data split and other preprocessing procedures were complete, ResNet-50 model was trained to do the classification task. Hence, the model was trained using the training dataset and validated with the validation dataset. Finally the learning and generalizability performance of the model was measured using a learning curve. Figure 4-8 and figure 4-9 show the training and validation accuracy and loss plot on Epoch versus Accuracy and Epoch versus Loss graph. The train learning curve is calculated from the training dataset. It shows how well the model is learning. While the validation learning curve is calculated from a hold-out validation dataset to see how well the model is generalizing.
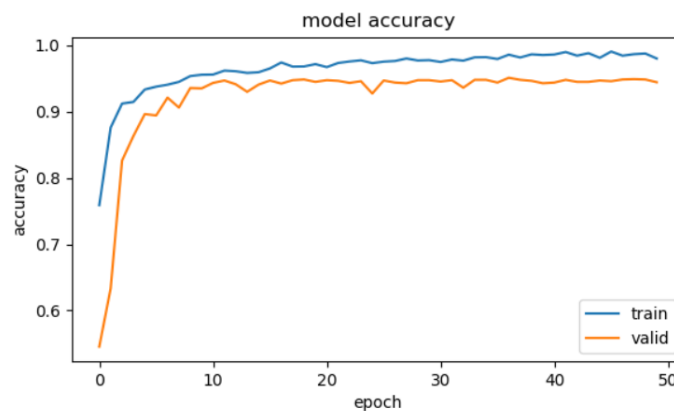

Figure 4-8: Training and Validation accuracy curve with respect to the Epoch
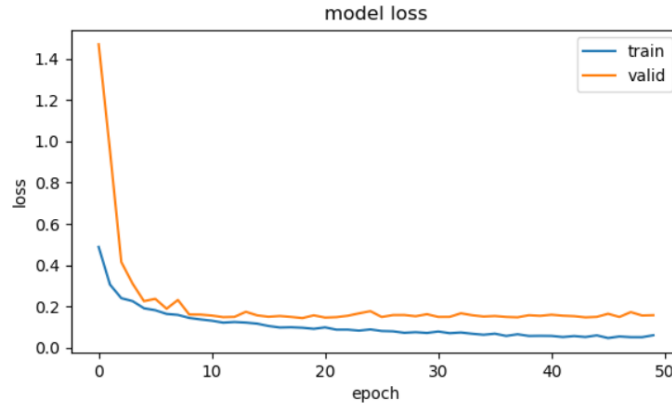
Figure 4-9: Training and Validation loss curve with respect to the Epoch

In this phase, better validation loss was achieved at the 37th epoch that is validation loss of 0.1490. Hence, the model has saved the weight value acquired at the 37th epoch for doing the classification task. As a result, 98.58 % training accuracy and 95.09 % validation accuracy were achieved.

4.1.2.2.   Multi-class classification of breast cancer

Multi class classification indicates the classification of benign and malignant types into their respective four subtypes. The systems performance was measured for both.

A. Benign type classification

The benign types of breast cancer in this study include four types. These are: adenosis, fibroadenoma, phyllodes and tubularadenoma. From the data augmentation process the total number of images present in each class were increased. As a result, five thousand six hundred eighty eight (5,688) number of data or images were used as a training dataset, one hundred seventy seven (177) number of data as a validation set and one hundred seventy seven (177) images were used as a test set for the benign class classification task.

Finally, the training data was given to the constructed model which is a fine tuned ResNet50 model. The models' learning and generalization ability performance was measured using the training and validation curves respectively. From all the values resulted in each epoch, the lowest validation loss, 0.1598, was achieved at the 27th epoch. As a result, the models weight value at the 27th epoch was saved for doing the classification task. As a result, an accuracy of 99.14%

training accuracy and 94.57% validation accuracy was achieved during training. Figure 4-10 and Figure 4-11 indicates the results obtained during the training phase.
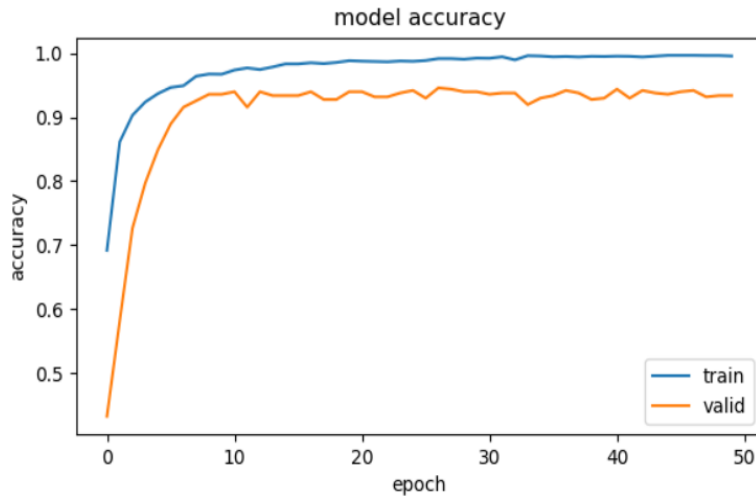

Figure 4-10: Training and Validation accuracy curve with respect to the Epoch
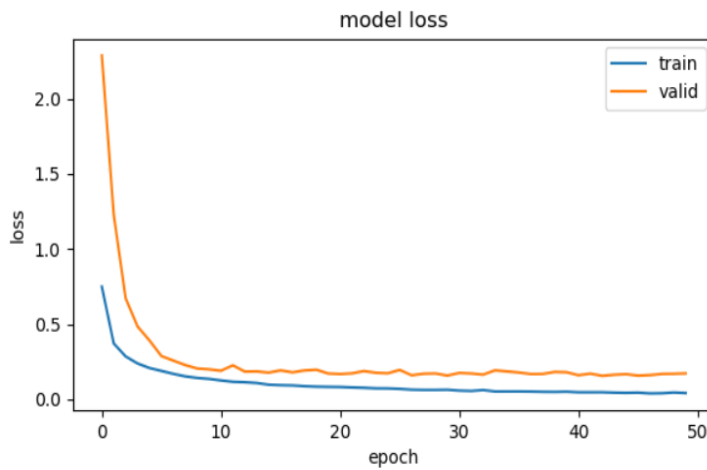
(Benign type classification)

Figure 4-11: Training and Validation loss curve with respect to the Epoch

(Benign type classification)

B. Malignant type classification

The malignant types of breast cancer are the invasive once. In this study Ductal Carcinoma, Lobular carcinoma, Mucinous carcinoma and papillary carcinoma are included. Once the images were collected they were arranged properly and were resized and preprocessed using the above stated preprocessing techniques. After data augmentation process the total number of images present in each class were increased. As a result, five thousand six hundred eighty eight (5,688) number of data or images were used as a training dataset, one hundred seventy seven (177) number of data as a validation set and one hundred seventy seven (177) images were used as a test set for the malignant class classification task. Finally, the training data was given to the constructed model and by using the learning curve the models performance in learning and generalization for a hold-out validation data set was measured. During training, the lowest validation achieved was 0.1657 at the 49th epoch. This indicates that, the weight value saved for doing the classification task is the value that was obtained at the 49th epoch. As a result, an accuracy of 93.56 percent validation accuracy was achieved. Figure 4-12 and Figure 4-13 indicate the results obtained during the training phase.
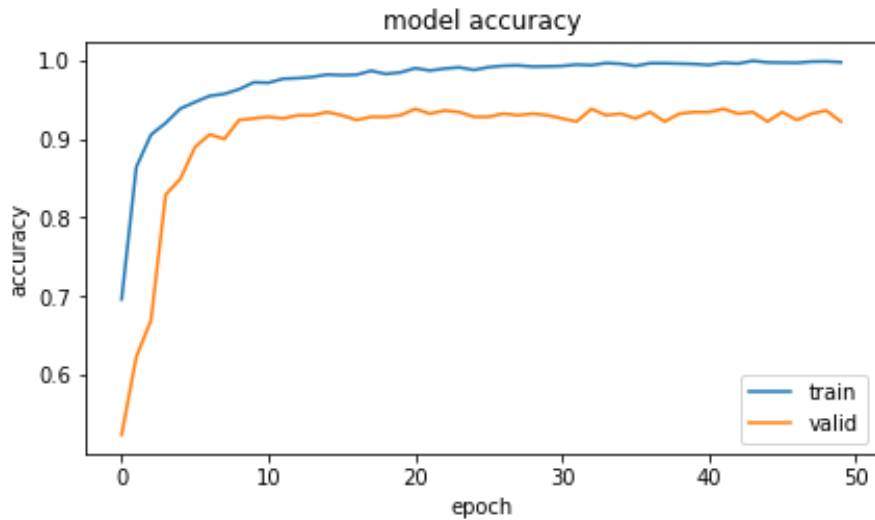

Figure 4-12: Training and Validation accuracy curve with respect to the Epoch

(malignant type classification)
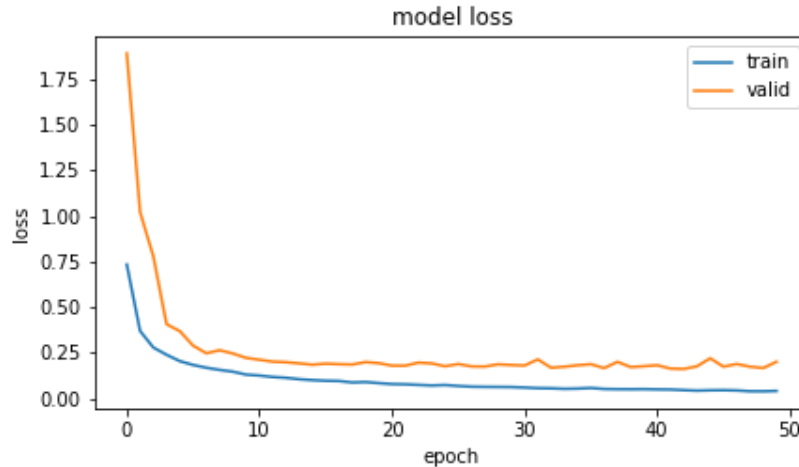
Figure 4-13: Training and Validation loss curve with respect to the Epoch

(Malignant type classification)

C. Grade classification for IDC malignant type of cancer

The grade indicates the invasiveness or aggressiveness of the cancer type. It is classified into three classes Grade 1, Grade 2 and Grade 3. Of the total images, three hundred forty three (343), three hundred twenty seven (327) and three hundred (300) images were taken for grade 1, grade 2 and grade 3 classes as a training set respectively. Moreover, thirty two (32) images were taken as a validation set from each class.

Finally the data was feed to the model and its' learning and generalization ability performance was measured using the training and validation curve respectively. During training time, of all the validation loss values achieved during the 50 epochs, the lowest validation value was realized at the 42th epoch, that is, 0.0652 validation loss values was attained. At this point, an accuracy of 97.44 percent validation accuracy was achieved during training. Figure 4-14 and Figure 4-15 indicate the results obtained during training phase for the grade level classification task.
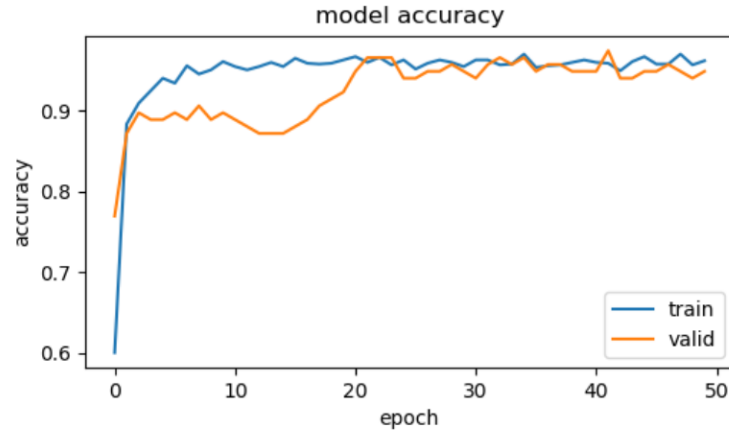
Figure 4-14: Training and Validation accuracy curve with respect to the Epoch

(Grade classification)
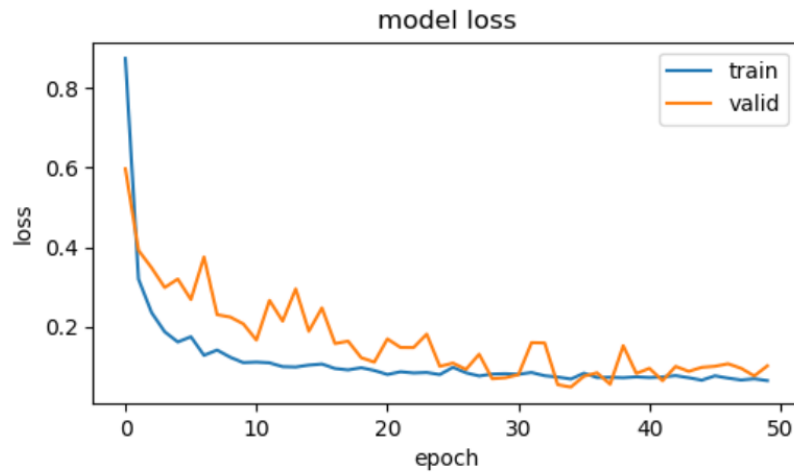
Figure 4-15: Training and Validation loss curve with respect to the Epoch (Grade classification)

4.1.2.3.   Summary of training results

The overall result of training phase of the system including the training accuracy, validation accuracy, training loss and validation loss for the binary and multiclass classifiers can be summarized as indicated in Figure 4-16.
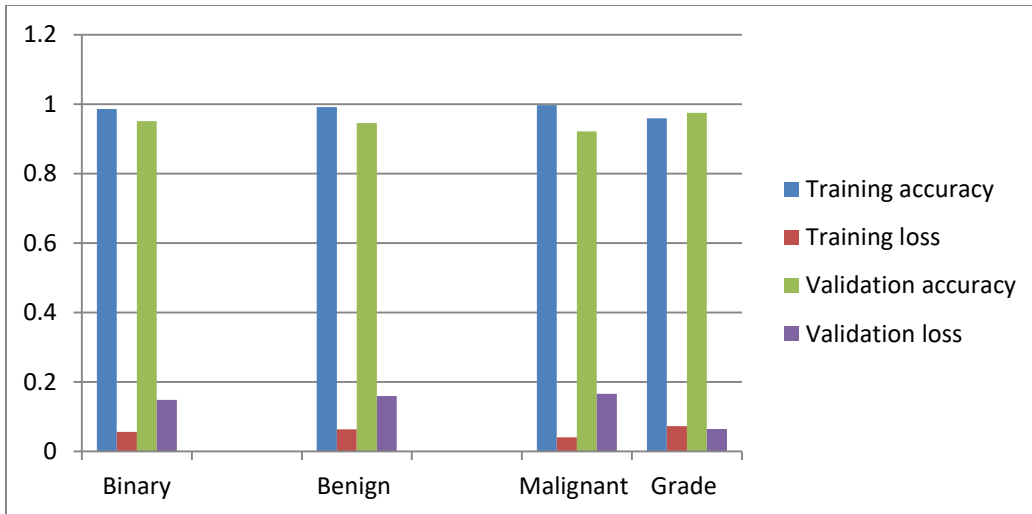
Figure 4-16: Training phase result summary

4.1.3. Testing Phase Results

4.1.3.1. Binary classification

In the binary classification, the model is expected to categorize given images into benign or malignant classes. To do this, seven hundred and eight (708) images were considered from each class. Based on the actual and predicted values a confusion matrix was generated. Figure 4-17 shows the resulted confusion matrix for the binary classifier. "0" represents the malignant class whereas "1" represents the benign class. Out of the 708 images in the malignant class, 687 images were correctly classified as malignant cancer types, while 21 of the images were predicted as benign class. On the other hand, for the benign class, out of the 708 images, 683 images were correctly classified as benign tumors, whereas 25 of the images were predicted as malignant types.
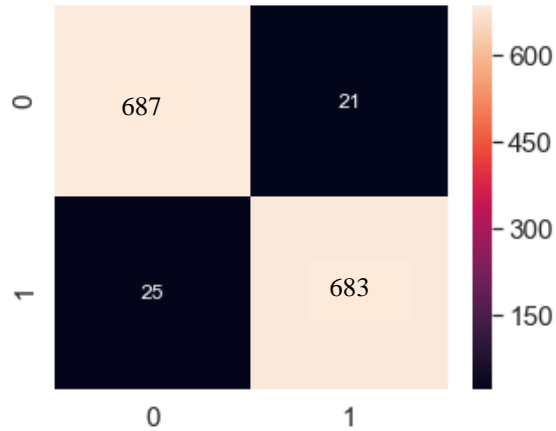
Figure 4-17: Confusion matrix for the binary classifier

Once the confusion matrix is done, the TP, TN, FP and FN values are easily known. And from those values the precision, recall, specificity, f1-score and test accuracy were calculated. Table 4-2 represents the overall result of binary classification.

Table 4-2: Result of binary class classifying model

	Precision	Recall	F1-score	Specificity	Accuracy
Benign	97.02	96.47	96.74	97.03	96.75
Malignant	96.49	97.03	96.76	96.47	96.75

Finally average test accuracy of the model for binary classification was calculated using the average of the accuracies of the two classes. As a result, 96.75 % test accuracy, 96.755 % precision, 96.75 % recall and 96.75 % specificity were achieved.

4.1.3.2. Benign type classification

The benign class consists of Adenosis tumor, fibroadenoma tumor, phyllodus tumor and the tubular adenoma. These classes are labeled as "0", "1", "2" and "3" respectively.

For testing the performance of the developed model, one hundred seventy seven new images were considered from each class. And the performance was measured by comparing the actual and predicted values using a confusion matrix. Figure 4-18 indicates confusion matrix for the benign type's classification.
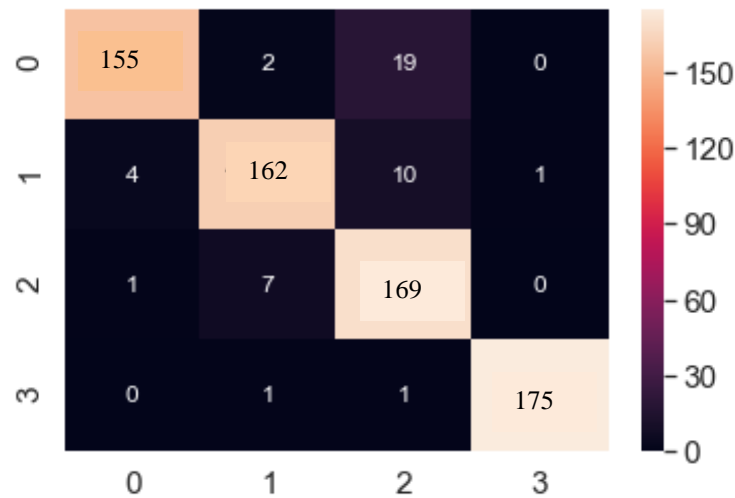
Figure 4-18: Confusion matrix of benign classification

From the confusion matrix it can be understood that the diagonal light colored values indicate the true positive values of the four classes, the x- axis indicates the predicted class values while the y-axis indicates the actual class values. Therefore, the column for each class in the confusion matrix except the true positive value indicates the false positive values, while the row for each class, except the true positive value, indicates false negative values.

4.1.3.3.   Malignant type classification

The malignant cancer types considered in this study are; ductal carcinoma, lobular carcinoma, papillary carcinoma and mucinous carcinoma. These classes are labeled as "0", "1", "2" and "3" respectively.

Similarly, a separate dataset was used to test the performance of the constructed model. As a result, the directory of the test images was given to the model and the system has predicted each images class one by one. Finally, the confusion matrix was built. The prediction's output for malignant class classification is shown in the Figure 4-19.
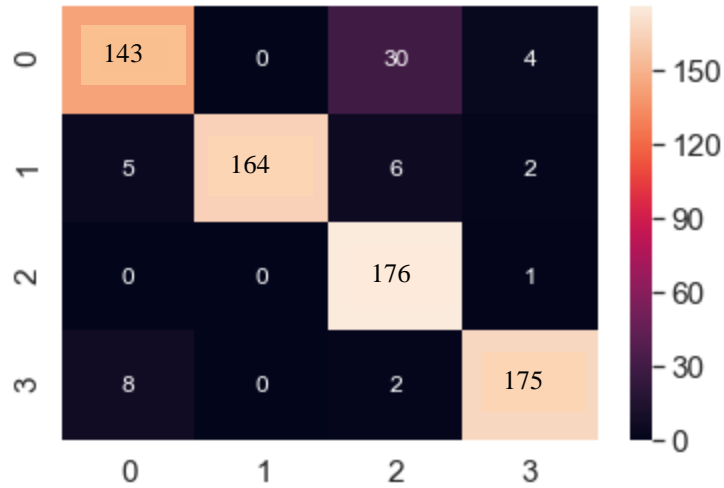
Figure 4-19: Confusion matrix of malignant classification

From the TP, TN, FP and FN values in the confusion matrix; the precision, recall, specificity, f1-score and test accuracy were calculated for both benign and malignant type classifications. Table 4-3 represents the overall result of multi class classification model.

Table 4-3: Result of multi-class classifying model

	precision	Recall	F1-score	specificity	Accuracy
Adensis	96.9	88.14	92.31	99.05	96.33
Fibroadenoma	94.2	91.53	94.19	98.03	96.36
Phyllodus	84.93	95.48	90	94.3	94.6
Tubularadenoma	99.43	98.87	99.15	99.8	99.5
Ductal	91.7	81	85.9	97.5	93.3
Lobular	100.0	92.6	96.19	100.0	98.04
Mucinious	96	94.35	95.2	98.57	97.45
Papillary	82.24	99.44	90.03	92.6	94.34

Lastly, the total models performance in multiclass classification task has been assessed by calculating the average precision, recall (sensitivity), specificity and test accuracy results.

Therefore, for the benign type classification, an average accuracy of 96.7 %, average precision of 93.87 %, average recall (sensitivity) of 93.51 % and 97.8 % specificity were achieved.

For the malignant type classification task, an average test accuracy of 95.78 %, average precision rate of 92.49 %, an average recall (sensitivity) of 91.85 % and 97.17 % of average specificity were achieved.

4.1.3.4. Grade classification

Once the data and the trained model were ready, the image data has passed through the preprocessing techniques and get preprocessed. Finally, it was fed to the model so as to evaluate the performance of the model. The result is indicated in the Figure 4-20 using a confusion matrix.
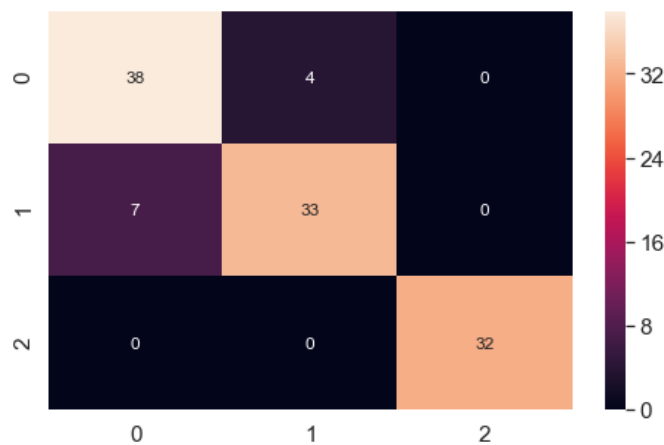
Figure 4-20: Confusion matrix for grade level identification

In the confusion matrix, the grade levels are labeled as, "0" representing grade 1, "1", representing grade 2 and "2" representing grade 3. From the result found using the confusion matrix, the precision, recall, specificity, F1-score and test accuracy results were calculated. Table 4-4 indicates the grade classification results.

Table 4-4: Result of grade level classifying model

	Precision	Recall	F1-score	Specificity	Test accuracy
Grade 1	84.44	90.48	87.4	91.14	90.91
Grade 2	89.2	82.5	85.71	94.87	90.68
Grade 3	100	100	100	100	100

Finally, the general performance of the developed model for classifying the grade of a given image was assessed and an average test accuracy of 93.86 percent, precision rate of 91.21 percent, recall rate of 90.99 percent and specificity of 95.34 percent was achieved.

Generally, based on the test results, the summery of the performance of the model (ResNet 50) is summarized in Table 4-5.

Table 4-5: Test result summary

	Precision (%)	Recall (%)	Specificity (%)	Test accuracy (%)
Binary classification	96.755	96.75	96.75	96.75
Benign type classification	93.87	93.51	97.8	96.7
Malignant type classification	92.49	91.85	97.17	95.78
Grade level identification	91.21	90.99	95.34	93.86

4.1.4. Graphical user interface (GUI)

The developed GUI was tested with respect to response time and ease of use. It is found to be easy to use and is convenient for users. Once initialized, result can be achieved within 3 seconds. Figure 4-21 shows the pictures for each step. Figure 4-21A shows the layout of the GUI.

"Acquire Image" and "Browse Image" buttons enable the user to load a histopathological image as shown in Figure 4-21B, "preprocess" button enables the user to preprocess the loaded image and get an output as indicated in Figure 4-21C. "Binary classifier", "benign subtype analysis", "malignant subtype analysis" and "Grade level analysis" buttons enable users to operate the needed classification task as indicated in Figure 4-21D. Finally, the result obtained can be saved using the "save" button and to continue on analyzing more images while the "clear" is used so as to clear the previous result and start freshly. Illustration is given in Figure 4-21E and Figure 4-21F.

A.

B.

C.

D.

E.

Breast Cancer Histology Image Analyzer

Acquire Image

Browse Image

preprocess

binary classifier

benign subtype analysis

malignant subtype analysis

Grade level analysis

clear result

save result

By: Elbetel Taye

F.
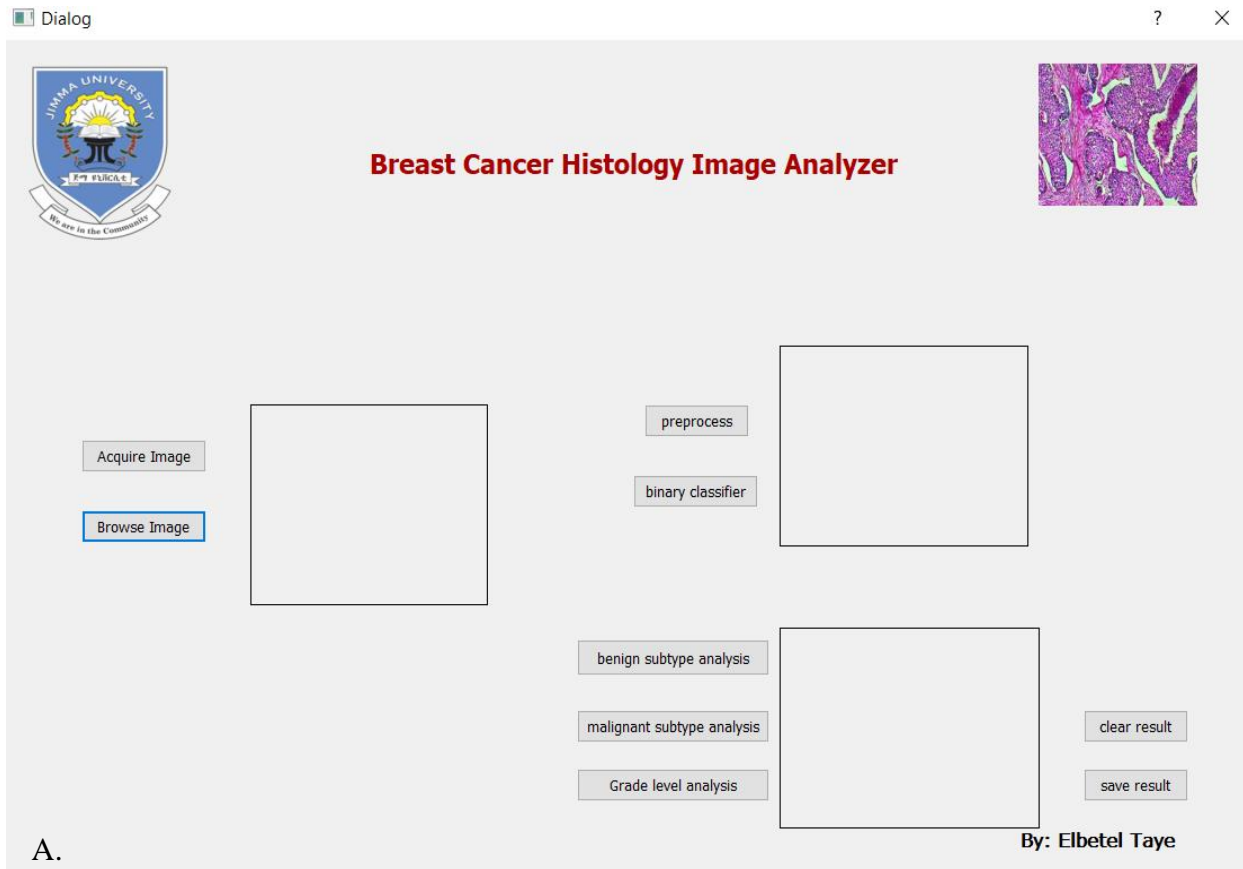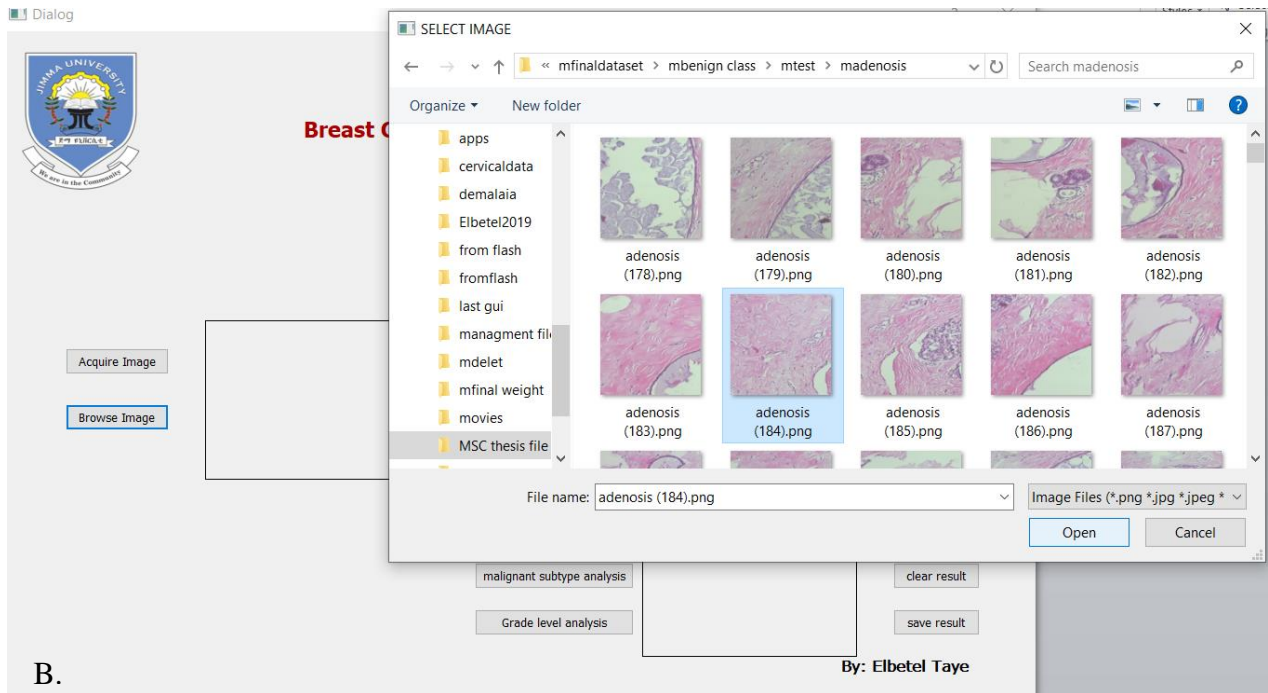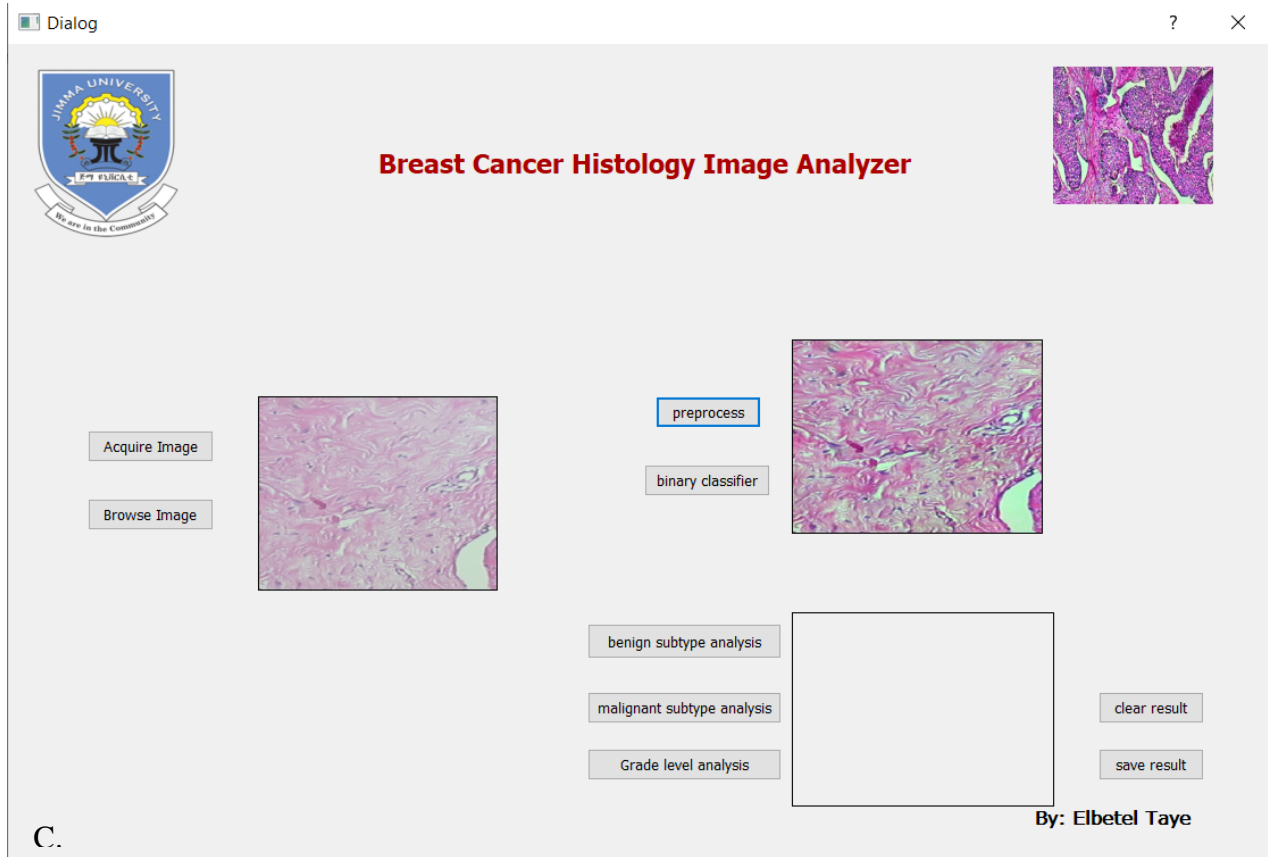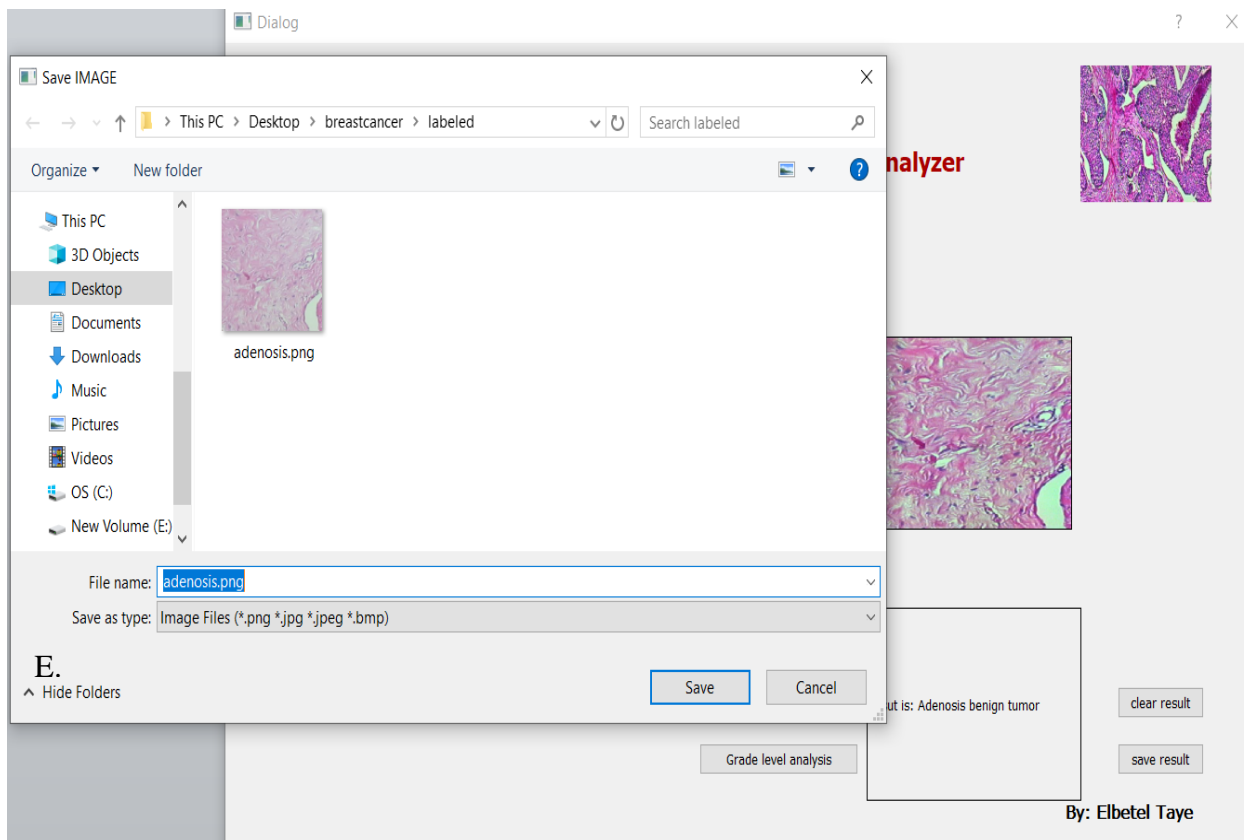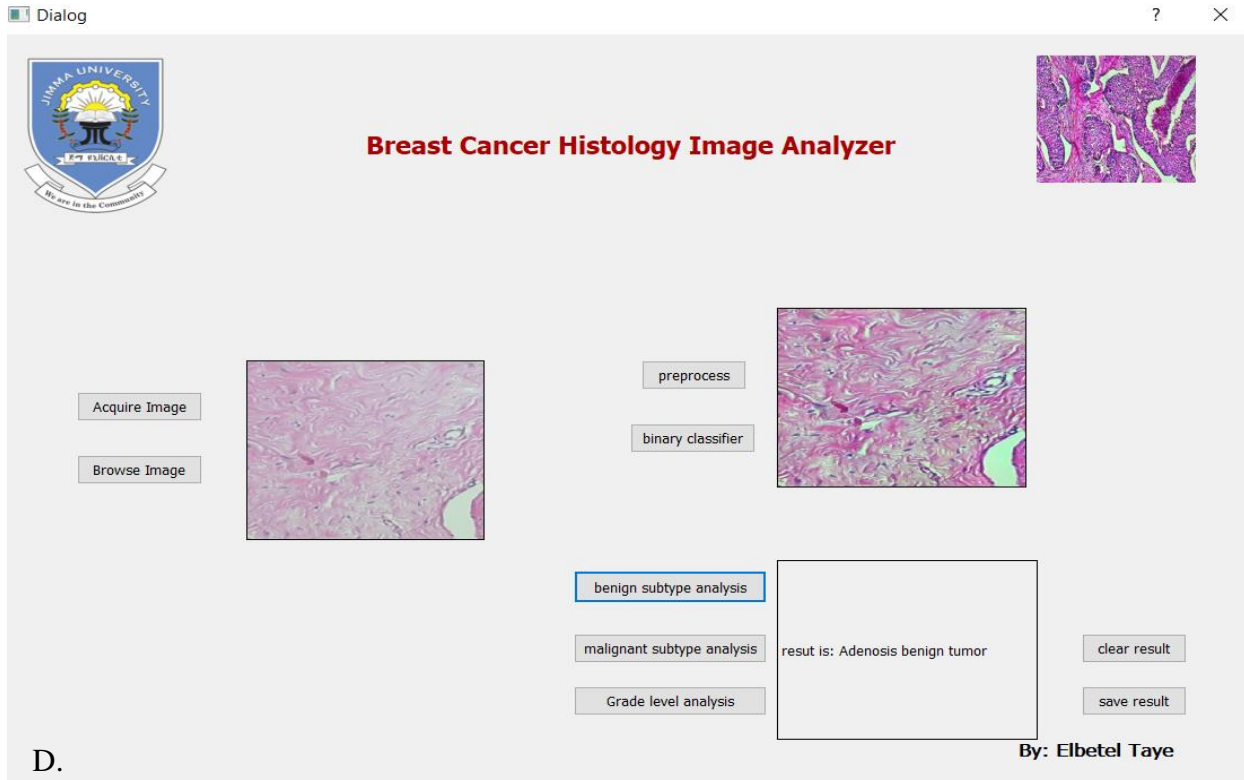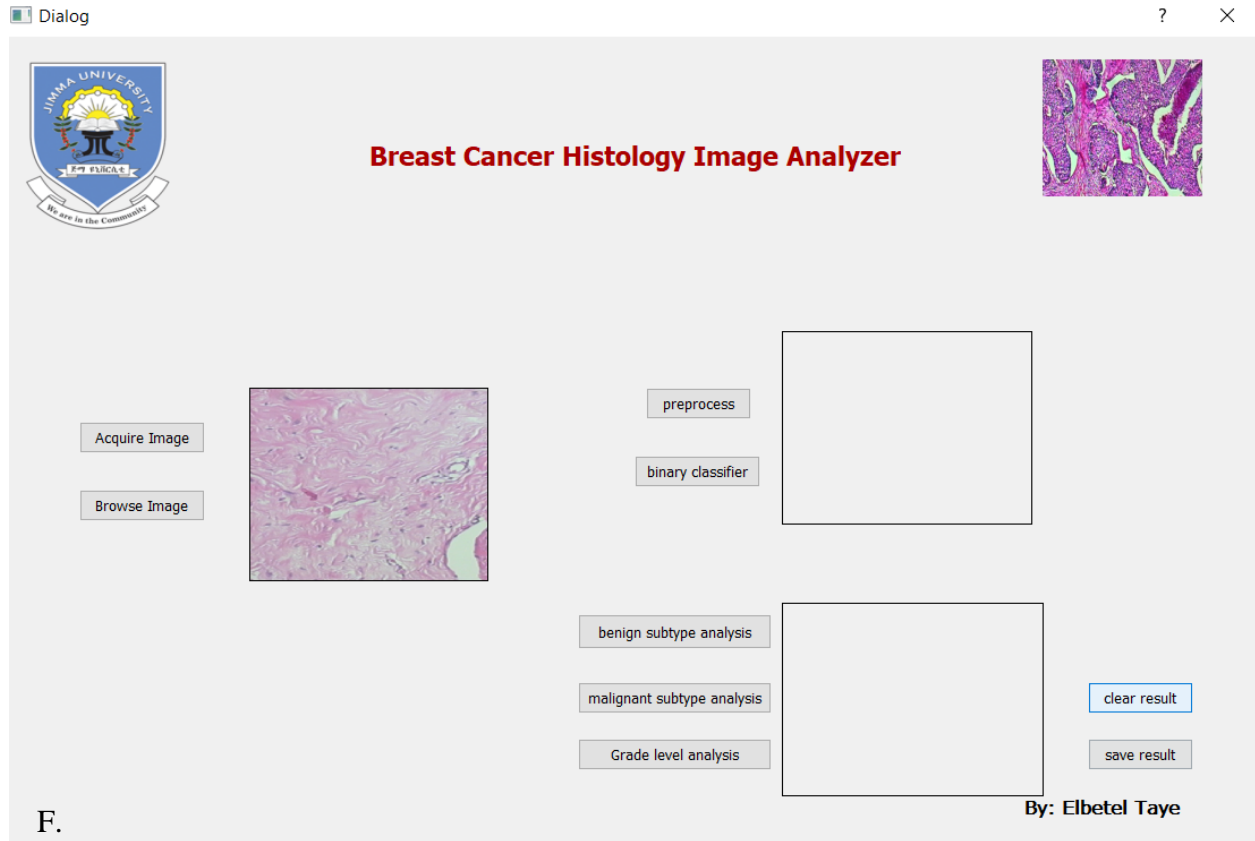
Figure 4-21: Test result of the GUI A) Layout of the GUI, B) Browse image C) Preprocess the image d) Classification result E) save result, F) Clear result

4.1. Discussion

BC is the uncontrolled growth of breast cells. According to world health organization, in 2018, the incidence of BC is ranked 2nd after lung disease and it was estimated to have risen to 18.1 million new cases and 9.6 million deaths globally [1]. Besides, according to the statistical data revealed by the Addis Ababa city cancer Registry Quadrennial report of 2012-2015 investigation, females accounted for two-third of the cancer burden, largely due to BC, which is the leading type of cancer highly affecting the women population in the city followed by cervical cancer [2]. This makes it the number one cancer disease which needs high concern.

BC is diagnosed using triple assessment test, which includes clinical examination, imaging and biopsy test. Of the examining techniques, biopsy test is the gold standard and the most common technique that is used to identify the cancer type and grade before performing a treatment plan

[8]. Unfortunately, there is still a shortage of pathologists in most developing countries including Ethiopia, limiting many breast cancer patients from being diagnosed at the early stage and baring the pathologists to examine a large number of biopsy samples per day making them exhausted. Moreover, the complexity of the cancer cells and their subjective decision, which is dependent on expert's performance, leads to misdiagnosis [12, 13, 14]. As a result, there is a high need of computer aided diagnosing system.

In this study automatic diagnosis of breast cancer from digital histopathological images in terms of binary and multi-class classification of cancer types has been proposed based on a deep learning model. For model training and testing histopathological images were acquired from two online data sets (break-his and the zendo) and local data from JUMC. Preprocessing step includes image resizing and data augmentation along with image enhancement and template matching has been applied to all images.

The aim of preprocessing digital images is to improve the images quality by suppressing unwanted distortions and/or enhancing some important image features, so that the model that will be used for further analysis can take advantage from the improved data. Deep learning models performance is highly dependent on the type of the data and the total number of training data [19]. As the number of data used for training and the quality, the number of features to be learnt by the model increase and boosting the classification performance. In this study, data augmentation, data resize, normalization, histogram equalization and histogram matching were used.

Data augmentation was performed by rotating the images in 90 degree, 180 degree and 270 degree to increase the number of available data without affecting its features. As a result, the total number of data was increased by 4 x (four times). But since we must have equal number of data in each class, the number of images in each class was limited to one thousand seven hundred seventy seven (1776). As a result a total of fourteen thousand two hundred and eight (14,208) images were generated from all the eight subtypes of BC. And one thousand two hundred (1200) images were generated for the grade type of IDC. Next, all image data were resized to 224 x 224 size in order to have a uniform image size and in order to make the input images size appropriate for the classifying model (ResNet50). Finally, by using the holdout method, which is the

simplest kind of cross validation, the whole data was split into 80% training, 10% validating and 10% testing set respectively.

In order to enhance the contrast of the images CLAHE algorithm was applied. The CLAHE method was chosen among other histogram equalization techniques, because of its ability to enhance the contrast of an image by disabling the amplification of noise [44, 45]. In this study a clip limit of two (2) and block size of 8 x8 has been used. These values were the default values for the algorithm in OpenCV. Better histogram equalization is obtained by the method described in Figure 3-4 using the said parameters. The contrast of the image has shown an improvement without loss in information as indicated in the result in Figure 4-6. Besides, the histogram of the images in Figure 4-6 indicates the intensity level distribution by using intensity level versus intensity frequency graph. That is, from the result, it is shown that the intensity of the image was clipped accordingly using the stated value and the clipped intensity was redistributed throughout the image. As a result, an image with a better contrast was achieved.

Moreover, histogram matching was applied on the images by choosing one image as a template image shown in Figure 3-5. This image was chosen by considering the color of the tissue components as per the staining protocol. That is, the cytoplasm part is pinkish in color while the cell nucleus is blue colored. As a result considering the image as a properly stained slide image, it has been taken as a template image. The histogram matching was done to correct the unnecessary color variation that occurred due to difference on the concentration and amount of dye used in the staining process and/or effect of brightness of the ambient light during image acquisition. Different literatures [24, 26, 28, 33 and 46] have used color normalization, histogram matching and random color augmentation to correct the color variation of histopathological images. In this study, after applying histogram matching on the images, the unnecessary color variation was eliminated as shown in Figure 4-6. This prevents the classifying model from being confused by the variable color feature of images in a class.

The major aim of this study was to classify BC histopathological images into the specific type of breast cancer and to determine the grade of invasive ductal carcinoma (IDC). To achieve this, four different classifying models were developed using ResNet 50 model. Better result was achieved by training a fine tuned ResNet 50 model which has an ADAM optimizer, learning rate 0.01 – 0.001, (Rectified Liner Unit) ReLu Activation function, batch size of 100, Loss function

for binary class classification was binary_crosentropy, while for multi class classification it was categorical_crossentropy for fifty (50) epochs.

The learning performance of the model over 50 epochs was evaluated by plotting a learning curve during each training phase as shown in Figure 4-8 to Figure 4-15. During the performance evaluation of the model using learning curve, two different metrics were used. These are the accuracy and loss metrics. These metrics help to see how the model is optimized according to cross-entropy loss and to evaluate the model according to the classification accuracy. Good result is obtained when the accuracy curve increases and the loss curve decreases with increasing number of epochs for both training and validation datasets. As a result both metrics plots were created for each training task. Hence, by reviewing the learning curves shown in Figure 4-8 to Figure 4-15 the performance of the model was evaluated for all classification tasks. As a result, from the plots it is concluded that the model has learned well and has good generalizability performance for all classification tasks.

To classify the subtype of BC, the first step was classifying a given image into BC's binary class (benign or malignant type). This was achieved by using the developed binary class classifying model. The model's performance was tested using a separate dataset and a promising result was achieved as indicated in Table 4-1. Finally average test accuracy of the model for binary classification was calculated. Hence, 96.75 % test accuracy, 96.755 % precision, 96.75 % recall and 96.75 % specificity was achieved. After binary class classification, the identified cancer type has been further classified into their subtypes. This was achieved by using the benign subtype classifying model and malignant subtype classifying model. Similarly, the performance of the developed model for multiclass classification task has been assessed by calculating the average precision, recall (sensitivity), specificity and test accuracy results. Therefore, for the benign type classification, an average accuracy of 96.7 %, average precision of 93.87 %, average recall (sensitivity) of 93.51 % and 97.8 % specificity were achieved. While, for the malignant type classification task, an average test accuracy of 95.78 %, average precision rate of 92.49 %, an average recall (sensitivity) of 91.85 % and 97.17 % of average specificity was achieved. Similarly, satisfactory results were achieved when the models performance were evaluated by using the local data. These indicates that, the classification accuracy of ResNet 50 model was improved comparing it with similar recent studies [28, 31] that have used same model. This

improvement was achieved by the advantage taken from the preprocessing techniques applied on the images.

IDC accounts for 80% of the total occurrence of BC [7]. Grade classification of IDC has been also performed in this study. Likewise, after model training for IDC grade classification, the performance was measured in terms of accuracy, precision, recall and specificity, and a result of 93.86 %, 91.21%, 90.99 % and 95.34 % , respectively, were achieved. This will have significant advantage in aiding physicians to have good treatment plans. The overall summary of the performance of the model (ResNet 50) is summarized in Table 4-4.

Comparing the result of this study with the studies done using the Break-His dataset [15, 25, 28, 29**,** 30, 32**,** 33, 40] as benchmark, we can conclude that, the developed system can classify whole slide breast cancer histopathology images with better classification accuracy. Moreover, the developed system has overcome the gap of further classification of histopathology images in to their sub types. In addition, in case of ductal malignant type of cancer, it has the ability to classify the images into their grade level without being dependent on the magnification factor of the images.

Finally, a GUI is provided in order to make the system user friendly and easily applicable so as to solve the problem on the existing BC diagnosis method. This will assist physicians in making diagnosis decisions and making good treatment plans at the early stage.

4.2. Limitation of the study

In this study, data has been collected from two online data sets and local data from JUMC. But while collecting the local data, it was difficult to find variety of histopathology images having different breast cancer types. This has been one challenge that hinders the study from being trained by variety of data sources.

The other limitation was lack of powerful computational resource that is a computer having higher processor and GPU. This has limited the work from using ensemble techniques (combining different models to have increased classification performance) and models having increased number of layers which enables networks to learn very deep features of the images.

Chapter Five

Conclusion and Recommendation

5.1. Conclusion

This study proposed an automatic system of breast cancer multiclass classification and grade level identification of IDC from histopathological images. A pre-trained ResNet50 model was used to classify whole slide breast cancer histopathological images into benign and malignant (binary classification) and further classifies benign class into four different sub types (adenosis tumor, fibroadenoma tumor, phyllodus tumor and tubularadenoma tumor) and malignant class into other four carcinoma types (ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma). The pre-trained model was used as a feature generator and the extracted features were fed to a soft-max classifier. The result obtained was much better than recently proposed studies. Moreover, in case of malignant carcinoma type the system has the ability to classify given image into its grade level (grade 1, grade 2 and grade 3). Accuracy of 96.75%, 96.7%, 95.78%, and 93.86 % were achieved for binary classification, benign sub-type classification, malignant sub-type classification, and grade identification, respectively.

This developed system can be used as a decision support system in the diagnosis of breast cancer, and this will have a great impact by helping pathologists, especially in those low resource settings where both the expertise and the means is in scarce.

5.2. Recommendation

It can be seen from the literatures that most works are done using specific dataset. Even in this study only small amount of local data was included due to unavailability of similar cases and lack of local data repositories. Therefore, using variety of data source is important in order to have a universal diagnosing system that is not biased to a specific condition like staining procedure.

This study is limited to identifying the grade of ductal carcinoma but it is also important to identify the grade of other malignant cancer types. Moreover, it is limited to classifying breast cancer histology images but such kind of research can be extended to other types of cancers too, this will help to have more impact on saving lives.

As indicated in [30, 33] the effect of using several neural networks can give better prediction result. Therefore it is recommended to combine different neural networks so as to get the advantage of multiple models and increase the classification performance.

Reference

[1] "World Health Organization Releases Latest Global Cancer Data - Cancer Health." [Online]. Available: https://www.cancerhealth.com/article/world-health-organization-releases-latest-global-cancer-data. [Accessed: 15-Dec-2019].

[2] G. Timotewos *et al.*, "First data from a population based cancer registry in Ethiopia," *Cancer Epidemiol.*, vol. 53, pp. 93–98, Apr. 2018.

[3] G. Zucca-Matthes, C. Urban, and A. Vallejo, "Anatomy of the nipple and breast ducts," *Gland Surgery*, vol. 5, no. 1. AME Publishing Company, pp. 32–36, 01-Feb-2016.

[4] "What Is Breast Cancer? | Breastcancer.org," 2018. [Online]. Available: https://www.breastcancer.org/symptoms/understand_bc/what_is_bc. [Accessed: 15-Dec-2019].

[5] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[6] S. R. Lakhani, I. O. Ellis, S. J. Schnitt, P. H. Tan, and M. J. van de Vijver, *WHO classification of tumors of the breast*, 4th editio. International Agency for Research on Cancer (IARC), 2012.

[7] Y. Feng *et al.*, "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis," *Genes Dis.*, vol. 5, no. 2, pp. 77–106, 2018.

[8] Y.-J. Zhang, L. Wei, J. Li, Y.-Q. Zheng, and X.-R. Li, "Status quo and development trend of breast biopsy technology.," *Gland Surg.*, vol. 2, no. 1, pp. 15–24, 2013.

[9] G. Okbazgi, A. Tsegaye, F. Hassen, and M. Negussie, "College of Health Sciences School of Allied Health Science Department of Medical Laboratory Sciences Assessment of Histopathology Services in Ethiopia Addis Ababa University School of Graduate Studies Approved by the Examining Board :," Addis ababa University, 2015.

[10] "Cancer grade and size | Breast Cancer Now." [Online]. Available: https://breastcancernow.org/information-support/facing-breast-cancer/diagnosed-breast-cancer/cancer-grade-size. [Accessed: 16-Dec-2019].

[11] "Tumor Grade Fact Sheet - National Cancer Institute." [Online]. Available: https://www.cancer.gov/about-cancer/diagnosis-staging/prognosis/tumor-grade-fact-sheet. [Accessed: 16-Dec-2019].

[12] "Breast Cancer Misdiagnosis | Diagnosis Delayed." [Online]. Available: http://www.diagnosisdelayed.com/breast-cancer-misdiagnosis/. [Accessed: 16-Dec-2019].

[13] "BreastCancer Screening." [Online]. Available: https://stephensoncancercenter.org/Types-of-Cancer/Breast/NCI-PDQ-Breast/docId/CDR0000062751. [Accessed: 16-Dec-2019].

[14] J. G. Elmore *et al.*, "Evaluation of 12 strategies for obtaining second opinions to improve interpretation of breast histopathology : simulation study," BMJ Publishing Group Ltd., vol. 353, pp. 1–10, 2016.

[15] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.

[16] D. R. Adrian, *Deep Learning for computer vision with python (starter Bundle)*, 1st editio. PYIMAGESEARCH, 2017.

[17] E. Klang, "Deep learning and medical imaging," *J. Thorac. Dis.*, vol. 10, no. 3, pp. 1325–1328, 2018.

[18] A. Maier, C. Syben, T. Lasser, and C. Riess, "REVIEW A gentle introduction to deep learning in medical image processing," *Elsevier*, vol. 29, no. 2, pp. 86–101, 2019.

[19] D. R. Adrian, *Deep Learning for computer vision (Practitioner Bundle)*, 1st Edition. PYIMAGESEARCH, 2017.

[20] B. Kieffer, M. Babaie, S. Kalra, and H. R. Tizhoosh, "Convolutional Neural Networks for

Histopathology Image Classification : Training vs . Using Pre-Trained Networks," *ResearchGate*, no. March 2018, pp. 1–7, 2017.

[21] "Keras : Deep Learning library for Theano and TensorFlow," *Edit on GitHub*. [Online]. Available: keras.io.

[22] Y. Guo, H. Dong, F. Song, C. Zhu, and J. Liu, "Breast Cancer Histology Image Classification Based on Deep Neural Networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10882 LNCS, pp. 827–836.

[23] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10882 LNCS, pp. 737–744, 2018.

[24] T. Araujo *et al.*, "Classification of breast cancer histology images using convolutional neural networks," *PLoS One*, vol. 12, no. 6, pp. 1–14, 2017.

[25] A. Golatkar, D. Anand, and A. Sethi, "Classification of Breast Cancer Histology Using Deep Learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10882 LNCS, no. January, pp. 837–844, 2018.

[26] M. H. Motlagh, M. Jannesari, H. Aboulkheyr, and P. Khosravi, "Breast Cancer Histopathological Image Classification : A Deep Learning Breast Cancer Histopathological Image Classification : A Deep Learning Approach," *ResearchGate*, no. May, 2018.

[27] A. Al Nahid, M. A. Mehrabi, and Y. Kong, "Histopathological breast cancer image classification by deep neural network techniques guided by local clustering," *Biomed Res. Int.*, vol. 2018, 2018.

[28] F. A. Spanhol, P. R. Cavalin, L. S. Oliveira, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," *2017 IEEE Int. Conf. Syst. Man, Cybern. SMC 2017*, vol. 2017-Janua, pp. 1868–1873, 2017.

[29] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018.

[30] J. Xie, R. Liu, J. Luttrell, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Front. Genet.*, vol. 10, no. FEB, pp. 1–19, 2019.

[31] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," *J. Digit. Imaging.*, vol. 32, no. 4, pp. 605–617, 2019.

[32] A. Cruz-Roa *et al.*, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, no. April, pp. 1–14, 2017.

[33] R. Rulaningtyas, A. S. Hyperastuty, and A. S. Rahaju, "Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method," *J. Phys. Conf. Ser.*, vol. 1120, no. 1, pp. 0–8, 2018.

[34] M. Ma, Y. Shi, W. Li, Y. Gao, and J. Xu, "A Novel Two-Stage Deep Method for Mitosis Detection in Breast Cancer Histology Images," *Proc. - Int. Conf. Pattern Recognit.*, vol. 2018-Augus, pp. 3892–3897, 2018.

[35] M. E. Vandenberghe, M. L. J. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak, and C. Barker, "Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer," *Sci. Rep.*, vol. 7, no. January, pp. 1–11, 2017.

[36] H. D. Couture *et al.*, "Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype," *npj Breast Cancer*, vol. 4, no. 1, 2018.

[37] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," *30th AAAI Conf. Artif. Intell. AAAI 2016*, pp. 1160–1166, 2016.

[38] K. Dimitropoulos, P. Barmpoutis, C. Zioga, A. Kamas, K. Patsiaoura, and N.

Grammalidis, "Grading of invasive breast carcinoma through Grassmannian VLAD encoding," *PLoS One*, vol. 12, pp. 1–18, 2017.

[39]   O. Marques, *practical image and video processing using matlab*. Ieee press, 2011.

[40]   R. P. Singh and M. Dixit, "Histogram Equalization: A Strong Technique for Image Enhancement," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 8, no. 8, pp. 345–352, 2015.

[41]   B. Kang, C. Jeon, D. K. Han, and H. Ko, "Adaptive height-modified histogram equalization and chroma correction in YCbCr color space for fast backlight image compensation," *Image Vis. Comput.*, vol. 29, no. 8, pp. 557–568, 2011.

[42]   N. M. Sasi and V. K. Jayasree, "Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images," Engineering, vol.5, pp. 326–331, 2013.

[43]   J. Ma, X. Fan, S. X. Yang, X. Zhang, and X. Zhu, "Contrast Limited Adaptive Histogram Equalization Based Fusion for Underwater Image Enhancement," International Journal of Pattern Recognition and Artificial Intelligence.,Vol. 32, pp. 1–27, 2018.

[44]   M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.

[45]   A. Canziani, A. Paszke, and E. Culurciello, "An Analysis of Deep Neural Network Models for Practical Applications,"Arxiv,abs/1605.07678, May 2016.

[46]   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)*, Las Vegas, NV, 2016, pp. 770-778.

[47]   P. Marcelino, "Transfer learning from pre-trained models - Towards Data Science," 2018. [Online].   Available: https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751. [Accessed: 20-Oct-2019].

[48]   P. Marcelino, "Detailed Guide to Understand and Implement ResNets – CV-Tricks.com," 2018. [Online]. Available: https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751. [Accessed: 20-Oct-2019].

[49]   D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[50]   T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *ResearchGate*, no. March, 2015.

Appendix 1

Source code

December 17, 2019

```
[ ]:
```

```python
[ ]: # this is to do histogram matching
     import glob
     import cv2
     import matplotlib.pyplot as plt
     from PIL import Image
     from skimage.transform import match_histograms
     import numpy as np
     from numpy import *
     import os

     image = cv2.imread('E:\MSC thesis file\delet\mbreast
      ↪cancer\mtempletimage\mductalca(2708).png') #templet image
     print(image.dtype)
     filenames = glob.glob("...\*.png") #directory of the input images
     filenames.sort()
     images = [cv2.imread(img) for img in filenames]

     matched = []

     for img in images:
         matched.append(match_histograms(img, image, multichannel=True))
         os.chdir(' ... ')

         length = len(matched)
         print(length)

         for i in range(length):
             cv2.imwrite("..."+str(i)+".png ", matched[i]) #matched image will be
      ↪saved
             fig, ax = plt.subplots(nrows=1, ncols=2, figsize=(8, 5),
                         sharex=True, sharey=True)
             ax[0].imshow(image)
             ax[0].axis('off')
```

```
        ax[0].set_title('reference')
        ax[1].imshow(matched[i])
        ax[1].axis('off')
        ax[1].set_title('matched')
        fig.tight_layout()
        plt.show()
```

[ ]:

```
#this is to do contrast adaptive histogram equalization
import glob
import cv2
import matplotlib.pyplot as plt
from PIL import Image
from skimage.transform import match_histograms
import numpy as np
from numpy import *
import os
filenames = glob.glob(" ... ")  # give the directory to the input images
filenames.sort()
images = [cv2.imread(img) for img in filenames]
equalized = []
for img in images:
    img_y_cr_cb = cv2.cvtColor(img, cv2.COLOR_BGR2YCrCb)
    y, cr, cb = cv2.split(img_y_cr_cb)
    clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(8,8))
    output2 = clahe.apply(y)
    output3 = clahe.apply(cr)
    output4 = clahe.apply(cb)
    img_y_cr_cb_eq = cv2.merge((output2, output3, output4))
    equalized.append(cv2.cvtColor(img_y_cr_cb_eq, cv2.COLOR_YCR_CB2BGR))
    os.chdir('...') # os.chdir is used to change the working directory to
   ↪the path where the processed image will be saved.
    length = len(equalized)
    print(length)
    for i in range(length):
        cv2.imwrite("..."+str(i)+".png ", equalized[i]) # will save the
   ↪processed image
```

[ ]:

```
#for training the model
# importing libraries
from tensorflow.python.keras.applications import ResNet50
from tensorflow.python.keras.models import Sequential
```

```python
from tensorflow.python.keras.layers import Dense
from tensorflow.python.keras import optimizers
from keras.applications.resnet50 import preprocess_input
from keras.preprocessing.image import load_img
from keras.preprocessing.image import ImageDataGenerator
from tensorflow.python.keras.callbacks import EarlyStopping, ModelCheckpoint
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import cv2
import os

#declaring variables
num_classes =4   #indicates the number of catagories
channels = 3   # indicates if the image is colour image or grey level image
image_resize =224
resnet50_pooling_average = 'avg'
objective_function = 'catagorical_crossentropy'
loss_metrics =['accuracy']
num_epochs = 10
early_stop_patience = 3
batch_size_training = 100
batch_size_validation =100

#model intialization
model = Sequential()
resnet_model = ResNet50(weights = 'imagenet')
model.add(ResNet50(include_top = False, pooling = 'avg', weights =
 'imagenet' ))
model.add(Dense(num_classes, activation = 'softmax'))
model.layers[0].trainable = False
model.summary()
opt = optimizers.adam()
model.compile(optimizer = opt, loss = 'categorical_crossentropy', metrics =
 ['acc'])


image_size = image_resize


train_datagen = ImageDataGenerator(
        preprocessing_function=preprocess_input)  #normalize the input image
 by bring it to center value
#generating training and validation inputs to be fed for the neural network
train_generator = train_datagen.flow_from_directory(
        'E:\MSC thesis file\mtypeclahe\mhisclachebenigntra',
        target_size=(224, 224),
        batch_size=batch_size_training,
```

```python
        class_mode='categorical')

validation_generator = train_datagen.flow_from_directory(
        'E:\MSC thesis file\mtypeclahe\mhistclahebenign',
        target_size=(224, 224),
        batch_size=batch_size_validation,
        class_mode='categorical')

(batch_size_training, len(train_generator),batch_size_validation,␣
 ↪len(validation_generator))

cb_early_stopper = EarlyStopping(monitor = 'val_loss', patience =␣
 ↪early_stop_patience)
cb_checkpointer = ModelCheckpoint(filepath = 'E:\MSC thesis␣
 ↪file\mtypeclahe\mjupresbenignclaheweight2.h5', monitor = 'val_loss',␣
 ↪save_best_only = True, mode = 'categorical')

#train the model
fit_history = model.fit_generator(
        train_datagen.flow_from_directory(
        'E:\MSC thesis file\mtypeclahe\mhisclachebenigntra',
        target_size=(image_size, image_size),
        batch_size=batch_size_training,
        class_mode='categorical'),
        steps_per_epoch= 142,
        epochs = 50,
        validation_data=train_datagen.flow_from_directory(
        'E:\MSC thesis file\mtypeclahe\mhistclahebenign',
        target_size=(image_size, image_size),
        batch_size=batch_size_validation,
        class_mode='categorical'),
        validation_steps= 17,
        callbacks=[ModelCheckpoint(filepath = 'E:
 ↪\mforcoding\colournormalizedgradeimage\mresbenignclaheweight2.h5', monitor␣
 ↪= 'val_loss', save_best_only = True, mode = 'categorical'),␣
 ↪EarlyStopping(monitor = 'val_loss', patience = early_stop_patience)]
)

# during training, the callbacks help to save only the best resulted weight␣
 ↪value
model.save_weights('mjupresbenignclaheweight2.h5')  #saves weight value
model.save('mjupodel_kerasacbr.h5')   # saves model with its weight value

#plot the result
print(fit_history.history.keys())
plt.figure(1, figsize = (15,8))
```

```python
plt.subplot(221)
plt.plot(fit_history.history['acc'])
plt.plot(fit_history.history['val_acc'])
plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'valid'])

plt.subplot(222)
plt.plot(fit_history.history['loss'])
plt.plot(fit_history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'valid'])

plt.show()
```

```python
[ ]: #for Testing trained model
    # importing saved model
    import tensorflow as tf
    from tensorflow.keras.applications import resnet50
    from keras.models import load_model
    model = tf.keras.models.load_model('model_kerasgradefinal.h5')
    model = tf.keras.models.load_model('model_kerasgradefinal.h5')
    model.summary()

    # prediction of grade set
    from keras.applications.resnet50 import ResNet50
    from keras.preprocessing import image
    from keras.applications.resnet50 import preprocess_input, decode_predictions
    import numpy as np
    import os
    import cv2 as cv
    from PIL import Image
    from numpy import *
    from cv2 import *
    import glob

    class_names = ['Grade1', 'Grade2', 'Grade3'] # label
    train_datagen = ImageDataGenerator(
            preprocessing_function=preprocess_input)


    test_generator = train_datagen.flow_from_directory(
```

```
        'E:\MSC thesis file\delet\previous␣
 ↪datasets\mforcoding\TRUEDATASETS\MATCHEDGRADE\mtest\grade1',
        target_size=(224, 224),
        batch_size=batch_size_validation)


model.evaluate_generator(test_generator, steps=None, max_queue_size=10,␣
 ↪workers=1, use_multiprocessing=False)
pred = model.predict_generator(test_generator, steps=None,␣
 ↪max_queue_size=10, workers=1, use_multiprocessing=False, verbose=1)
predicted = np.argmax(pred, axis=1)


print('Confusion Matrix')
cm = confusion_matrix(test_generator.classes, np.argmax(pred, axis=1))
plt.figure(figsize = (30,20))
sn.set(font_scale=1.4) #for label size
sn.heatmap(cm, annot=True, annot_kws={"size": 12}) # font size
plt.show()
print()
```