MAMMOGRAM BREAST MASS CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

MOHD NAFIE BIN MASLAN

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

MAMMOGRAM BREAST MASS CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

MOHD NAFIE BIN MASLAN

RESEARCH REPORT SUBMITTED PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BIOMEDICAL ENGINEERING

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MOHD NAFIE MASLAN**

Matric No: **S2015935/1**

Name of Degree: **MASTER OF BIOMEDICAL ENGINEERING**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

MAMMOGRAM BREAST MASS CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

Field of Study:

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 10 / 01 / 2022

Subscribed and solemnly declared before,

Witness's Signature Date: 10/1/2022

Name:

Designation:

MAMMOGRAM BREAST MASS CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

ABSTRACT

The aim of this study is to investigate a deep convolutional neural network (DCNN) image classification technique to classify breast mass (i.e., benign or malignant) on mammogram image. DCNN architecture model of AlexNet, GoogLeNet and VGG-16 was compared to evaluate the performance of classifying breast mass patch. The dataset of breast mass was obtained from public database, Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). Transfer learning which pre-trained the DCNN model on large-scale ImageNet, and data augmentation to increased number of data technique, were applied in this study. In total, 13,260 images of training dataset were imported into the NVIDIA deep learning GPU training system (DIGITS). A trained model of AlexNet, GoogLeNet and VGG-16 was created and selected based on the high accuracy and low loss. 540 images of testing dataset were applied to the trained DCNN model. The classification performance between the prediction and actual images were evaluated using confusion matrix and receiver operating characteristic (ROC) curve, to calculate accuracy, sensitivity/ recall, precision, specificity, F1-score and area under the curve (AUC). As a result; AlexNet and GoogLeNet have the highest specificity (71.48%); GoogLeNet have the highest accuracy (73.89%) and precision (72.79%); VGG-16 have the highest sensitivity/ recall (80%) and F1-score (75.26%). VGG-16 (0.8207) has the highest AUC, trailed by GoogLeNet (0.8064) abd AlexNet (0.7601). As conclusion, the overall ability of VGG-16 and GoogLeNet was superior over AlexNet in discriminated benign and malignant of breast mass patch.

Keywords: deep convolutional neural network (DCNN), mammogram breast mass, image classification, transfer learning, data augmentation.

KLASIFIKASI JISIM PAYUDARA MAMMOGRAM MENGGUNAKAN RANGKAIAN NEURAL KONVOLUSI DALAM

ABSTRAK

Matlamat kajian ini adalah untuk menyiasat teknik pengelasan imej menggunakan *deep convolutional neural network* (DCNN), untuk mengklasifikasikan jisim payudara pada imej mammogram. Model DCNN iaiu AlexNet, GoogLeNet dan VGG-16 dibandingkan untuk menilai prestasi mengelaskan tompok jisim payudara. Dataset jisim payudata diperolehi daripada pangkalan data awam, *Curated Breast Imaging Subset of Digital Database for Screening Mammography* (CBIS-DDSM). Teknik pemindahan pembelajaran yang melatih model DCNN dengan skala besar ImageNet, dan penambahan data dengan meningkatkan bilangan data, telah digunakan dalam kajian ini. Secara keseluruhan, 13,260 imej set data latihan telah diimport ke dalam NVIDIA *deep learning GPU training system* (DIGITS). Model terlatih AlexNet, GoogLeNet dan VGG-16 telah dicipta dan dipilih berdasarkan ketepatan tinggi dan rendah kehilangan. 540 imej set data ujian digunakan pada model DCNN terlatih. Prestasi pengelasan antara imej ramalan dan sebenar dinilai menggunakan *confusion matrix* dan lengkung ROC, untuk mengira ketepatan, kepekaan, kejituan, kekhususan, skor F1 dan AUC. Keputusannya, AlexNet dan GoogLeNet mempunyai kekhususan (71.48%) tertinggi; GoogLeNet mempunyai ketepatan (73.89%) dan kejituan (72.79%) tertinggi; VGG-16 mempunyai kepekaan (80%) dan skor F1 (75.26%) tertinggi. VGG-16 (0.8207) mempunyai AUC tertinggi, diikuti dengan GoogLeNet (0.8064) dan AlexNet (0.7601). Kesimpulannya, keupayaan keseluruhan VGG-16 dan GoogLeNet adalah lebih baik berbanding AlexNet dalam pengelasan *benign* dan *malignant* pada jisim payudara.

Kata kunci: *deep convolutional neural network*, jisim payudara mammogram, klasifikasi imej, pemindahan pembelajaran, penambahan data.

ACKNOWLEDGEMENTS

Bismillahirrahmanirrahim,

In the name of Allah the Merciful, I would like to express my thanks for the blessings in giving me the strength, health and ability to complete my research project report and consequently complete my study, I owe it all for His blessings.

First, I would like to express my utmost gratitude to my supportive supervisor, Ir. Dr. Khairunnisa binti Hasikin in favor of her guidance to complete this research project report entitled “Mammogram Breast Mass Classification Using Deep Convolutional Neural Network”. I am grateful to work under her supervision since her enthusiasm, encouragement and critique, motivates me to complete my research within planned time. I am truly blessed that I got her as my supervisor. Also, special thanks to Ir. Dr. Lai Khin Wee, who is Programme Coordinator at that time, for guiding and proposing method of artificial intelligence in biomedical study, which given me some time to do some research before the actual project.

A special mentioned to my beloved wife, Nor Faisza binti Ismail for continuous support and encouragement during this study. I am grateful that this study creates awareness about the cancer since my wife have been diagnosed with breast cancer during my research. Alhamdullillah, she is recovering well and I praise to the Almighty for giving me a chance to live a little longer with her. Also, to my wonderful children, Fatin Hannani, Faiqul Haris and Faiqul Hadif, who is being supportive during the study.

My deepest gratitude also goes to my parent, Dr. Hj. Maslan and Hjh. Sakdiah, and parent-in-law, Ismail and Sharifah, who gives me so much love and support throughout the study. Finally, I would like to express my thanks to my siblings and friends who helped me during the process of completing this programme.

Ir. Ts. Mohd Nafie bin Maslan, Jan 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

3D	:	Three-Dimensional
AI	:	Artificial Intelligence
Adam	:	Adaptive Moment Estimation
AUC	:	Area Under the Curve
BI-RADS	:	Breast Imaging Reporting and Data System
BCDR	:	Breast Cancer Digital Repository
CAD	:	Computer-Aided Detection or Diagnosis
CBIS-DDSM	:	Curated Breast Imaging Subset of Digital Database for Screening Mammography
CC	:	Craniocaudal
CNN	:	Convolutional Neural Network
CT	:	Computerized Tomography
DCIS	:	Ductal Carcinoma in Situ
DCNN	:	Deep Convolutional Neural Network
DICOM	:	Digital Imaging and Communications in Medicine
DIGITS	:	Deep Learning GPU Training System
EDA	:	Exploratory Data Analysis
FN	:	False Negative
FP	:	False Positive
GAN	:	Generative Adversarial Network
GPU	:	Graphic Processing Unit
IARC	:	International Agency for Research on Cancer
IBC	:	Inflammatory Breast Cancer
IDC	:	Invasive Ductal Carcinoma

ILC	:	Invasive Lobular Carcinoma
ILSVRC	:	ImageNet Large-Scale Visual Recognition Challenge
JPEG	:	Joint Photographic Experts Group
LCIS	:	Lobular Carcinoma in Situ
MIAS	:	Mammography Image Analysis Society
MLO	:	Mediolateral Oblique
MRI	:	Magnetic Resonance Imaging
PET	:	Positron Emission Tomography
PReLU	:	Parametric Rectified Linear Unit
ReLU	:	Rectified Linear Unit
RNN	:	Recurrent Neural Network
ROC	:	Receiver Operating Characteristic
ROI	:	Region of Interest
RReLU	:	Randomized Rectified Linear Unit
SGD	:	Stochastic Gradient Descent
SVM	:	Support Vector Machines
TCIA	:	The Cancer Imaging Archive
TN	:	True Negative
TP	:	True Positive
VGG	:	Visual Geometry Group
VOI	:	Volume of Interest
WHO	:	World Health Organization

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background

Breast cancer is the most prevalent cancer, responsible for the fifth leading cause of cancer deaths among women globally (Sung et al., 2021). According to World Health Organization (WHO), there were around 2.3 million new cases of women diagnosed with breast cancer, and 685,000 deaths globally in 2020. As of 2020, there are over 7.8 million women living with breast cancer globally in the last 5 years ("Breast Cancer," 2021).

In Malaysia, 1 out of 19 women are at risk with breast cancer. Based on National Cancer Registry from 2003 to 2005, Chinese women seemed to be most vulnerable with a rate of 59.7 per 100,000, followed by Indian women (55.8 per 100,000) and Malay women (33.9 per 100,000) (Lim, Rampal, & Yahaya, 2008). The death result from breast cancer is due to the lack of early detection. Early detection of breast cancer increases the life expectancy of women diagnosed with it. An effective diagnostic approach is critical for detecting breast cancer to improve the survival rate through an appropriate treatment resulting reduction of mortality rates.

For diagnosis and early detection, many imaging modalities are used to visualize the breast such as ultrasound (sonography), mammograms (x-rays), magnetic resonance imaging, etc. Mammography is commonly used method to diagnose for breast cancer in clinical practice. As for ultrasound, the imaging modalities employs acoustic waves to view the women’s breast with no ionizing radiations, but the disadvantage is that it lacks specific information which disable microcalcification detection. Microcalcification is required to diagnose such malignancy in timely manner. Microcalcification in breast cells is characterized by a little deposit of calcium that difficult to detect but can be quickly detected using imaging modalities. Other imaging modalities such as magnetic resonance

imaging (MRI), radiologists often used it to confirm the presence of mass in the breast. However, the disadvantage is that patient may experience an irrational fear of confined spaces or claustrophobia. Due to the disadvantage of other imaging modalities, mammography is the most popular imaging method to screen breasts, and detect cancer due to cost effectiveness and rapid acquisition for large population.

Mammography is a non-invasive and specialized imaging modality for breast image. In this approach, the images usually obtained using two different views which are craniocaudal (CC), and mediolateral oblique (MLO) views. Mammography is extensively used due to lower x-ray energy to detect microcalcification in the breast. The main benefit of the mammography imaging technology is that it can detect cancer even before the patient can detect physically. The nature of abnormality identified within the breast is determined by the degree of microcalcification which can be benign or malignant. The benign type is usually non-cancerous, and will not infiltrate nearby tissue. However, the malignant form is commonly referred to as cancerous since it will force infiltrate and spread to nearby tissue. As a result, it is necessary to be able to distinguish the different type of severities to improve the breast cancer detection.

In order to be able to detect, distinguish, and diagnose the breast cancer accurately (i.e., benign or malignant), radiologist face challenges due to vast number of breast images where they have to examine or analyze the image daily, and the difficulty of reading the images for detecting the breast tumours which may appeared as microcalcification or distortion, and masses on mammograms image. The tumours masses can be readily disguised by overlaying with the patient's dense breast tissue which make it difficult to detect them. Therefore, the procedure of manual reading is required but this method is time consuming, and use many resources. The manual classification of vast number of breast images by radiologist can cause visual fatigue which may easily make error and result in high recall rate for additional screening. Furthermore, radiologist

require years of experience to detect and diagnose breast cancer accurately, especially for breast masses that are not obvious in early stage. So, it is necessary to use other techniques that will help radiologist to make their diagnosis since the classification into malignant and benign of breast tumours is a difficult task.

In order to overcome such constraints of breast tumours diagnosis, deep learning methods shows a superior classification accuracy for screening of breast mammograms compared with conventional computational imaging methods. Deep learning which is a subset of machine learning and artificial intelligence, refers to complex multi-layered statistical learning methods for extracting representation from multiple layers of input data, which are not user-defined but learnt directly from the data by the algorithm. It is become a common that machine learning technique have the capacity in nonlinear modelling tasks such as feature extraction, and classification process from large dataset. In recent years, many deep learning techniques have been studied for classification tasks, and cancer diagnosis including convolutional neural network (CNN).

Deep learning with convolutional neural network, or also known as deep convolutional neural network (DCNN), is a multi-layer neural networks that are meant to extract features that represent contexts of images with minimal preprocessing. Because of this benefit, DCNN has become the most used method of interpreting images. Furthermore, DCNN has the ability to performed various task such as image classification, object detection, and semantic segmentation which can apply to medical imaging fields. DCNN gained a lot of attention in recent years, and many models of DCNN have been used for cancer diagnosis.

The purpose of this study was to evaluate the performance of a deep convolutional neural network (DCNN) image classification technique for breast tumours (i.e., benign or malignant) on mammography image. The method was using DCNN model for image

classification such as AlexNet, GoogLeNet, and VGG16 on NVIDIA deep learning GPU training system (DIGITS).

1.2 Problem Statements

Diagnosing breast tumours from mammogram images is a challenging task for radiologist. Breast cancer are hard to identify and radiologist require years of experience to detect accurately. They need to manual read a large number of images which can cause visual fatigue and making error in diagnosing the tumours. Furthermore, breast tumours hard to quantify since mammogram limit the image contrast between the tumours and their surrounding areas. It is also hard to locate the specific features of the breast tumours since they have complex characteristic. Therefore, medical image analysis study was required for image classification of breast tumours.

At the beginning of the study for medical image analysis, the researchers attempted to apply the DCNN directly on the whole breast mammogram for image classification of breast tumours. However, the size of the image usually large around 3500 x 5500 pixels for whole breast mammogram compared to the input image of DCNN which is usually small around 256 x 256 pixels. Large size mammogram needs to be resized into small size resulting many valuable features lost. Due to the reduction of the image, the breast tumours may become invisible which can severe the classification effectiveness and performance of the model. So, classification using whole breast mammogram is much more challenging, and study by Zhou, Zaninovich, and Gregory (2017), the best performance for the whole breast mammogram classification is 60.90% using DCNN.

In order to enhance the use of whole breast mammogram, some researcher cropped the tumour patches known as region of interest (ROI) from the image that classify as benign or malignant of breast cancer. An ROI is the location of the breast tumour and usually based on clinical information. The performance of classification is improved by using

different DCNN model and training strategies. From the ROI, more features can be extracted compare to whole breast mammogram which enhance the network performance. But the information that can be extracted from single image is usually limited. So, some scientist attempts to enhance the performance of classification by extracting features from multiple ROI. In general, the classification of mammograms using ROI or patches of abnormalities yields good results but necessitates substantial preprocessing work.

In this study, various deep convolutional neural network (DCNN) models such as AlexNet, GoogLeNet, and VGG-16 shall be used, and evaluated for region of interest (ROI) mammogram using NVIDIA deep learning GPU training system (DIGITS). The technique of data augmentation, and transfer learning will be explored to optimize the classification performance.

1.3 Aim and Objectives

The aim of the study is to investigate a deep convolutional neural network (DCNN) method using NVIDIA Deep Learning GPU Training System (DIGITS), and DCNN model to classify breast tumours on mammogram image. The specific study objective include: -

i. To classify breast tumours (i.e., benign or malignant) from mammogram image using DCNN architecture model.
ii. To compare the performance between deep convolutional neural network models in classifying breast masses.
iii. To optimize the classification performance using the analyses attained from (i) and (ii).

1.4 Scope of Study

The scope of the study focuses on the development of learning model, and evaluate the performance of a deep convolutional neural network (DCNN) of a deep learning image classification technique for classification of breast tumours (i.e., benign or malignant) on mammogram image. The study also features on technique combined with DCNN model to optimize the classification performance includes transfer learning, and data augmentation.

The study begins with an extensive literature review on breast anatomy, breast cancer symptom, types, stage, test and treatments. The reviews also include breast cancer detection using imaging diagnostic test and classification of mammogram image with convolutional neural network. Reviews on method and technique to improve and optimize the performance of the classifiers also conducted.

The training procedures for image classification performed with use of DCNN model which is AlexNet, GoogLeNet, and VGG16 which implemented with the NVIDIA DIGITS deep learning training system version 6.1.1 on the Caffe framework. The training and test images imported into the DIGITS library as dataset. NVIDIA DIGITS will be used to train the network to get the learning/ training model.

The proposed system is developed and tested on a machine with Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz, 16GB RAM and graphic card NVIDIA GeForce GTX 1060 with 6GB memory.

1.5 Organization of Thesis

The thesis consists of five chapters and the outline of the thesis is structured as follows:-

- Chapter 1 – Introduction

  This chapter presents an overview of detecting, and diagnose of breast cancer using mammogram image with the aid of deep learning with DCNN. Problem statement, aim, objective, and scope of the study are presented in this chapter.

- Chapter 2 – Literature Review

  This chapter presents an extensive review on breast cancer, and deep learning classification method. It explained about the breast anatomy, and physiology. This chapter also explained about the breast cancer symptom, types, stage, tests, treatments, and various of breast cancer imaging diagnostic test. Other than that, this chapter includes the analysis of convolutional neural network model performance to classify the breast tumours, and distinguish between benign and malignant using mammogram image. This includes the review on the accuracy of image classification of breast tumours. This chapter also review the technique to optimize, and improve the classification of breast tumours.

- Chapter 3 – Methodology

  This chapter presents the overall approach used in this study to perform the development of learning model using DCNN model and NVIDIA DIGITS. The technique to improve the image classification also highlighted in this chapter. Other than that, preparation of the dataset or image preprocessing, hardware specification and architecture of DCNN is also be discussed in this chapter.

- Chapter 4 – Results and Discussion

  The results and discussion of the image classification of different DCNN model and performance are discussed. The effectiveness of technique to improve the performance of classification is evaluated.

- Chapter 5 – Conclusion

  Finally, the last chapter draws the conclusion and highlight the overall performance, and accuracy of image classification of breast tumours using mammogram images. As future works, several methods and topics to be explored are outlined.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Cancer is a severe public issue and the most significant cause of death in the world. Cancer, also known as neoplasms or malignant tumours, refers to a general term of diseases that can infect any part of the body. One distinguishing aspect of cancer is the rapid formation of abnormal cells that expand beyond their normal borders thus infect adjacent sections of the body and migrating to other organs. This process which is known as metastasis are the cause of mortality due to cancer. **Figure 2.1** and **Figure 2.2** shows the most common new cases, and the most common causes of death due to cancer in 2020 by the World Health Organization (WHO) website.

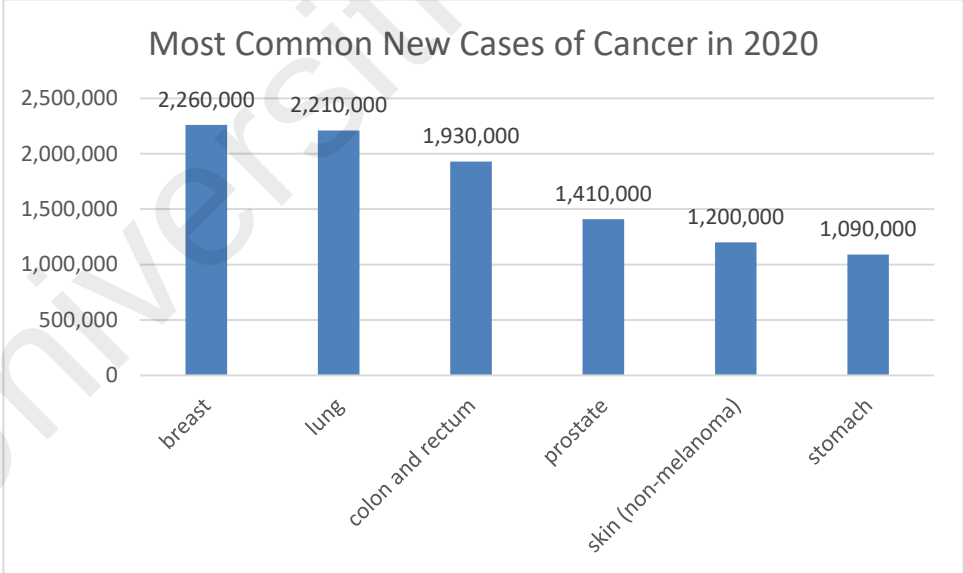


Figure 2.1: Most common new cases of cancer in 2020

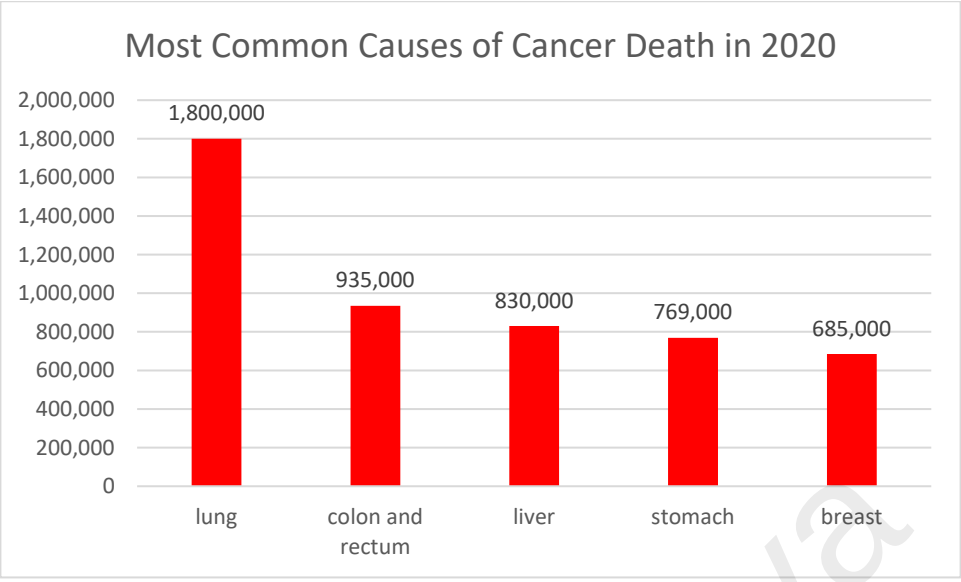


Figure 2.2: Most common causes of cancer death in 2020

In December 2020, International Agency for Research on Cancer (IARC) reported that breast cancer is the highly prevalent diagnosed cancer globally, overtook lung cancer from previous year ("Breast cancer now most common form of cancer: WHO taking action," 2021). The most diagnosed cancer is female breast cancer (11.7%), with an estimated 2.3 million new cases, follow by lung cancer (11.4%), colorectal cancer (10%), prostate cancer (7.3%) and stomach cancer (5.6%) (Sung et al., 2021). However, the leading cause of cancer death are still lung cancer (18%), with an estimated 1.8 million deaths, followed by colorectal cancer (9.4%), liver cancer (8.3%), stomach cancer (7.7%) and female breast cancer (6.9%).

Breast cancer incidence rates are rapidly increasing in transitional nations such as Asia, Africa and South America and high-income Asian countries such as Japan and South Korea, where rates have previously been low. In Malaysia, breast cancer has a high incidence risk for women and is the highly prevalent cancer among women of all ethnic groups. The overall lifetime risk in Malaysia is 1 in 27 women which, Chinese is 1 in 22, Indians is 1 in 23, and Malays is 1 in 30 (Azizah et al., 2019). The increased cases are due

to drastic changes in sociocultural, lifestyle and environment. Most of the women that participated in industrial workforce have impact in breast cancer risk factors such as temporary suspension bearing children, lower birth rate (less breastfeeding), insufficient physical activity, obesity due to unhealthy diet, harmful use of alcohol, use of tobacco and breath air pollution. However, cancer may be averted between 30% to 50% by avoiding risk factors and can also decrease by early detection and applying adequate treatment for patients that developed cancer. Cancers are most likely curable once they are early detected and adequately treated.

Early detection comprised of two methods: early diagnosis and screening. Early diagnosis is being able to recognize the symptoms of various cancer, having access to clinical examination for diagnosis, and treatment to be made as soon as possible. Meanwhile, screening aims to detect patients who have the results that are indicative of certain cancer or pre-cancer before they develop symptoms. Screening techniques include mammography screening for breast cancer settings.

As recommended by WHO, women between the aged of 50 to 69 years who has an average risk of developing breast cancer, shall have mammography screening for every two years. American Cancer Society guideline recommends that women between the aged of 45 to 54 years begin screening annually and above 55 years to screen every two years or continue screening annually. It is also recommended that women between the aged of 40 to 44 years to take the benefit to initiate screening annually.

However, there are drawbacks to mammographic screening including as overdiagnosis or overtreatment. Therefore, there are potential to increase screening effectiveness by using the assistance of artificial intelligence to predict the classification of cancer by using mammography images. Artificial intelligence (AI) technology has becoming popular due to advanced processing power, a vast amount data and innovative algorithm. Machine

learning is a part of artificial intelligence that used the data itself to learn the classification categories or forecast the future or uncertain situations with minimum human intervention. Since machine learning is data-driven learning, it is able to anticipate unseen and unknown data. In this case, machine learning able to do the tasks such as regression, semantic segmentation, object detection, image classification, and etc. Machine learning dataset are generally consisting of training, testing and validation datasets. It learns the features of the data through training dataset and validate the features through validation dataset. The accuracy of machine learning can be check by testing dataset (Kim et al., 2019).

Deep learning is a specialized form of machine learning, performed “end-to-end learning”, which a neural network learns automatically to perform task, such as image classification, object detection, and segmentation, by using raw dataset. Deep learning models trained using a vast amount set of labeled data and neural network architecture that learn to extract features, and process modelling directly from the data. Compared with machine learning, the features and classifier of images are manually extracted. The significant benefit of deep learning is the networks continue to improve as the amount of data increases. Due to the increased public availability of datasets and growth in computational power, deep learning is widely used in research fields especially in medical imaging processing.

There are 3 types of deep learning system: the most common is supervised deep networks, unsupervised deep networks, and hybrid deep networks. Deep learning networks algorithms consists of convolutional neural network (CNN), generative adversarial network (GAN) and recurrent neural network (RNN). Among the deep learning algorithms, CNN is the most commonly used by the researcher as the backbone

of neural network architecture and has achieved interpreting medical images efficiently, similar with human experts.

The remainder of this chapter is dedicated to the review of breast anatomy and breast cancer symptoms, types, stage, tests and treatments. Section 2.3 reviews the diagnostic tests using imaging modalities available to detect breast cancer. This section elaborates on the imaging modalities advancement and limitations to diagnose breast cancer. Detailed findings on mammogram diagnostic tests also reviewed. Reviews on artificial intelligence using machine learning and deep learning, in aiding the classification of breast tumour are presented in section 2.5. In addition, image classification of breast cancer and optimization method also reviewed. Finally, the summary of reviews in this chapter concludes at Section 2.8.

2.2 Breast Cancer

2.2.1 Female Breast Anatomy and Physiology

Breasts are the organs for both male and female sexual anatomy. Female breasts have a function for lactation with milk ducts and glandular tissue, meanwhile male breasts don't serve any purpose. The nipples and areolae are the components that are visible on breast anatomy, which connects with many nerves that increase sexual pleasure. Breast cancer may affect both men and female. However, women are likely to develop benign (non-cancerous) and malignant (cancerous) breast illnesses compared to males.

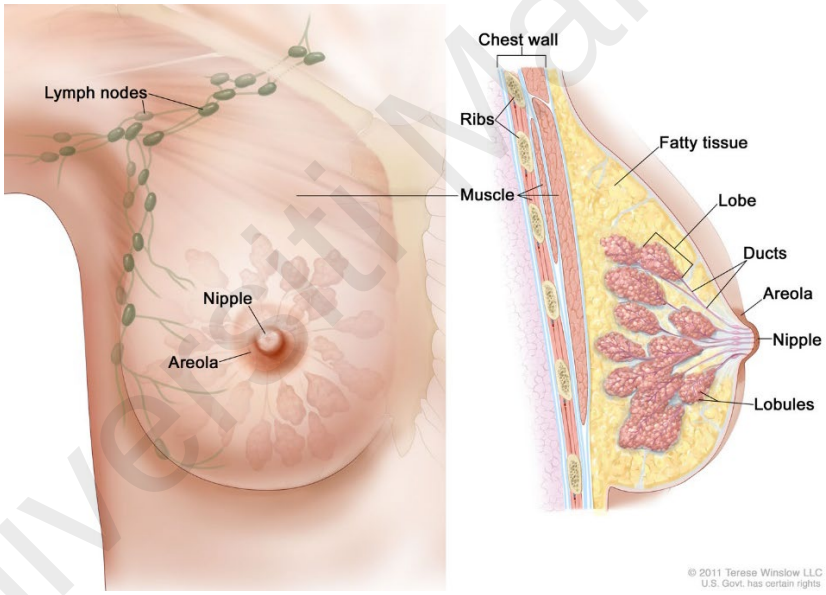


Figure 2.3: Female breast anatomy
(For the National Cancer Institute © (copyright 2011) Terese Winslow LLC, U.S. Govt. has certain rights)

In **Figure 2.3**, female breast is consisting of many tissue types including fatty tissue that regulates the size of the breast, lobules (known as glandular tissue) that produce milk, and fibrous tissue responsible for holding fatty tissue and lobules in place. A healthy female breast comprises of 15 to 20 lobes embedded in each breast's fatty tissue, which radiate around the nipple (Aydiner, İğci, & Soran, 2015). Each of these lobes has small

regions of tissue called lobules which have a glandular structure that produce milk. The milk ducts are linked with lobes and lobules to transport milk to the nipples. The nipple is located at the center of the areola and estimated to have 9 ducts, as well as nerves. Montgomery's glands can be found in areola which release a lubricating fluid that protects the nipple and skin from chafing during breastfeeding. The lymph nodes are tiny organs located at the chest, armpit and all across the body that produced a fluid called lymphocytes. Lymphocytes are immune cells that aid the fight against infections and diseases. Filtered lymph fluid is transported from the breast to lymph nodes through lymph vessel which is part of lymphatic vessels. Lymphatic vessels and blood vessels can be found throughout the breast.

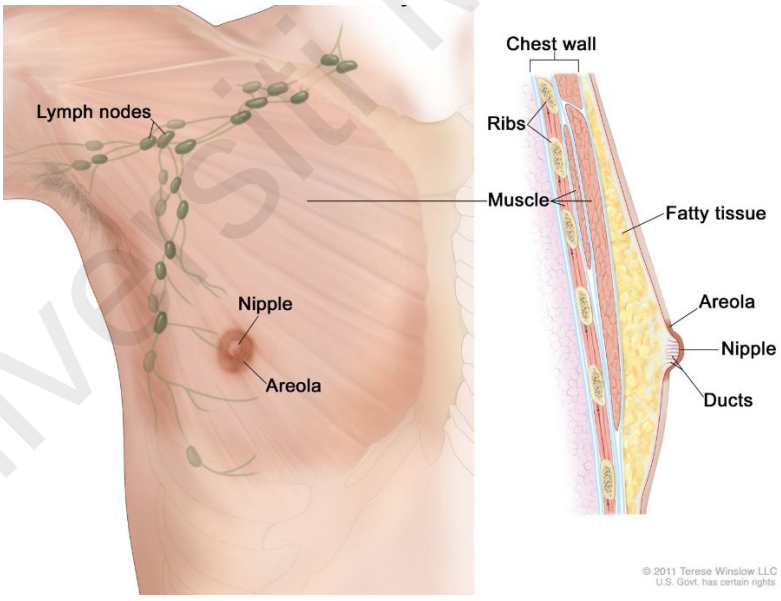


Figure 2.4: Male breast anatomy
(*For the National Cancer Institute © (copyright 2011) Terese Winslow LLC, U.S. Govt. has certain rights)*

During puberty, both male and female breasts will start developing. However, as for male breast, male hormone testosterone will prevent the breast from growing as female breast. Male breast has immature ducts and lacks a specialized structure of lobe and

lobules to produce and secrete milk as shown in **Figure 2.4**. On the exterior, male breast has nipple and areola which is similar to female breast. Even though male breast did not grow specialized lobules, their breast tissue and cell can still develop cancer although it's rare. Study by Anderson, Jatoi, Tse, and Rosenberg (2010), male breast cancer only account less 1% of total female and male breast cancer cases. Mostly male breast cancer developed later in life compared to female breast cancer. Male breast may also develop a condition called gynecomastia, medical term for a benign condition that cause abnormal enlargement of the male breast to grow.

2.2.2 Breast Cancer Symptom, Types and Stage

Breast cancer develops when breast cells proliferate at accelerated and unpredictable rate due to mutation or changes in genes that control the cell's growth which response to hormone estrogen. The majority cases of cancer developing at the cells from the lobules and milk ducts, the structures that containing glands. Cancer can spread to different regions of the body parts via lymphatic or blood vessel.

Breast cancer can also develop in the axillary tail of Spencer, a tail of breast tissue that extends into the underarm area, even though it is not detected within the actual breast. Breast cancers metastasize or spread to the lymph nodes through lymphatic vessel. Majority number of lymphatic vessel flow through axillary (underarm) lymph nodes, and only small amount flows to mammary lymph nodes that situated deep inside the breast. It is crucial to understand the lymphatic vessel because cancer usually begins with the first lymph node in the lymph node network when it's spreads. The lymph node that causes cancer to metastasize known as sentinel lymph node, need to be removed by surgery.

2.2.2.1 Breast Cancer Symptom

Various patients encounter various breast cancer symptoms, and some may not have any symptoms. The most prevalent indication of breast cancer is a painless breast lump. Breast cancer symptoms also include abnormal nipple discharge, change of appearance of a breast, abnormal nipple discharge, change of appearance of nipple and skin such as dimpling. Breast lumps can form for a variety of reason which majority of them are benign (noncancerous). So, it is critical to visit a health practitioner as soon as the symptom has been discovered for medical evaluation. Medical evaluation includes imaging of the breast and tissue sampling (biopsy) to determine benign (non-cancerous) and malignant (cancerous) is essential for more effective treatment.

2.2.2.2 Breast Cancer Types

Carcinoma is the prevalent type of breast cancer that originate in the skin or tissue that surround the internal organs and tissues throughout the body. There are four main types of breast cancer which is invasive breast cancer, non-invasive or in situ cancer, inflammatory breast cancer (IBC) and metastatic breast cancer.

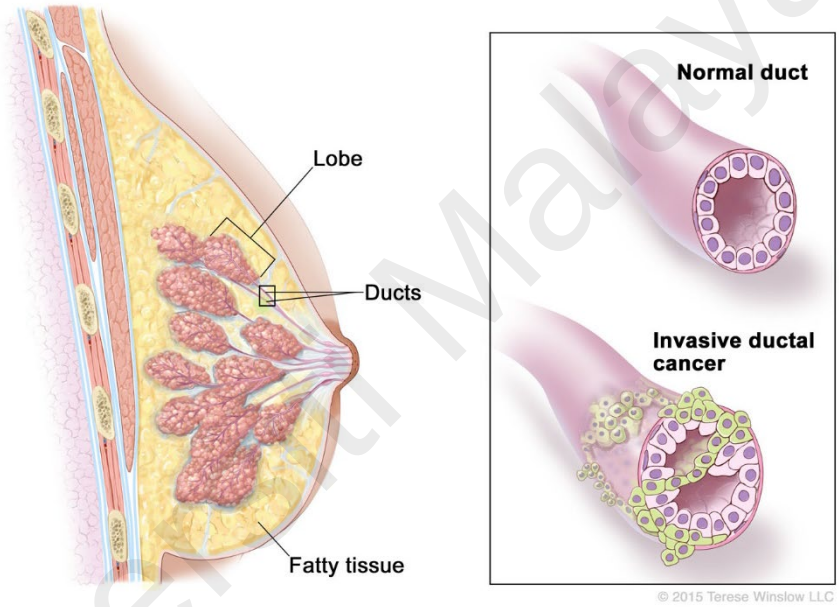


Figure 2.5: Invasive ductal carcinoma breast cancer
(For the National Cancer Institute © (copyright 2015) Terese Winslow LLC, U.S. Govt. has certain rights)

The most prevalent type of invasive breast cancer are invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). The invasive breast cancer spreads from the initial site to other regions such as neighboring breast tissue, lymph node, and other organs through blood vessels and lymphatic vessels. IDC is the most prevalent type of breast cancer which account around 70% to 80% occurrences. IDC begins from the milk ducts and spreads to the other breast regions, as shown in **Figure 2.5**. Over time, it may metastasize or spread further to other areas if not detected.

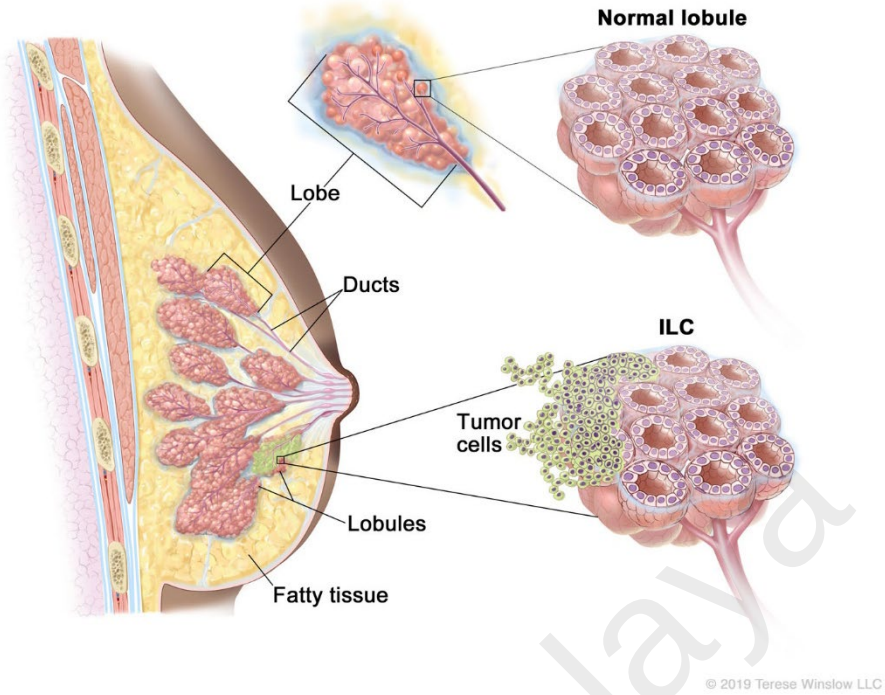


Figure 2.6: Invasive lobular carcinoma breast cancer
(*For the National Cancer Institute © (copyright 2019) Terese Winslow LLC, U.S. Govt. has certain rights)*

The second most prevalent type of breast cancer is ILC which accounts around 5% to 10% occurrences of breast cancers. ILC originates from the lobules and spreads to the surrounding breast tissue as shown in **Figure 2.6**. Similar to IDC, ILC also have the potential to metastasize further other regions of the body, but harder to detect on mammograms compared to IDC.

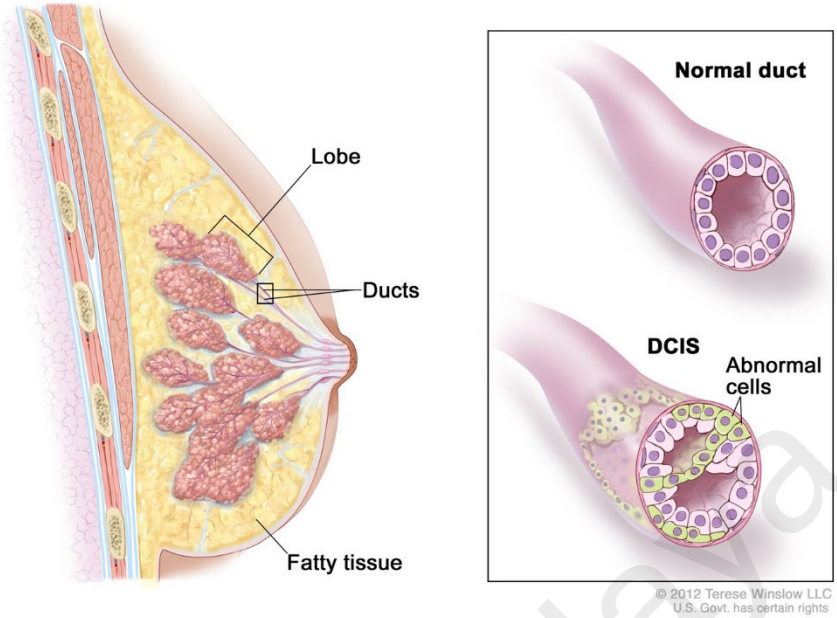


Figure 2.7: Ductal carcinoma in situ breast cancer
(*For the National Cancer Institute © (copyright 2012) Terese Winslow LLC, U.S. Govt. has certain rights)*

Ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) are prevalent for non-invasive or in situ cancer types. DCIS and LCIS is an early and noninvasive breast cancers (also known as stage 0 tumours) in which abnormal cells are developing in the milk ducts or the lobules as show in **Figure 2.7** and **Figure 2.8**. The abnormal cells will not spread outside of other regions surrounding the breast tissue. DCIS and LCIS are relatively curable cancers and seldom become invasive cancer. However, patients with DCIS are more likely to have cancer return or develop new cancer. The majority of recurrences are within 5 to 10 years originating from diagnosis and chances of recurrence are less than 20% (Ottesen, Graversen, Blichert-Toft, Christensen, & Andersen, 2000).

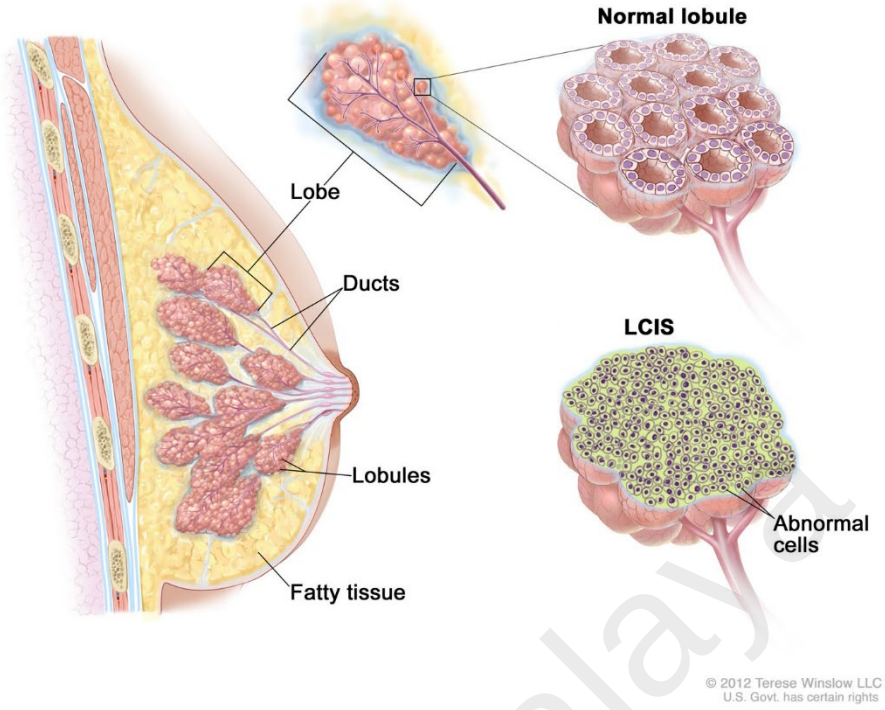


Figure 2.8: Lobular carcinoma in situ breast cancer
(For the National Cancer Institute © (copyright 2012) Terese Winslow LLC, U.S. Govt. has certain rights)

In **Figure 2.9**, inflammatory breast cancer (IBC) is a type of fast-growing and aggressive cancer where cancer cells invade lymphatic vessels and breast's skin. The disease is rare where it's only account for 1% to 2% of all breast cancer cases according to the American Cancer Society (Cristofanilli et al., 2007). Instead of a visible lump or tumour, inflammatory breast cancer (IBC) usually develops with swelling and reddening of the breast due to obstruction of lymphatic vessels by cancer cells. IBC tends to spread and grow rapidly in the matter of days or even hours.

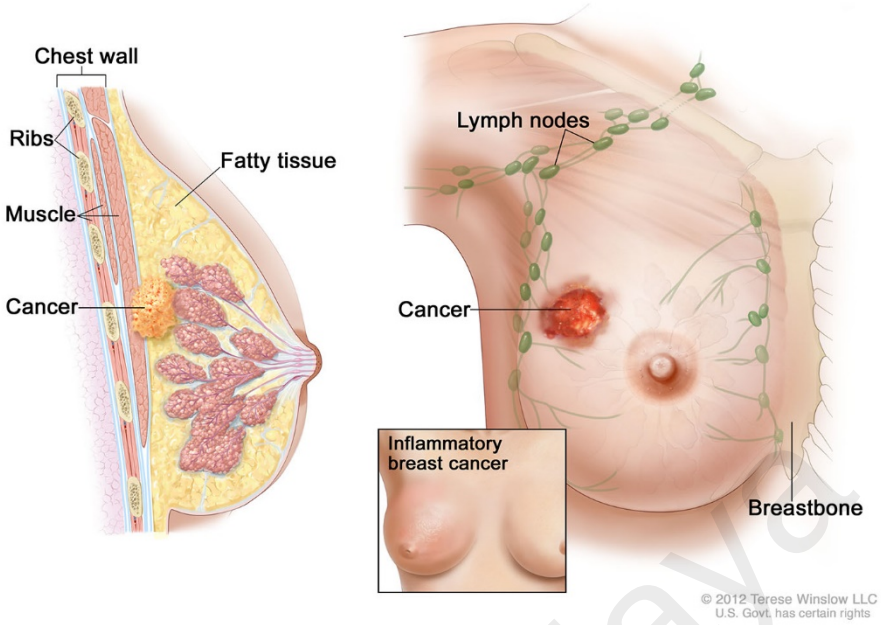


Figure 2.9: Inflammatory breast cancer
(For the National Cancer Institute © (copyright 2012) Terese Winslow LLC, U.S. Govt. has certain rights)

As for metastatic breast cancer, cancer has spread to other regions of the body such as the liver, lungs, brains or bones, referred to Stage 4 breast cancer. Cancer cells metastasize to other region through lymphatic vessels and blood vessels. Cancer can reoccur at other regions of the body in months or years after it has been diagnosed and treated in the first place. It indicates the presence of metastatic tumour made from breast cancer cells in other part of the body.

Other than 4 type of breast cancer mention above, there are also other types cancer can occur but not so common such as medullary carcinoma, phyllodes tumours of the breast, triple-negative breast cancer, HER2-positive breast cancer, hormone receptor-positive breast cancer, recurrent breast cancer, and Paget's disease of the nipple.

2.2.2.3 Breast Cancer Stage

Breast cancer stage represents the level of cancer, including the size of tumours, lymph node migration and spread distant to other regions of body. Staging conducted prior to

the surgery. Diagnostic tests used to establish the stage of cancer after all necessary tests are completed. Mammography, blood tests, bone scan, positron emission tomography (PET) scan, breast magnetic resonance imaging (MRI), and computerized tomography (CT) scan, are all used to stage breast cancer. Accessing the staging of breast cancer aids the doctors to determine the best course of treatment or possibility of recovery. **Table 2.1** shows the breast cancer stage and description for different type of cancer.

Table 2.1: Breast Cancer Stage

Staging		**Description**
Stage 0	Stage 0	Cancer only limited in the terminal ducts of breast tissue and not spread to other regions of the body. This stage also known as non-invasive or in situ cancer.
Stage I	Stage IA	Size of cancer is tiny, invasive and not metastasize to the lymph nodes.
	Stage IB	No sign of tumour or size of tumour less than 20 mm at the breast. Cancer metastasize to the lymph nodes and the cancer size in the lymph node is around 0.2 mm and 2 mm.
Stage II	Stage IIA	Any of the following conditions: • No sign of tumour at the breast. Cancer metastasizes from 1 to 3 axillary lymph nodes but not metastasize to other regions of the body. • The tumour size is 20 mm or less at the breast. Cancer metastasizes from 1 to 3 axillary lymph nodes. • The tumour size is more than 20 mm but less than 50 mm at the breast. Cancer not metastasize to axillary lymph nodes.
	Stage IIB	Any of the following conditions: • The tumour size is more than 20 mm but less than 50 mm at the breast. Cancer metastasizes from 1 to 3 axillary lymph nodes. • The tumour size is more than 50 mm at the breast. Cancer not metastasize to axillary lymph nodes.

Table 2.1 continued

Staging		Description
Stage III	Stage IIIA	Any of the following conditions: • Any tumour size at the breast. Cancer metastasizes from 4 to 9 axillary or internal mammary lymph nodes. • The tumour size is more than 50 mm at the breast. Cancer metastasizes from 1 to 3 axillary lymph nodes.
	Stage IIIB	Tumour has metastasized to the chest wall or swelling or ulceration of the breast, or diagnosed as inflammatory breast cancer. Cancer metastasizes up to 9 axillary or internal mammary lymph nodes.
	Stage IIIC	Any tumour size at the breast. Cancer metastasizes up to 10 or more axillary or internal mammary lymph nodes and/ or the lymph nodes under the collarbone.
Stage IV	Stage IV	Any tumour size at the breast and has metastasize to other region of the body including the liver, lungs, bones and brain, distant lymph node and chest wall. Cancer also known as metastatic cancer, meaning it has metastasize beyond the region of the body from where it was originated.

2.2.3 Breast Cancer Tests and Treatments

2.2.3.1 Breast Cancer Tests

The most common breast cancer test fall into one of three methods: screening tests, diagnostic tests and monitoring tests. Before the test, doctor will conduct a breast physical exam and extensively inquire the patient's personal and family history of breast cancer, or genetic mutation or history of chest radiation therapy.

Screening tests are performed routinely such as yearly mammograms to persons who are considered as healthy and has not been diagnosed with breast cancer. The screening test's objective is to detect breast cancer early before showing any symptoms (such as lump) so doctors can perform effective treatment. Mammography still remains the primary screening in the community compared to clinical breast examination or self-examination, and unlikely to be replaced by other imaging modalities for screening general population in near future (Elmore, Armstrong, Lehman, & Fletcher, 2005). However, there is no scientific proof that the self-examination method reduced breast cancer detection or mortality, and most expert groups no longer recommended using it as a screening test (Hackshaw & Paul, 2003).

Diagnostic tests are performing to persons who are considered as having breast cancer based on symptoms or the results of a screening test. The tests are used to identify cancer in the breast and to check the spreading of the tumour. The result of diagnostic tests is used to gather information about cancer to guide treatment options. Diagnostic tests and methods include breast examination, ultrasound, mammogram, breast MRI and biopsy. In terms of breast examination, doctor will be examining for tumours or other abnormalities in both of the breast and lymph nodes at the armpit. Mammogram are x-ray that are used to evaluate the breast, and common to diagnose for breast cancer. Ultrasound creates images of the area and structures deep within the body and breast using sound

waves or also known as sonograms. Ultrasound may be used to determine the lump as fluid-filled cyst or solid mass. Biopsy is a procedure of taking a breast cell sample to test for breast cancer. During biopsy, a specialised needle device guided by ultrasound, x-ray or MRI is used to retrieve core tissue or fluid of suspicious area include lymph nodes in axilla (underarm). The biopsy specimen then submitted to laboratory for examination to determine if the cells are cancerous and evaluate the type of cells implicated in breast cancer, aggressiveness and to check cancer cell's receptors that may impact the treatment available. Breast magnetic resonance imaging (MRI) produce the breast image using magnet and radio waves without the use of radiation, and the patients was given dye injection before the scan.

Monitoring tests are performed during or after treatment to evaluate the efficiency of the therapy once breast cancer has been detected. Monitoring tests may also be performed to check any sign of breast cancer recurrence. To determine if cancer has spread, chest x-ray, CT scan and liver ultrasound may be required. The MRI of the breast may be required to clarify the results or determine the depth of the condition in the breast. The bone scan is necessary if cancer has spread to the bones. The isotope, a material containing a small quantity of radiation injected into a vein at the arm before the scan. The isotope absorbed by the bone that has been affected by cancer and appears as highlighted areas on the scan.

2.2.3.2 Breast Cancer Treatments

The result of the test responds to the specific type of treatment. The main objective of treatment is to cure cancer and extend patient's life. It is also important to improve patient's quality of life include support patient's physical, mental and spiritual health, and palliative care (treatment to relieve, rather than cure).

Breast cancer treatments consists of chemotherapy, radiotherapy, surgery, hormone therapy and targeted therapy. The treatment's type determined how cancer was diagnosed

and breast cancer stage. Breast cancer detection through screening test is considered early stage, however breast cancer discovered after symptoms may be in later stage and necessitate different therapy.

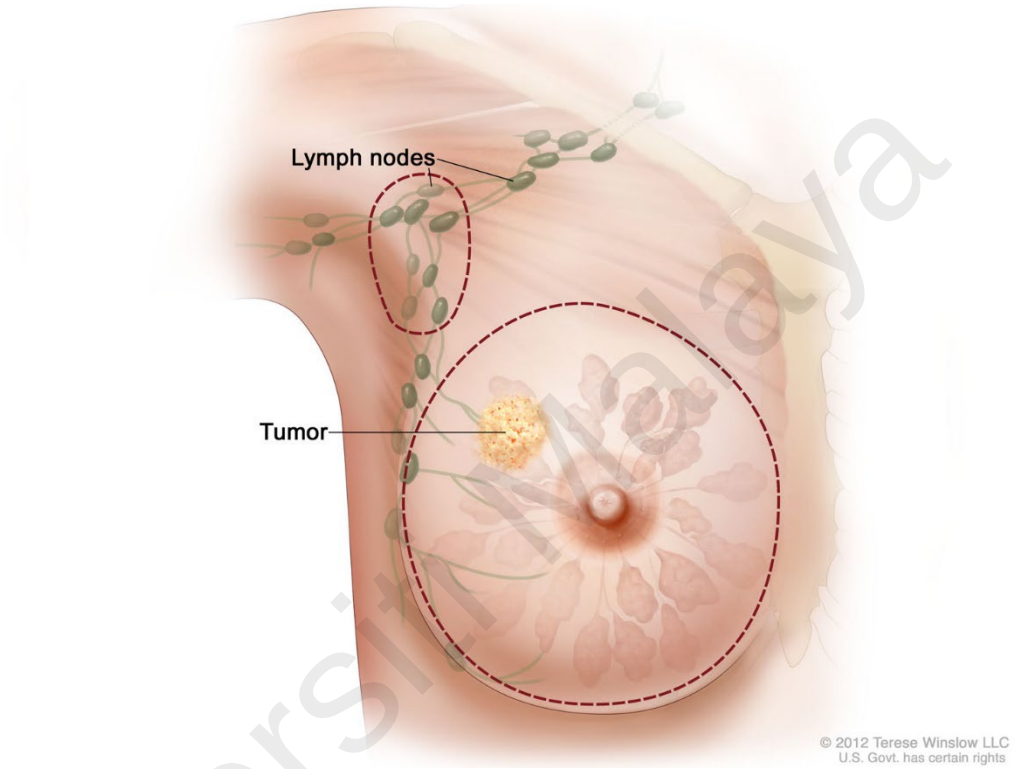


Figure 2.10: Mastectomy surgery
(*For the National Cancer Institute © (copyright 2012) Terese Winslow LLC, U.S. Govt. has certain rights)*

Surgery is the common type of therapy for breast cancer and the type of surgery determined by the stage of the breast cancer. Radiotherapy, chemotherapy, hormone therapy or targeted therapy usually administered after surgery. In some case, chemotherapy or hormone therapy used as the initial treatment before surgery. Breast cancer surgery can be classified into two type which is mastectomy or breast-conserving surgery (also known as lumpectomy). Mastectomy is a surgery required to remove the whole breast including the nipple and often followed by reconstructive surgery to rebuild

a breast as shows in **Figure 2.10**. Meanwhile, breast-conserving surgery is a surgery required to removed cancerous lump (tumour) and lymph nodes only as shows in **Figure 2.11**. The breast-conserving surgery includes lumpectomy (wide local excision), a surgery to remove the tumour and small quantity of surrounding breast tissue, and partial mastectomy (quadrantectomy), a surgery to remove up to quarter of the breast. Breast-conserving surgery accompanied by radiotherapy has demonstrated in studies to be effective as total mastectomy in the treatment of early-stage breast cancer.

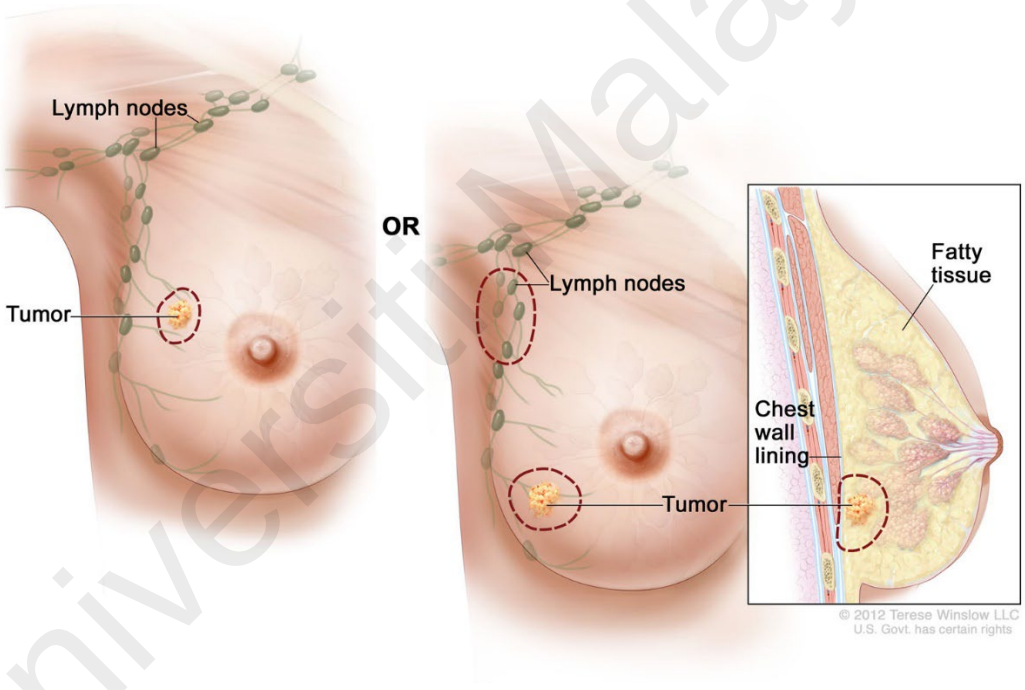


Figure 2.11: Breast-conserving surgery
(For the National Cancer Institute © (copyright 2012) Terese Winslow LLC, U.S. Govt. has certain rights)

In **Figure 2.12**, a procedure called sentinel lymph node biopsy may be used to determine the spread of breast cancer. It is used for patients that already been diagnosed with cancer. The sentinel lymph nodes are the first lymph nodes reach in contact that is located at the underarm (axillary lymph nodes) is identified and removed. The removed sentinel nodes are analyzed in the lab to detect the presence of cancer cells. The negative

result indicates that cancer has not metastasized to adjacent lymph nodes or organs. Meanwhile, a positive result indicates that cancer has metastasised to other surrounding lymph nodes and organs. This information is helpful to evaluate the cancer stage and developing an appropriate treatment program.

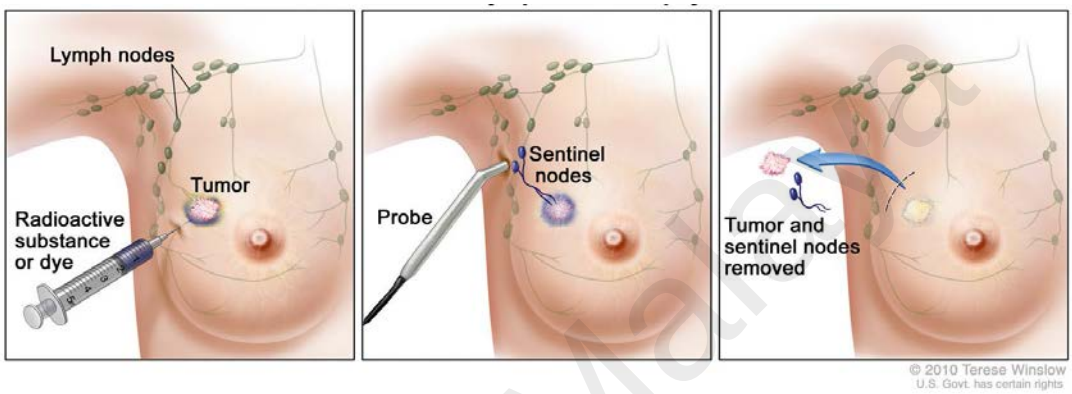


Figure 2.12: Sentinel Lymph Node Biopsy
(*For the National Cancer Institute © (copyright 2010) Terese Winslow LLC, U.S. Govt. has certain rights)*

Radiotherapy employs regulated amounts of radiation to eliminate cancer cells. Radiotherapy often administered following surgery and chemotherapy, to eliminate remaining cancer cells. Radiotherapy's type includes radiotherapy to the lymph nodes (applied radiation at the armpit and surrounding area), breast boost (applied high-dose radiation to the area where cancer was removed), chest-wall radiotherapy (applied radiation to the chest wall after mastectomy surgery), and breast radiotherapy (applied radiation to the remaining breast tissue after breast-conserving surgery).

Chemotherapy use anti-cancer (cytotoxic) medicine to eliminate cancer cells. Chemotherapy can be administered before the surgery to decrease the tumour size (known as neo-adjuvant chemotherapy) or after surgery to eliminate cancer cells that was not removed (known as adjuvant chemotherapy). Medications are often administered by a

drip directly into the vein. The treatment session is known as cycle of chemotherapy, which depends on therapy program.

Hormone therapy is a medicine used to eliminate or reduce the levels of estrogen or progesterone hormones, which can lead to the formation of breast cancers. During hormone receptor test, a sample of the cell that obtained will be tested and analyze if they react to estrogen or progesterone. Hormone-receptor positive cancer occurs when a hormone attaches to cancer cells. Hormone therapy administered after surgery or chemotherapy, and sometimes used before the surgery to decrease the tumour size and make it easier to remove. Type of hormone therapy include tamoxifen, aromatase inhibitors and ovarian ablation.

Targeted therapy is medicine that change the way cells work and aid the prevention of cancer growth and spread. However, targeted therapy cannot be used to treat all cases of cancer. Other cases of breast cancer are stimulated by a protein cell that are known as human epidermal growth factor receptor 2 (HER2). Tratuzimab is the most often used targeted therapy for breast cancer treatment. Targeted therapy treatment usually administered by a drip into a vein and also available as tablets.

2.3 Breast Cancer Imaging Diagnostic Test

2.3.1 Mammography

Mammograms are likely the most essential tool doctors have to do screening, diagnosing, evaluating and monitoring patients with breast cancer. Mammograms are low-dose x-ray of the breast. Mammography typically involves taking two x-rays of each breast from opposite angles: top to bottom (craniocaudal) and side to side (mediolateral oblique). To ensure that the entire breast is seen, the breast is pulled away from the body, squeezed and held between two glass plates. Regular mammograms can aid to diagnose breast cancer in its early, where early treatment can be applied effectively. Years before physical symptoms appears, a mammography can detect breast abnormalities that could identify as cancer. Persons who get frequent mammograms are more prone to have breast cancer detected early, and less likely require aggressive treatment such as surgery to remove the entire breast (mastectomy) and chemotherapy. However, there are some images that mammogram failed to detect. In certain cases, patients require further testing test to determine cancer. There are also the possibility of being diagnosed with cancer that never would have caused any problems had it not been discovered during screening, which also known as overdiagnosis. Study by Gøtzsche and Jørgensen (2013) shows screening with mammography leads to significant overdiagnosis and the effect of reduce breast cancer mortality remains uncertain.

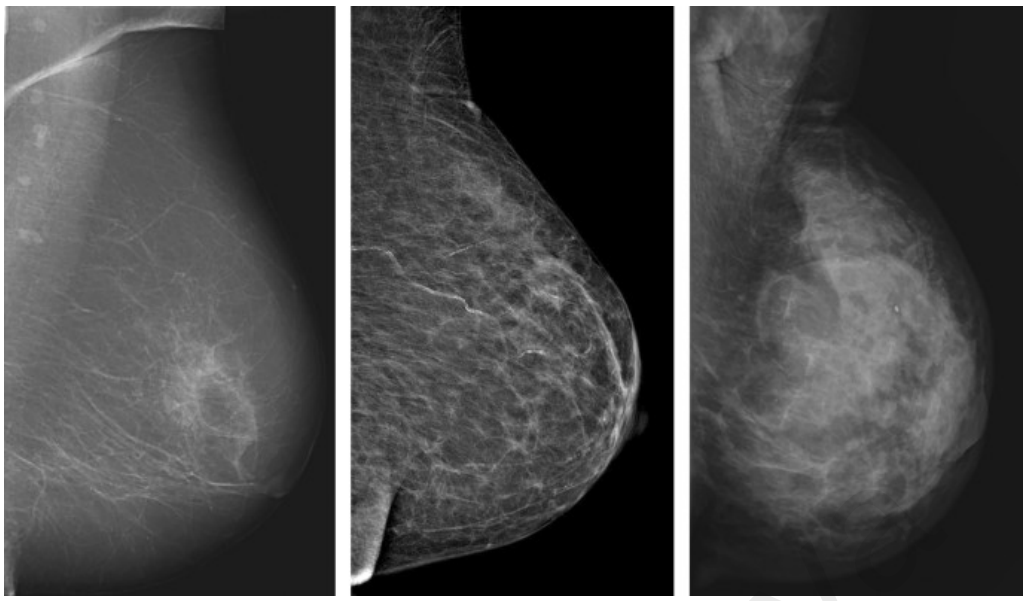


Figure 2.13: The difference of breast density from left to right: fat involuted, moderate and dense breast images acquired by mammogram (Sturesdotter et al., 2020)

Although mammographic screening reduced breast cancer mortality, the mammography's sensitivity decreased and unable to identify cancer in dense breast tissue. Women with dense breast tissue describes as the image of breast tissue appears on a mammogram. As mention before, breast tissue consists of glandular tissue (tissue that produce milk such as lobes), fibrous tissue and fatty tissue. Dense breast tissue has a lot of glandular tissue and fibrous tissue compared to fatty tissue. On mammogram, fatty tissue (non-dense breast tissue) appears dark and transparent, meanwhile glandular tissue and fibrous tissue (dense breast tissue) appears solid white area. **Figure 2.13** shows the difference of breast density from fat involuted images into dense breast images. Dense breast tissue makes it difficult to detect breast cancer on mammograms and increase the risk of cancer undetected. Dense breast tissue is frequently associated with an increased risk of breast cancer symptoms, including larger tumours, node-positive disease and likelihood of developing breast cancer.

There are two type of mammography which is screen-film mammography and full-field digital mammography. In term of cancer detection, there is no substantial difference between screen-film mammography and full-field digital mammography. However, full-field digital mammography has less recalls than screen-film mammography (Hoff et al., 2012; Lewin et al., 2002). In near future, full-field digital mammography gradually replacing screen-film mammography as breast cancer screening technique due to the implementation of advanced application includes computer-aided detection or diagnosis (CAD) and tomosynthesis.

2.3.2 Digital Breast Tomosynthesis

Digital breast tomosynthesis, also known as three-dimensional (3D) mammography, is a modern version of mammogram. Digital breast tomosynthesis captures several x-ray images of each breast from various angles. The position of the breast is similarly of mammography, but only enough of pressure to maintain the breast stable during the operation. During the examination, the x-ray tube travels in the arc all around the breast to take 11 images in 7 seconds. The images transferred to the computer for processing to create 3-dimensional images of the breast. As a result, breast imaging is improved by reduced tissue overlap and effect of superimposition may allow better visualization of detecting breast cancer. Thus, researcher anticipate that this approach of breast imaging increase the sensitivity and specificity to diagnose cancer in dense breast tissue (Hadadi, Rae, Clarke, McEntee, & Ekpo, 2021) and by combining screening using 2D and 3D mammograms may significantly enhance breast cancer detecting (Houssami & Skaane, 2013). In **Figure 2.14**, digital mammogram shows small spiculated mass on left CC view but does not show on left MLO due to overlapping breast tissue. Meanwhile tomosynthesis images show clearly small spiculated mass on both left CC and MLO view. Studies have found that digital breast tomosynthesis have lower risk of being called back for follow-up examination to detect additional breast cancer compared with conventional

mammography (McDonald et al., 2016; Skaane, 2016). Digital breast tomosynthesis has the possibility to overcome the limitation of mammography by improve diagnostic performance in identifying benign and malignant features, increase cancer detection in dense breast and masking effect caused by overlapping breast tissue (Gilbert & Pinker-Domenig, 2019).

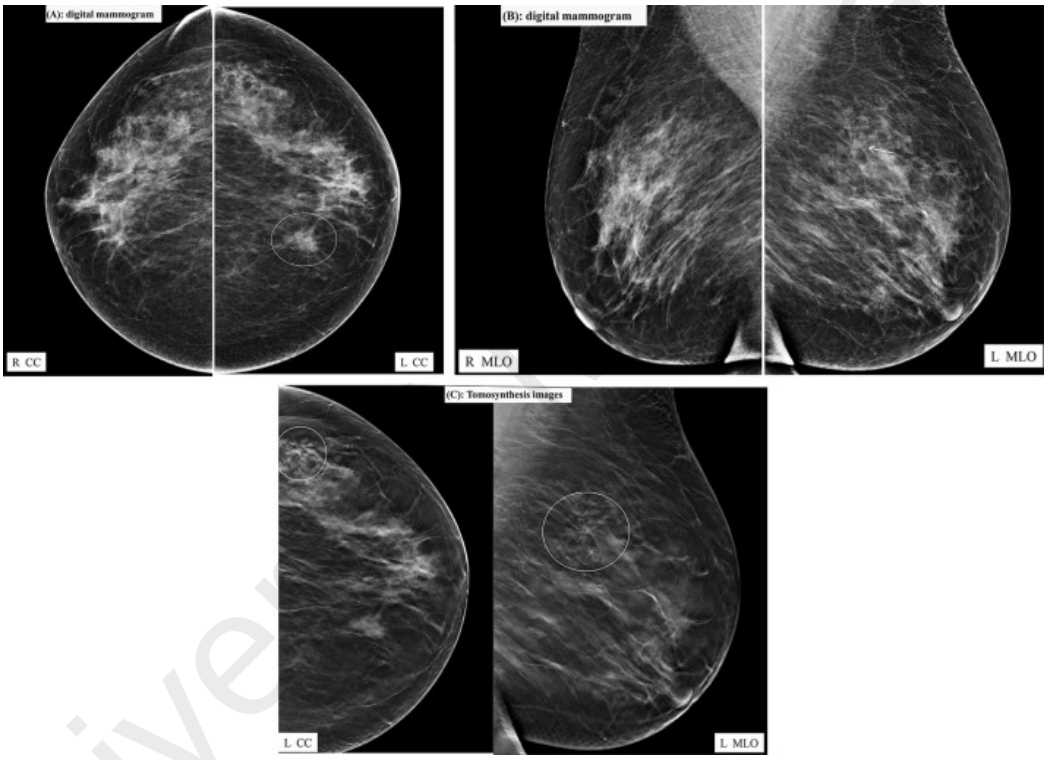


Figure 2.14: Digital mammogram (top) able to detect small spiculated mass (circle) left CC view only, compared with digital breast tomosynthesis (bottom) able to detect both left CC and MLO view (Houssami & Skaane, 2013)

2.3.3 Ultrasound

Ultrasound has dramatically improved in the past years with higher resolution and rapid processing. While mammography is a standard imaging method for screening breast cancer, it tends to miss cancers of women dense breast tissue and ultrasound have the ability to capture the images by using cross-sectional technique, displaying tissue without overlap (Kelly, Dean, Comulada, & Lee, 2010). Ultrasound is a type of imaging diagnostic test that transmits high-frequency sound waves instead of radiation through the breast and converts it into images. To perform the test, sound-emitting probe was put on the breast. Ultrasound is not performed as a stand-alone breast cancer diagnostic test, instead it is used to complement other diagnostic tests. If a physical exam or mammogram indicates an abnormality, ultrasound is the best approach to use as diagnostic examination on specific area of concern and determine the abnormality are solid masses or fluid-filled. **Figure 2.15** shows breast images for normal, benign and malignant. However, ultrasound unable to determine whether solid lump is cancerous, nor can it detect calcifications especially microcalcifications and small tumours might be ignored if cancer is invasive lobular. However, ultrasound may detect ductal carcinoma in situ (DCIS) due to enhance resolution in recent years.

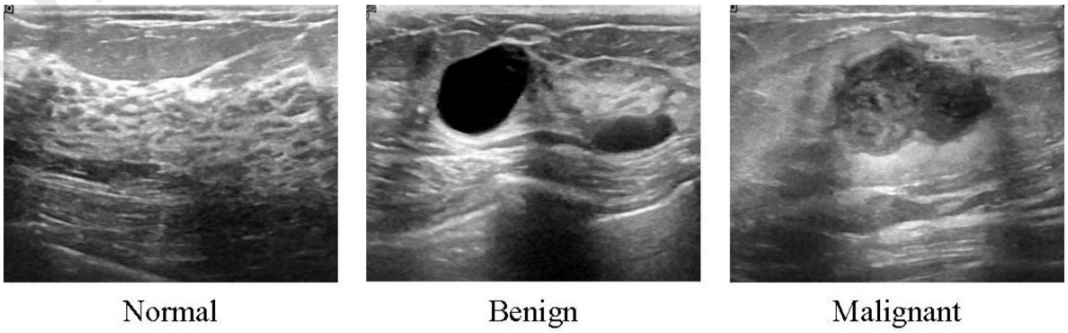


Figure 2.15: Ultrasound breast images from left to right: normal, benign and malignant
(Al-Dhabyani, Gomaa, Khaled, & Fahmy, 2020)

The important aspect of using ultrasound is the ability to perform image-guided biopsy rapidly, accurate and safe with no precautions other than checking for a bleeding diathesis. Ultrasound can accurately guide biopsy needles to the suspicious areas in the breast tissue. A biopsy is a minor surgical procedure used to remove tissue from the suspicious area in the breast that discovered by using imaging scan. A tiny needle with a hollow core is used by the surgeon to retrieve the tissue sample from the suspicious area. Most of the time, the needle guide to the location of the lump. In circumstances when a lump cannot be felt, the surgeon may need to use imaging studies such as ultrasound (also known as ultrasound-guided biopsy) and mammography (also known as stereotactic needle biopsy) to direct the biopsy to the suspicious area. As for ultrasound-guided biopsy, the surgeon use ultrasound monitors to guide the needle to the suspicious area. Meanwhile, with stereotactic mammography, mammograms are obtained from various angles to locate the suspicious area before surgeon inserts the needle to remove the tissue sample. Ultrasound can also be utilised as a second-look technique in patients with MRI anomalies.

2.3.4 Breast Magnetic Resonance Imaging

Breast Magnetic Resonance Imaging (MRI) offers a variety of applications for breast cancer including screening for high-risk patients, performing in high sensitivity in dense breast tissue, acquiring detail information about the suspicious area obtained from mammogram or ultrasound, monitoring for recurrence after treatment. Magnetic resonance imaging (MRI) is a technology that uses magnets and radio waves to produce comprehensive cross-sectional images inside in the body, without exposure of radiation. With the use of contrast solution (dye), MRI may provide comprehensive 3D images of breast tissue. The contrast solution enhances the images where the contrast becomes more concentrated in the location of cancer cell proliferation in breast tissue, resulting as white areas on black background. This assists the radiologist determine cancer region and

tumours may be detected without being affected by dense breast tissue. However, contrast solution may harm the kidney and some patients may experience brief discomfort.

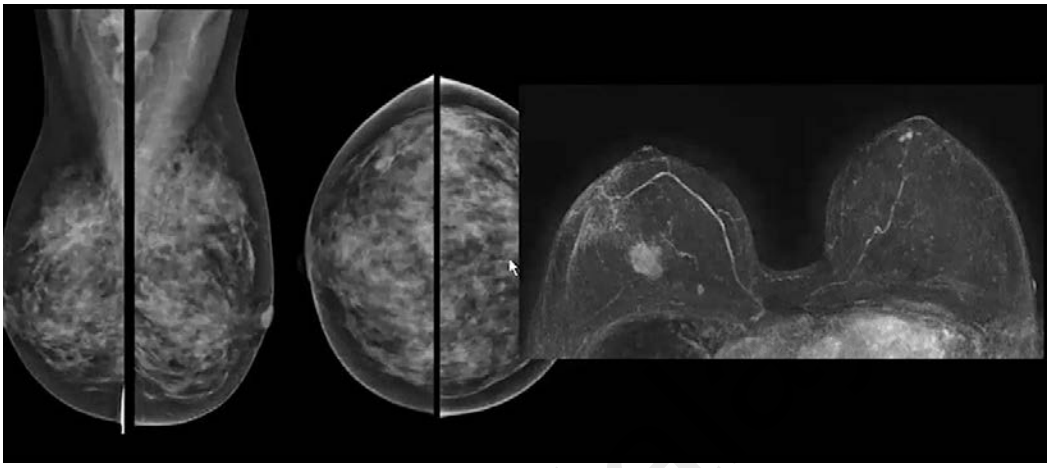


Figure 2.16: A comparison of mammography imaging (left) and breast MRI imaging (right) to detect cancer for women with dense breast tissue (Fornell, 2021)

Due to the technology of MRI using strong magnets, the patients need to remove any metal item and MRI is not suitable for patients with metal implants in the body such as pacemaker or artificial joint. Furthermore, patients may experience discomfort due to loud thumping sound and being confined in small spaces for extended of time especially for claustrophobic patients. During screening, breast MRI are more sensitive in detecting cancer compared to mammograms and ultrasound, however, the main disadvantage is screening results is more to false positives. Therefore, breast MRI is not suggested as a screening tool for average risk and only reserve for high risk especially for young women with familial breast cancer and carriers of BRCA1 or BRCA2 mutations (Elmore et al., 2005; Kuhl et al., 2005; C. H. Lee et al., 2010). Moreover, screening MRI more benefit as secondary screening for mammographic image of dense breast tissue (Hylton, 2005) as shows in **Figure 2.16**. As for diagnosis, breast MRI are better to detect normal breast and breast cancer gland tissue compared to other imaging modality. However, the

operation of breast MRI required highly trained experts, time-consuming, variability in performance, may cause problematic for claustrophobic patients and the cost is expensive due to highly specialized equipment probably prohibits to use routinely for screening general populations. Furthermore, MRI not capable to detect calcification (calcium deposits in breast tissue that could be cancer) and have relatively low specificity compared to mammography. In some situation, breast MRI is capable to gather more information about the suspicious area evaluate patient who has palpable mass, evaluate glandular breast and etc. After treatment, breast MRI can be beneficial to evaluating scar tissue for patients who have a lumpectomy surgery and detect changes for any recurrence of breast cancer. Other than that, MRI can scan the other part of the body that may infected by cancer especially for patient that have metastatic breast cancer.

2.3.5 Computerized Tomography

Computerized tomography (CAT or CT) scan is an x-ray method that provide information in 2D slices or cross-section images of body internal organs. The information is collected by collects various angles of body x-ray and computer will generate the detailed images of the internal organ of the body. A dye as contrast solution will injected into the patient's arm through intravenous line before the test to feed better images of cancer.

CT scans are not used routinely to evaluate the breast especially during an early-stage breast cancer. Instead of evaluate early-stage breast cancer, CT scans is useful to investigate for breast cancer staging and follow up for new and recurrent metastatic breast cancer (James et al., 2019). CT scan used to evaluate cancer spread to the chest wall which determine if cancer can be removed with mastectomy surgery. Other than chest wall, CT scan also help to investigate other regions of the body such as spine, brain, lungs, lymph nodes, head, chest and abdomen. CT scans used to check whether cancer is responding

to the therapy during treatment and the spread or recurrence of breast cancer after treatment.

2.3.6 Positron Emission Tomography

Positron emission tomography (PET) scans can identify cancer region by capturing images of cell in the body while they function. A substance that contains sugars and small amount of radioactive material is administered into the body and as a result, cancer cells absorb more radioactive sugar than normal cells. The body then scan by a special camera to detect any highlighted region on the screen which identify as suspiciously active and may indicate as cancer. PET scans are not used to screen for breast cancer and has limited capacity to identify small tumours. Usually, PET scans beneficial after breast cancer has been diagnosed for examine the cancer metastasize to lymph nodes and other regions of the body, to check if cancer responding to the treatment or recurrence of breast cancer after the treatment. PET scans are useful if another testing is inconclusive to determine cancer spreading to other regions of the body.

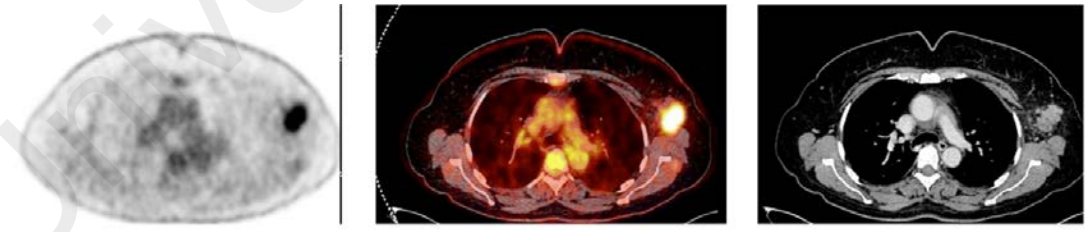


Figure 2.17: Left mass breast cancer located at the axillary tail of the breast from different imaging modalities from left to right: PET, PET/CT and contrast-enhanced CT
(Shawky, Ali, Hashem, & Houseni, 2020)

2.3.7 PET/CT

PET/CT scans is a hybrid imaging modality that combine positron emission tomography (PET) with computerized tomography (CT), which offers more accurate anatomic definition for pathologic and physiologic of cancer. PET/CT scans outperforms other imaging modalities in detecting possible regional lymph node recurrence such as axillary tail of the breast or risk of additional distant metastasis as shows in **Figure 2.17**. The detailed information enhanced the sensitivity and specificity compared to PET or CT scans alone, and played an important role on the breast cancer patient's management (Piperkova et al., 2007; Shawky et al., 2020).

2.3.8 Chest X-ray

The chest x-ray is routinely check for the spread of cancer to the lungs before beginning any treatment. The test may also perform to evaluate the heart and lungs before receive general anaesthesia, chemotherapy or radiotherapy. However, chest x-ray provides little information about the breast cancer and have a high false positive rate during testing (Hurria, Leung, Trainor, Norton, & Hudis, 2003). During treatment, the chest x-ray may be used to evaluate how disease responds to the treatment, presence of pneumonia during chemotherapy and lung inflammation during radiotherapy.

2.4 Mammography Breast Cancer Test

Mammography is the gold standard for screening and diagnosing breast cancer. When patient have mammogram, the breast will be positioned and compresses between two clear plates. The plates are connected to specialized camera which takes two images of the breast from two directions to identify features that indicate breast cancer include clasifications and masses. In **Figure 2.18,** the routine screening of the two directions is known as craniocaudal (CC) and mediolateral oblique (MLO) views of each breast. It is required to compress the breast to flatten and reduce the thickness as the x-ray beam

should be able to penetrate as few overlapping tissues as possible. The images will be taken of both breasts, even if lump detected on one breast, so that the breast can be compared and examined for abnormalities. The radiologist can also compare the old mammogram to the new mammogram to check the changes. There are breast changes type includes masses (large abnormal areas) and calcifications (small white spots) and other suspicious regions that may be cancerous.

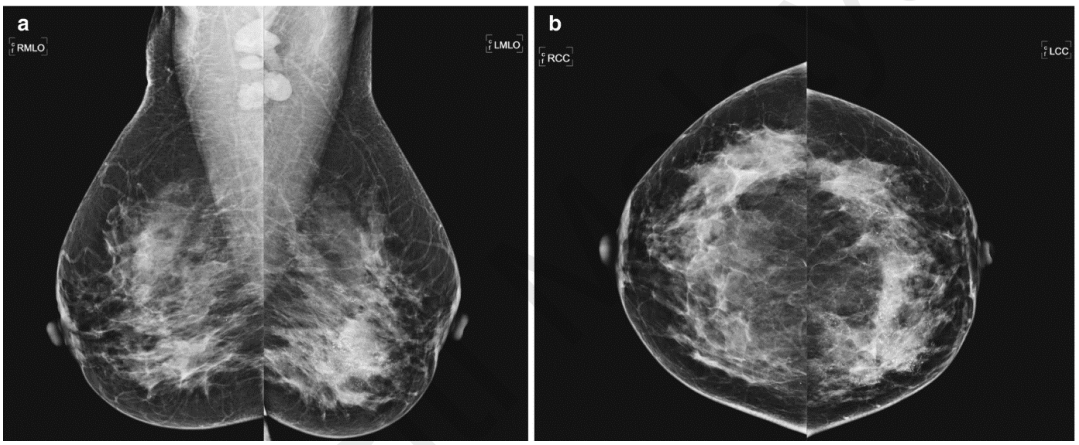


Figure 2.18: Mediolateral oblique (MLO) and craniocaudal (CC) view of both breast
(Gilbert & Pinker-Domenig, 2019)

2.4.1 Calcification

Calcification are small calcium deposits inside the breast tissue. On mammograms, they have an appearance as small white spots and may indicate the presence of an early breast cancer as shows in **Figure 2.19**. Calcification are normally can't be felt, although they can be seen on a mammogram. Calcifications are classified into two categories which is macrocalcifications and microcalcifications. Macrocalcifications are bigger calcium deposits created by abnormalities in the aging of breast arteries, past traumas and inflammation. Macrocalcifications deposits caused by non-cancerous disease and do not require to be tested by biopsy. It is common for older women after 50 years old.

Microcalcifications, on the other hand, are tiny calcium deposits in the breast which are more concerning than macrocalcifications. Microcalcifications not necessarily indicate the presence of cancer, but the form and pattern such as clusters of microcalcification may determine the likelihood that the changes occur in the area of early cancer. A biopsy may need to check for cancer if microcalcification have suspicious form and pattern. Studies shows that about 54.5% cancer-related calcifications might possibly detected earlier and critical to develop strategies that allow patient recall earlier while reducing false positives and invasive diagnostic examination (Mordang et al., 2018).

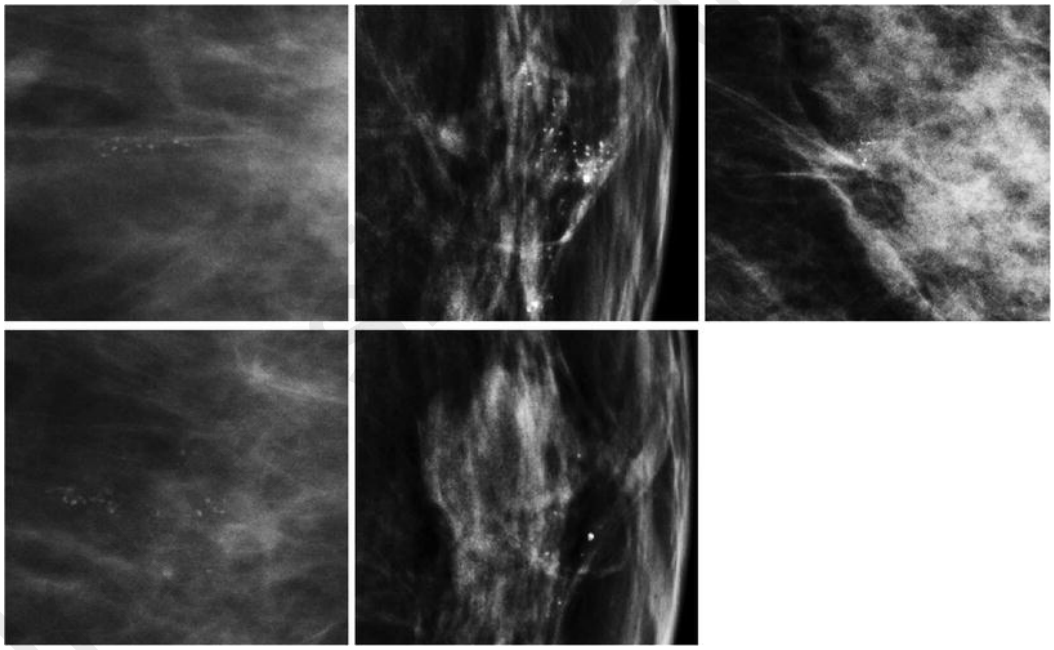


Figure 2.19: Calcifications detected in mammogram (Mordang et al., 2018)

2.4.2 Masses

A mass is a dense region that having distant shapes and edges, from the rest of the breast tissue. Masses can be cysts or non-cancerous solid tumours, but they can also be an indicator of cancer. Cyst are sacs filled with fluid which are very common and are mostly not related with cancer. Simple cysts are non-cancerous and do not require a

biopsy. Biopsy may require if it's not simple cyst to ensure it isn't cancer such as complex cysts or complicated cysts.

Although solid masses are more concerning, the majority of breast tumour are not cancer. Cells in the body normally regenerate themselves in an orderly process of cell proliferation, with healthy new ones taking over as old cells die off. However, mutation can change specific genes in the cells which responsible to regulate cell development and to maintain the healthy cell. As shown in **Figure 2.20**, the changed cells develop the ability to keep proliferate uncontrollably, resulting in the formation of a lump or tumour, which can be benign or malignant. Benign tumour is not cancerous because the cells seem be normal, develop slowly and do not metastasize to other region of the body. Although benign tumour is rarely life threatening, some types can raise the risk of develop breast cancer. Meanwhile, malignant tumour is cancerous and can metastasize to other region of the body.

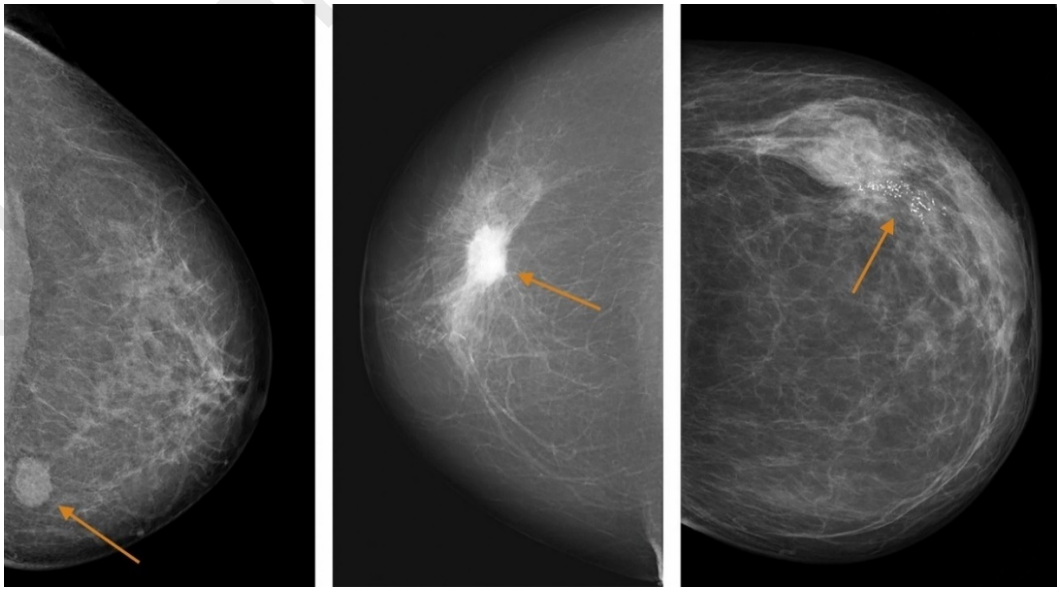


Figure 2.20: Tumours appearance in different density breast on mammogram from left to right: distinct mass on moderate dense breast, spiculated mass on fat involuted breast and calcifications on moderately dense breast (Sturesdotter et al., 2020)

2.4.3 Breast Imaging Reporting and Data System

The American College of Radiology developed the Breast Imaging Reporting and Data System (BI-RADS) to standardise breast imaging reporting and provide clarity in the interpretation of breast imaging findings. The BI-RADS atlas guidelines assist radiologist identifying and interpretation of suspected calcifications. The patients will be recall for additional clinical evaluation for suspected calcifications include biopsy. **Table 2.2** shows the BI-RADS assessment categories table and provide information about the management and the likelihood of cancer.

Table 2.2: BI-RADS assessment categories

Category		**Management**	**Likelihood of Cancer**
0	Need additional imaging or prior examinations	Recall for additional imaging and/ or awaits prior examinations	n/a
1	Negative	Routine screening	Essentially 0%
2	Benign	Routine screening	Essentially 0%
3	Probably Benign	Short interval-follow-up (6 month) or continued	> 0% suspicious ≤ 2%
4	Suspicious	Tissue diagnosis	4a. low suspicious for malignancy (> 2% to ≤ 10%) 4b. moderate suspicious for malignancy (> 10% to ≤ 50%) 4c. high suspicious for malignancy (>50% to <90%)
5	Highly suggestive of malignancy	Tissue diagnosis	≥ 95%
6	Known biopsy-proven	Surgical excision when clinically appropriate	n/a

2.5 Machine Learning/ Deep Learning in Medical Image Application

In biomedical research and clinical practice, artificial intelligence (AI) models are increasingly taking on a role with variety of application include risk modelling and classification, personalized screening and diagnostic (Castiglioni et al., 2021). The source includes medical image (the primary part of patient's data) and also disease risk factor, therapy procedure, follow-up data and etc. Therefore, AI application are frequently employed in cancer research and prediction, particularly when using machine learning and deep learning approach. Due to the success of machine learning, there has been an increase of research into prognostic and diagnostic medical imaging. Meanwhile, as compared to machine learning, deep learning has improved the ability to analyse and detail interpretation of medical image.

2.5.1 Classic Machine Learning

As for classic machine learning in **Figure 2.21 a)**, regions of interest (ROI) or volume of interest (VOI) were either semi-automatically or manually described in segmented medical images before large-scale hand-crafted feature extraction. The characteristics of ROI/ VOI are captured by morphometric measurement such as size, diameter, shape, and measurements of tissue, before going through the process of feature pre-processing and normalization, feature selection and lastly create predictive modelling. Typically, the features are not robust to medical image acquisition factors including spatial resolution, through-plane resolution and image extracting setting. Therefore, classic machine learning is largely focused on preset attributes and often designed to solve specific issue only.

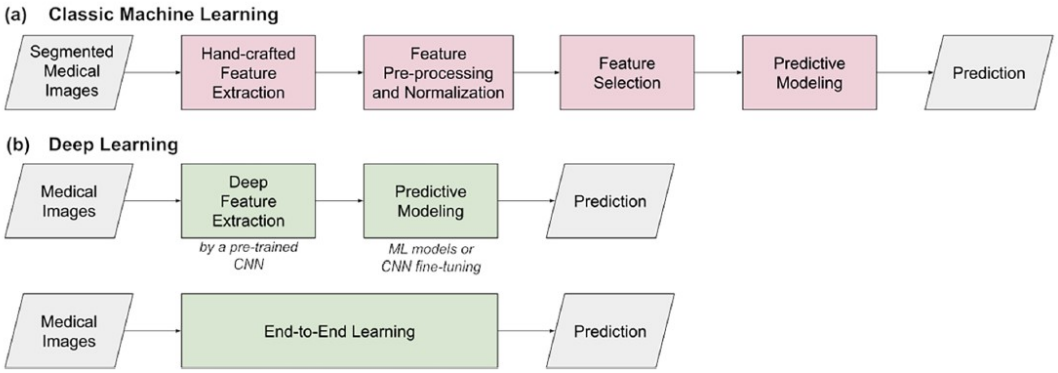


Figure 2.21: Common workflow of artificial intelligence a) classic machine learning b) deep learning either used as end-to-end learning or image feature extraction
(Castiglioni et al., 2021)

2.5.2 Deep Learning

Meanwhile, deep learning models as shown in **Figure 2.21 b)**, allows for the automatically image feature extraction to maximize the performance of the model. Deep learning is an artificial neural network that allows direct processing of raw data. Deep neural networks enable the development of predictive models either through end-to-end learning or deep feature extraction. Deep learning achieved the result by interpreting features by using and merge the data. Due to its capacity to model complex connections by utilising vast datasets, deep learning is widely used in extracting features from medical imaging. Deep learning has replaced conventional machine learning algorithms includes decision tree, support vector machines (SVM), etc. due to greater efficiency and popularity in image classification.

2.6 Diagnostic for Image Classification of Breast Cancer using Deep Learning

In mammogram images, to be able to detect and diagnose the breast mass into benign or malignant is the most challenging task to the radiologist. **Figure 2.22** displays some examples of mammogram image of region of interest (ROI) for benign or malignant masses. The benign masses usually well-defined, round and smooth boundary. Meanwhile, malignant masses usually blurred, spiculated and rough (Liu & Tang, 2014).

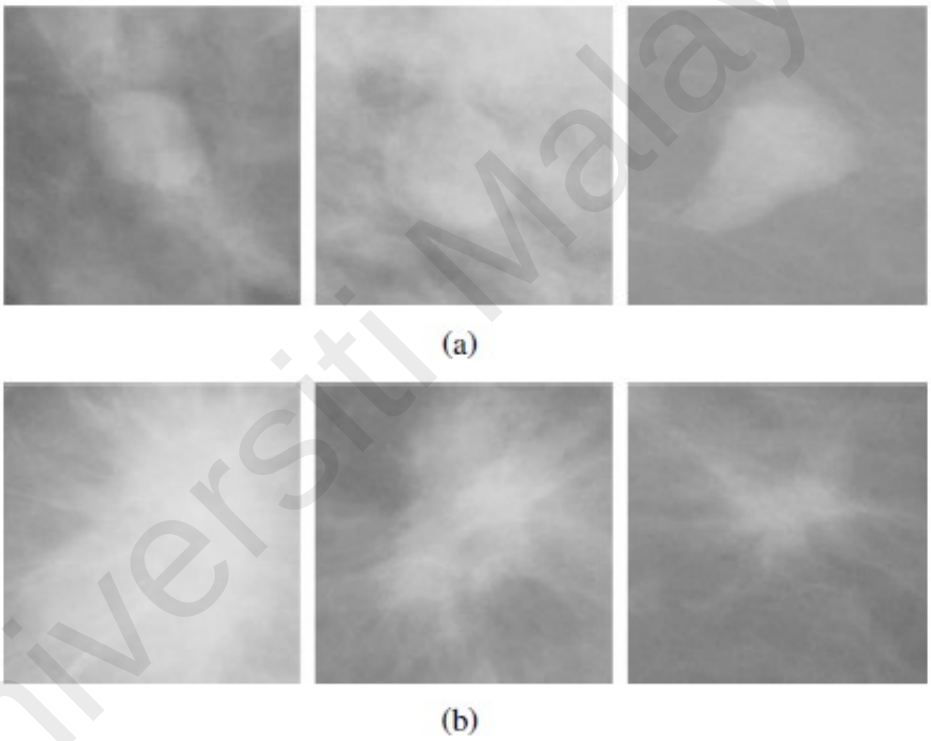


Figure 2.22: Example of a) benign and b) malignant masses in mammogram region of interest (ROI)
(Xiaoyong Zhang et al., 2017)

Conventional image classification of breast mass aim to be able to discriminate between benign or malignant based on the above description. However, due to the complex characteristic, it's hard to locate specific feature of the breast masses that indicate benign or malignant. Moreover, this specific feature is hard to quantify since

mammogram have limited image contrast between the masses and their surrounding tissues.

2.6.1 Convolutional Neural Network

Convolutional neural network (CNN) is a deep learning network architecture that learns directly from input data without the need for manual feature extraction, and commonly utilised in medical image processing tasks. CNN provide an effective architecture to detect and extract key features in interpret medical imaging knowledge, and has an advanced technology in the application of medical imaging to detect the presence of cancer from medical image and pathology reports. CNN have demonstrated knowledge acquisition in semantic segmentation, object detection, and image classification, to improve diagnostic accuracy and clinical workflow quality. These findings will assist radiologist to enhance sensitivity and reduce false positives.

CNN with a supervised deep network type is the most prominent and attracted the researcher into medical image processing experiments (Li et al., 2020; Pacal, Karaboga, Basturk, Akay, & Nalbantoglu, 2020; Tian & Fu, 2020). In experiments, CNN can incorporate feature extraction from medical imaging into the training process automatically, and has the benefits of shared weights, pooling and preferred method of analysis. Image classification and object detection is the most commonly used to diagnose cancer from medical images in which, image classification able to classify the abnormalities into specific categories automatically, and object detection allow to determine the location of abnormalities in the images. These functions have been shown to work effectively and successfully in classifying and detection such as supernumerary teeth in maxillary incisor region (Kuwada et al., 2020) and radiolucent lesions in the mandible (Ariji et al., 2019).

The main factor CNN is appropriate for medical image processing is because the spatial relationship is preserved even after the input medical image has been processed and modified. Spatial relationship is important in medical imaging analysis for distinguish between normal and cancerous tissue. Therefore, CNN have the ability to perform well in many areas of medical field including diagnosing for breast tumour. The three main CNN layers is convolutional layer, pooling layer and fully connected layer. The extracted features transfer the information obtained in convolutional layer within the network, and the classification executed by fully connected layer.

2.6.2 Convolutional Neural Network Architecture

Convolutional neural network (CNN) often utilised for classification, computer vision, and most frequently used deep learning algorithms in medical image analysis, since the network preserves the spatial relationship even after filter the image. It consists of multiple layer such as convolutional layer, rectified linear unit (ReLU) layer, pooling layer, and fully connected layer as shown in **Figure 2.23**.

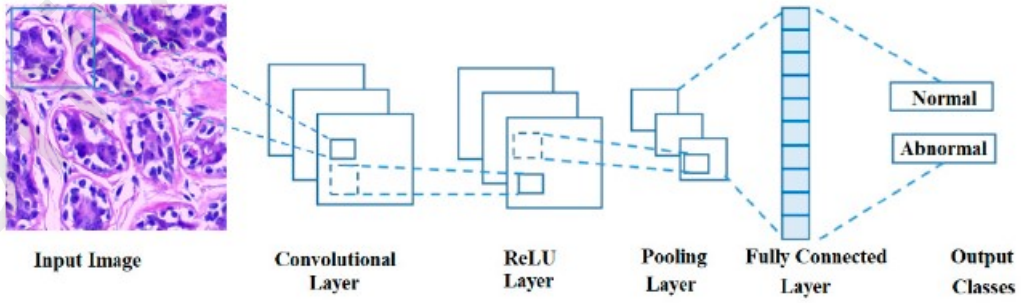


Figure 2.23: Convolutional neural network mail layers (Alzubaidi et al., 2020)

2.6.2.1 Convolutional Layer

The convolutional layer employs kernel (filter) to an input image and create a feature map or activation map, that summarize the presence of detected features in the image such as straight line, a dot or curve edge. This layer performs a convolutional operation on the input, a process that calculate pixel value in the respective area into single value based on the stride length and kernel size as it moves across the image's respective field, and send the output to the next layer. The image size was reduced as this layer combine information into single value.

2.6.2.2 Pooling Layer

The pooling layer known as down-sampling operation, and typically used to minimise the number of parameters in the input, performing dimensionality reduction in the network. Max pooling and average pooling are the most common pooling layer. As the filter moves across the input, max pooling selects the highest value to send to the output array, meanwhile average pooling takes the average value within the respective field to send to the output array. Pooling layer usually located between convolutional layer and rectified linear unit (ReLU) layer. Although a lot of information lost, this layer provides the benefit to reduce complexity, improve efficiency and reduce the risk of overfitting.

2.6.2.3 Rectified Linear Unit (ReLU) Layer

The rectified linear unit (ReLU) is an activation function that adjusts the value into zero if there is negative value in the input, with mathematic equation of f(x) = max. f(0, x). It intends to introduce non-linearities into the network. Therefore, ReLU layer simplify the training which make it easier to compute, and prevents the vanishing gradient problem in sigmoid activation function. After each convolution operation, CNN applied ReLU transformation to the feature map.

2.6.2.4 Fully Connected Layer

The fully connected layer placed at the end of CNN architecture and flatten the input prior to classification, with all neurons in this layer linked to neurons with previous layer. This layer used the previous layer output and generates the probability for classification into one of the output classes using SoftMax activation function, based on features extracted through previous layer and different filter.

2.6.3 Deep Convolutional Neural Network Image Classification

Deep learning with convolutional neural network or deep convolutional neural network (DCNN), has accomplished remarkable performance in classification task, which is the most effective method in deep learning. DCNN can automatically acquire classification rules and feature representation from large training dataset.

The convolutional technique allows simplifies image with millions of pixels into compact feature map, decreasing the input data dimensions while preserving the significant features. The use of DCNN to classify breast tumours is not a new development in research. During early of the research, the researcher attempted to execute the DCNN on the whole breast mammogram. Due to the reduction of the image, it have severe the classification effectiveness based on the research done by Zhou et al. (2017) with the result of 60.90%. However, the classification performance of whole breast mammogram improved by combining DCNN with data augmentation and transfer learning method (Xiaofei Zhang et al., 2017; Zhang et al., 2018).

Most of the research of DCNN to classify the breast tumours is focusing on small patches, known as region of interest (ROI) (Bakkouri & Afdel, 2017; Lévy & Jain, 2016; Xiaoyong Zhang et al., 2017). ROI is an area that probably have tumour. ROI often created on the basis of either clinical information or automatic segmentation from the whole breast mammogram. Xiaoyong Zhang et al. (2017) employed DCNN using pre-

trained AlexNet architecture on ROI achieved the result that is closed to the performance of radiologist. However, study done by Lévy and Jain (2016), based on three models architecture which is DCNN baseline model, AlexNet and GoogLeNet for breast masses classification, GoogLeNet outperforms other model with accuracy of 92.9% compared with DCNN baseline model (60.4%) and AlexNet (89.0%). Other method that Bakkouri and Afdel (2017) used is to normalization and scaling of ROI using Gaussion pyramid processing which can enhance the detail of feature extraction in different scales. The reason is because GoogLeNet have deeper and robust architecture where the networks can learn more information of the image and discriminant features (Rampun, Scotney, Morrow, & Wang, 2018). Overall, mammography classifying of breast tumours using ROI yield a rather good performance.

2.7 Optimization Technique

In recent years, DCNN are still continuously research extensively where they offer support for classification challenges. The main concern in employing DCNN models is that they require a large number of training data to give high accuracy performance. Sometime, acquire vast number of training data can be tedious. Therefore, optimizing technique need to be employed to solve the lack of training data.

The amount of training data determines the performance of the DCNN. DCNN requires a vast amount of data for it to perform well. The biggest barrier of implementing the DCNN is the shortage of training images in the medical field. It is tedious to collect medical images and require a specialist to label the images.

2.7.1 Transfer Learning

Transfer learning is technique which re-use the information collected by a trained model. Trained DCNN is a process of model trained in a large training data, and then the model is fine-tuned on a smaller training data as shown in **Figure 2.24**.

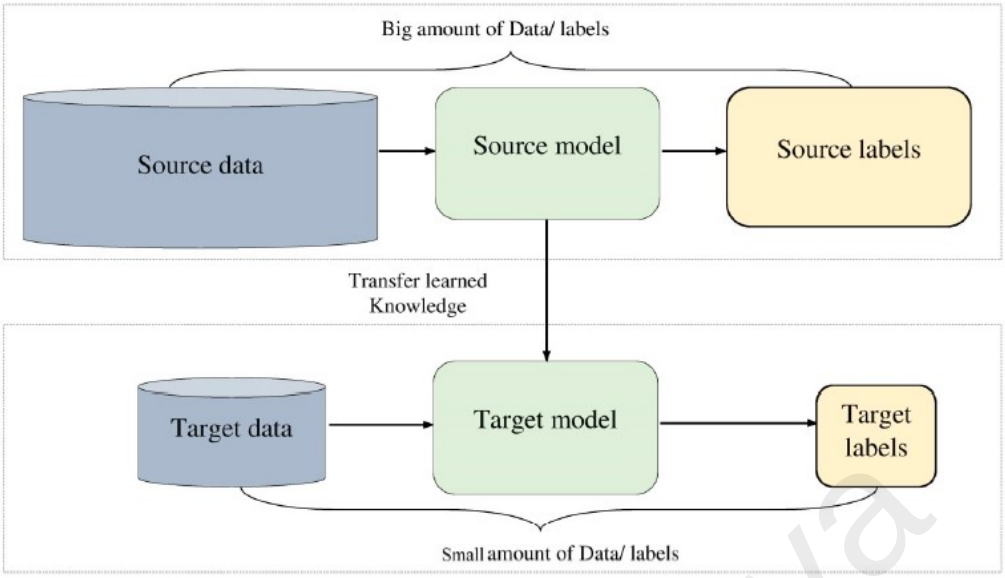


Figure 2.24: Transfer learning technique (Alzubaidi et al., 2020)

The first method of transfer learning is to fine tune certain parameter in some layers at the trained DCNN. Lévy and Jain (2016), Xiaoyong Zhang et al. (2017) and Rampun et al. (2018) pre-trained AlexNet using ImageNet database with 1,000 classifications which slightly improve the performance by using different source and target domain; and then fine-tuned by replacing 1000 neurons in the last of fully connected (fc8) AlexNet with 2 neurons which define as benign and malignant as show in **Figure 2.25**. The result shows a fine-tuned and pre-trained AlexNet is capable of extracting the features of benign and malignant masses. Lévy and Jain (2016) used pre-trained GoogLeNet on the ImageNet dataset and fine-tuned the last fully connected layer with binary output classes, and also removed two auxiliary classifiers from the GoogLeNet as they can weakened the training.

However, Alzubaidi et al. (2020) found that pre-trained the DCNN can significantly improve the performance if the transfer learning using the same source and target domain compared to different source and target domain.

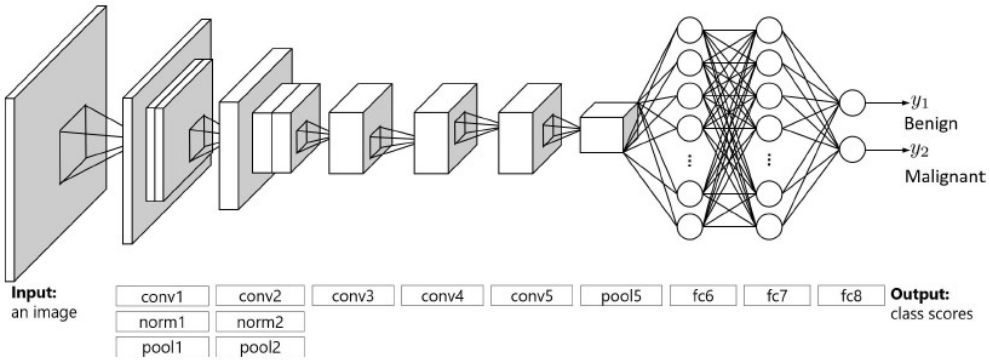


Figure 2.25: Pre-trained AlexNet architecture that have been modified by replacing fully connected layer fc8 from 1000 neurons to 2 neuron that defines as benign and malignant (Xiaoyong Zhang et al., 2017)

The second method is using the trained DCNN to do the calculation of the feature maps of new datasets. Due to the limited color structures and distribution, mammography image differs from the natural image data, but the fundamental of image features may still be leveraged by its shapes and edges which can be identified by trained neural network algorithms.

2.7.2 Data Augmentation

Deep convolutional neural network normally requires a vast amount of data to achieve high performance. However, only small number of training data found in medical field. The data can be increased by producing additional data from the original input data, also known as data augmentation. This technique is utilised to increase the number of training data by applying different view of image processes such as rotation, flipping, contrast and etc. This method also help to prevent overfitting problem that usually occurred on deeper neural networks (Krizhevsky, Sutskever, & Hinton, 2017). Zhang et al. (2018) and Zhou et al. (2017) used horizontal flipping and rotation by 90º, 180º and 270º for each of training data. Meanwhile, Alzubaidi et al. (2020) used horizontal flipping, vertical flipping and rotation technique where the image rotates by the angles of 45º, 90º, 135º, 180º, 225º, 270º and 315º as shown in **Figure 2.26**. As for Rampun et al. (2018), for each

of the ROI bounding box (P1), they also performed; (1) double the dimension of P1 (P2), (2) the tumour region of P1 only by removing background (P3), (3) zoom-in 50% effect of P1 (P4), (4) zoom-in 70% effect of P1 (P5), (5) double of dimension of P2 (P6) and (6) five random rotation for each patch (P1 to P6) as shown in **Figure 2.27**. Overall, geometric transformation is popular technique is effective to increase the size of data such as rotation, shifting, flipping, mirroring and etc., due to the simplicity and fast technique (Bakkouri & Afdel, 2017; Lévy & Jain, 2016).

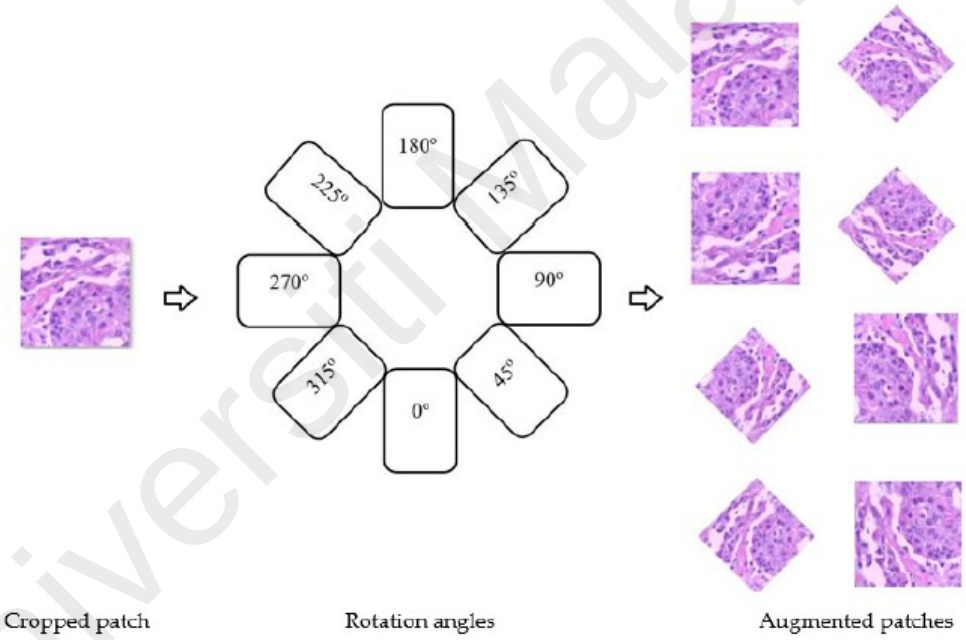


Figure 2.26: Data augmentation rotation process involves rotation (Alzubaidi et al., 2020)

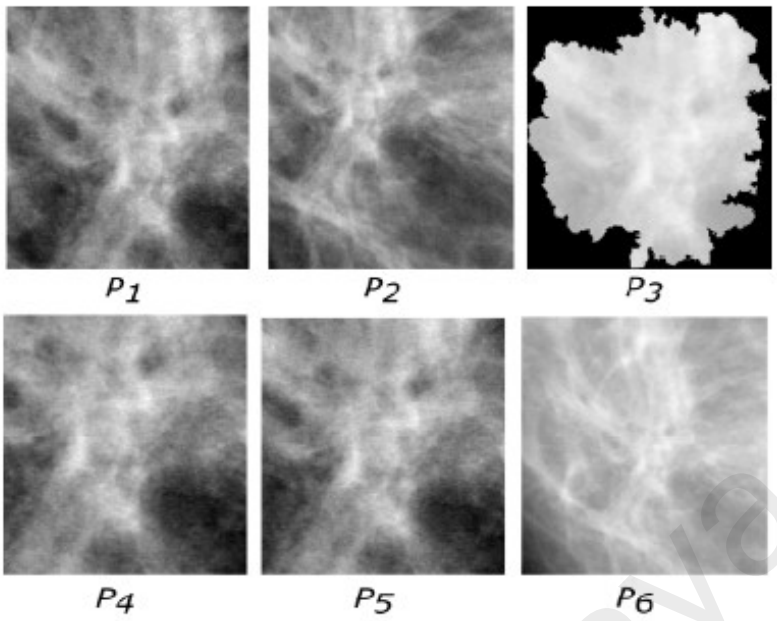


Figure 2.27: Example of breast mass after data augmentation (Rampun et al., 2018)

Such data augmentation provides relevant and justified training data since the breast tumours might occurs in different direction and their diagnosis is invariant to these transformation (Lévy & Jain, 2016; Zhang et al., 2018). The data augmentation can be done before or during the training of data.

2.7.3 Mass Context

Another technique shown by Lévy and Jain (2016) is mass context which is the region around the ROI of breast tumours can also provide the useful context for diagnosis. There are two approaches to provide the context to the network, the ROI region (known as small context) and a region twice the size of ROI that provide the features of its surrounding (known as large context). The result shows that a large region slightly increases the performance compared with the ROI region only.

2.8 Summary

Cancer is a rapid formation of abnormal cells that expand their normal boundary and spread to neighboring region of the body. Breast cancer surpassed lung cancer as the highly prevalent diagnosed cancer globally in 2020. Malaysia has a high incidence rate of breast cancer with overall lifetime risk is one in every 27 women. Women are more prone to acquire breast illness such as benign (non-cancerous) and malignant (cancerous) compared to males. Majority of women breast cancer cases develop in lobules, milk ducts and also an axillary tail of Spencer. There are many types of breast cancer includes invasive lobular carcinoma (ILC), invasive ductal carcinoma (IDC), ductal carcinoma in situ (DCIS), lobular carcinoma in situ (LCIS), inflammatory breast cancer (IBC), metastatic breast cancer, etc. Before identifying as breast cancer, a doctor conducts tests and results of the test determine the type of treatment. The main treatment of cancers includes surgery, chemotherapy, radiotherapy, targeted therapy, and hormone therapy.

Imaging modalities aid in detect breast cancer includes mammography, digital breast tomosynthesis, breast magnetic resonance imaging (MRI), computerized tomography (CT), positron emission tomography (PET), hybrid PET/CT and chest x-ray. Mammograms are the primary imaging modality tools to do screening, diagnosing, evaluating and monitoring for breast cancer. However, the sensitivity of mammography was reduced to detect cancer with dense women breasts. Digital breast tomosynthesis is able to capture multiple angles of the breast, which increase the sensitivity and specificity to detect cancer even with dense breast. Ultrasound able to detect cancer in the dense breast but is unable to detect calcifications. Ultrasound is useful to perform an image-guided biopsy. MRI are more sensitive in detecting breast cancer cause more false positive. Therefore, MRI does not suit the average risk as screening and is more suitable for high risk. Hybrid PET/CT provide detailed information of cancer with enhanced sensitivity and specificity, compared to CT or PET alone. Mammography is the gold

standard for diagnosing breast cancer among all diagnostic imaging modalities. With images from the CC and MLO view, radiologist can evaluate different types of breast changes such as calcifications and masses.

In biomedical research, artificial intelligence (AI) is increasingly participating in classification to detect breast abnormalities with the using of medical imaging. Deep learning is a subfield of machine learning and frequently employed in research such as medical imaging processing to execute the task as image classification. CNN is the typical deep learning network algorithms and has an advanced technique to extract hidden patterns and interpret medical imaging knowledge to classify breast cancer. CNN comprises of a fully connected layer, rectified linear unit (ReLU), pooling layer and convolutional layer to analyze medical images. Optimization technique also discussed since deep CNN require a vast number of the training dataset. Based on the presented reviews, the development model of image classification of breast mass using mammography could be achieved.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter discusses the process of developing breast cancer classification using deep learning approach from mammogram images. The chapter comprises four main stages from overall flowchart as shown in **Figure 3.1**: data collection, preprocessing data, DCNN model and training the DCNN model. This chapter begins with data collection, where the data are obtained. The preprocessing data of Section 3.2 describe the detail of the raw dataset and preparation of the data including data augmentation. The dataset was organized and structured to create a training and test dataset. Section 3.4 describes the specification of the hardware, and the deep learning system software is elaborated in Section 3.5. The next section explains about DCNN architecture, transfer learning with fine-tuning of pre-trained AlexNet, GoogLeNet and VGG-16. Section 3.7 explain about importing of the dataset and train configuration of the pre-trained model. Finally, this chapter concludes the methodology in Section 3.8.

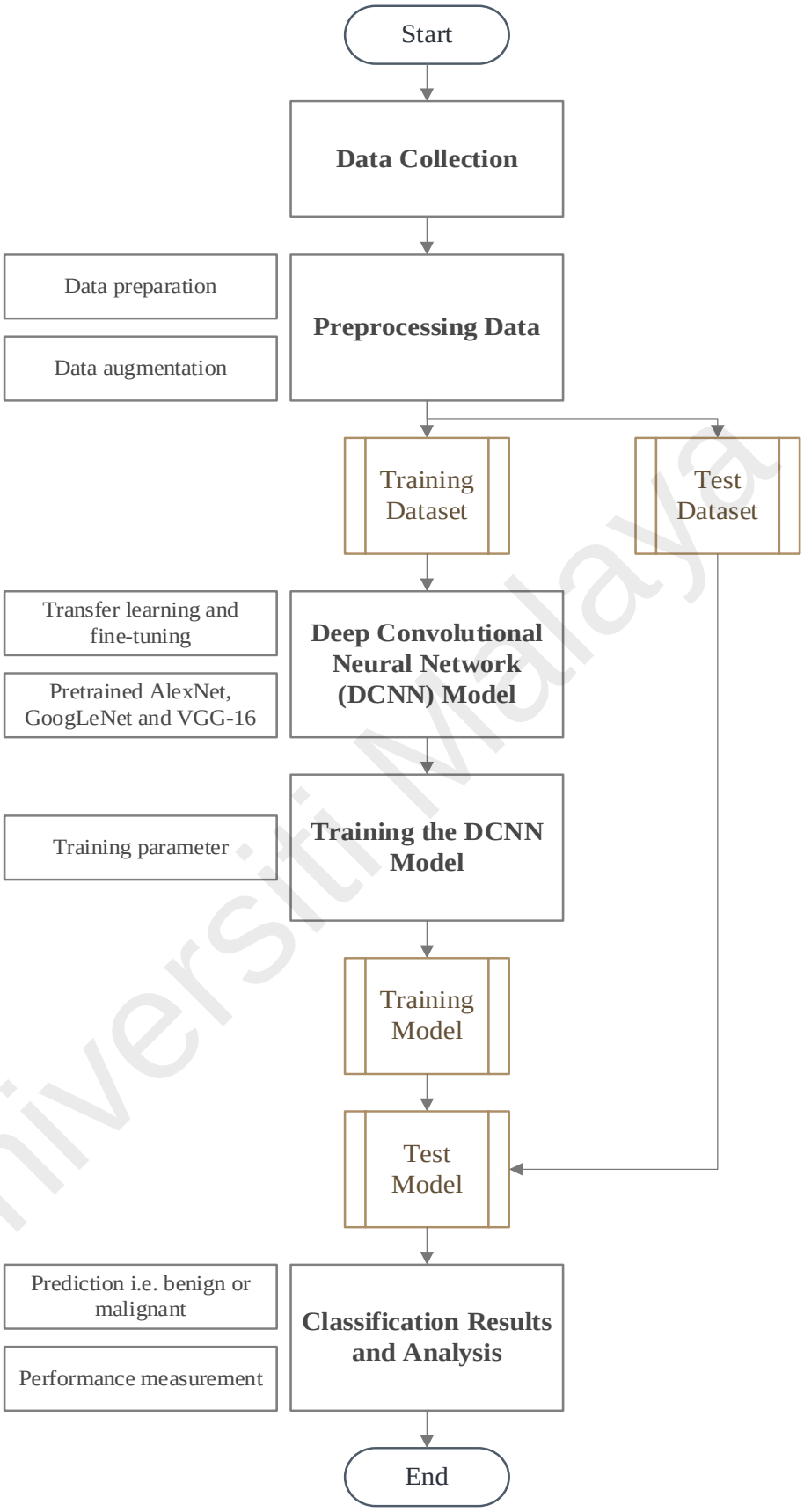


Figure 3.1: Overall flowchart of the study

3.2 Data Collection

In this study, the dataset was obtained from the publicly available Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) (R. S. Lee et al., 2017). CBIS-DDSM is an upgraded version of DDSM that comprises of digitized film mammogram in outdated lossless-JPEG format. Previously, DDSM was a database of 2620 mammographic images that contain normal, benign and malignant, as well as valid pathology information. However, due non-standard file compression and non-precise segmentation of tumours, most of the researcher before need to construct segmentation algorithms to extract the precise features and make it impossible to compare with performance of previous results (Zhou et al., 2017). Therefore, CBIS-DDSM database was created to overcome the issue by standardised the DDSM data and have been curated by trained radiologist for medical research especially for computer-aided diagnosis and detection (CAD).

3.3 Preprocessing Data

3.3.1 Raw Dataset

CBIS-DDSM contain 6,775 studies of grayscale mammographic images from 1,566 women in standard Digital Imaging and Communications in Medicine (DICOM) format with the total size of 163.6 GB. The CBIS-DDSM database images were obtained from The Cancer Imaging Archive (TCIA) website (https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM). The total number of images as stated in the website is 10,239 images. The CBIS-DDSM database includes the images of whole breast mammograms in CC and MLO view, region of interest (ROI) mask images and ROI cropped images, as shown in **Figure 3.2**. Range sizes of whole breast mammograms is 9.7 - 71.18 MB, ROI cropped images is 93.77 KB – 14.08 MB and ROI mask images is 4.91 – 41.25 MB. Each ROI label with calcification or mass, and

pathologically validated category as benign, benign without callback and malignant. For this study, benign without callback shall be categorized as benign.

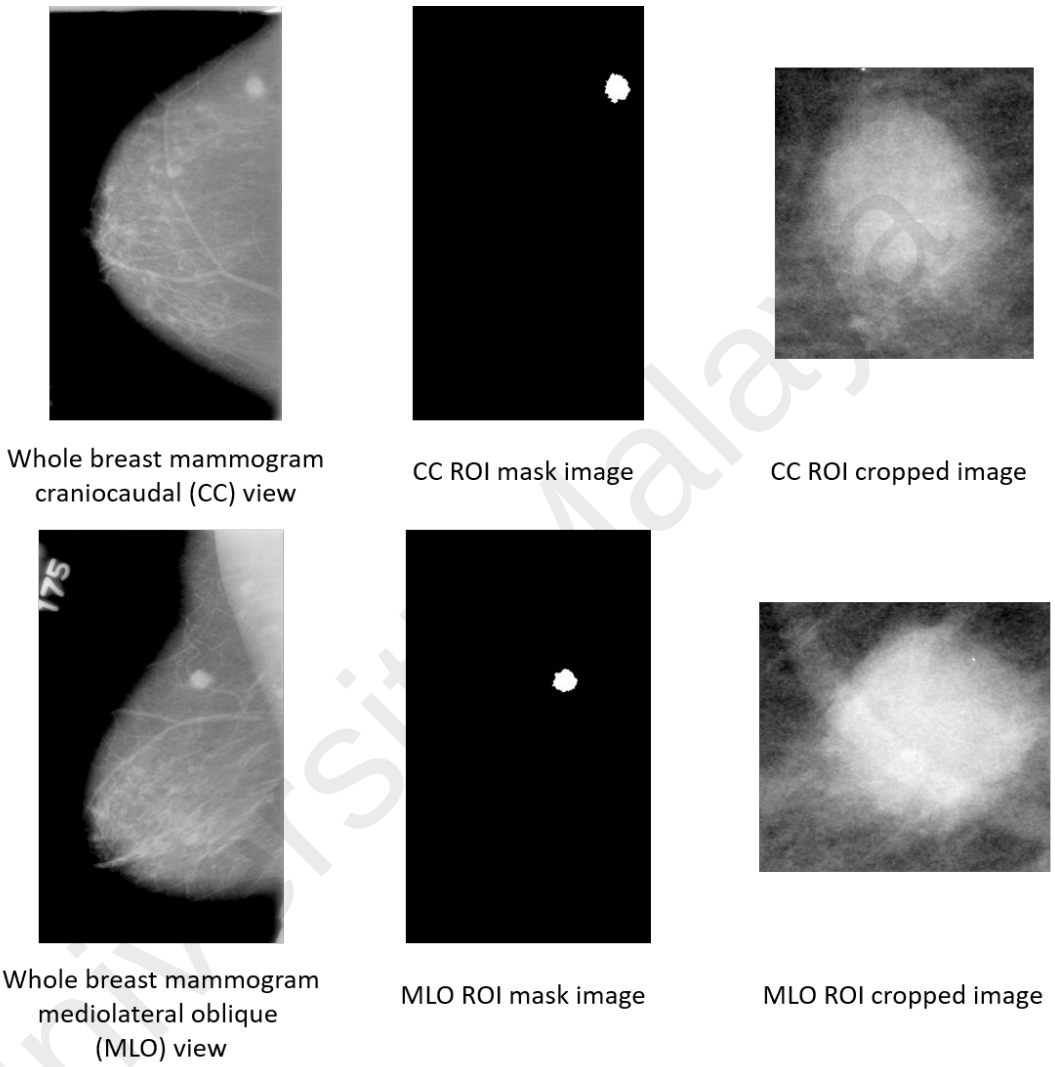


Figure 3.2: Types of images in CBIS-DDSM database

The database consists of 12-bit DICOM format and .csv files including metadata of images, and description for calcification training, calcification test, mass test, and mass training as **Table 3.1**. 4 types .csv datasets file provide important information includes patient ID, breast density, breast location, image view (CC or MLO), abnormality type (calcification or mass), type or distribution of cancer and pathology of disease (benign or

malignant). The method of downloading DICOM images is by using NBIA Data Retriever software provided in the website after TCIA provides the list of images in a manifest file.

Table 3.1: Types of data available in CBIS-DDSM

No.	Dataset Images Type in DICOM	CSV File
1	Calcification-Test Full Mammogram, ROI cropped and mask images	Calcification-Test Description
2	Calcification-Training Full Mammogram, ROI cropped and mask images	Calcification-Training Description
3	Mass-Test Full Mammogram, ROI cropped and mask images	Mass-Test Description
4	Mass-Training Full Mammogram, ROI cropped and mask images	Mass-Training Description

3.3.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is a technique for examining data and summarize essential properties using data visualization approaches such as bar charts, line charts, pie charts, etc. The main objective of EDA is to uncover data, greater understanding of dataset variables and relationships among the variables. EDA may assist in determining confidence intervals, categorical variables, and standard deviations. After EDA analyzed, the results can be used for more complex data modelling or analysis on deep learning and machine learning.

In this study, exploratory data analysis (EDA) conducts in Python 3 language using JupyterLab. JupyterLab is the next-generation notebook interface of Jupyter Notebook,

an open-source web-based interactive user interface for Project Jupyter to create development environment for notebooks, code and data. JupyterLab version 3.2.4 has the capabilities of in-line visualizations, which is necessary to observe a specific output and beneficial while performing EDA.

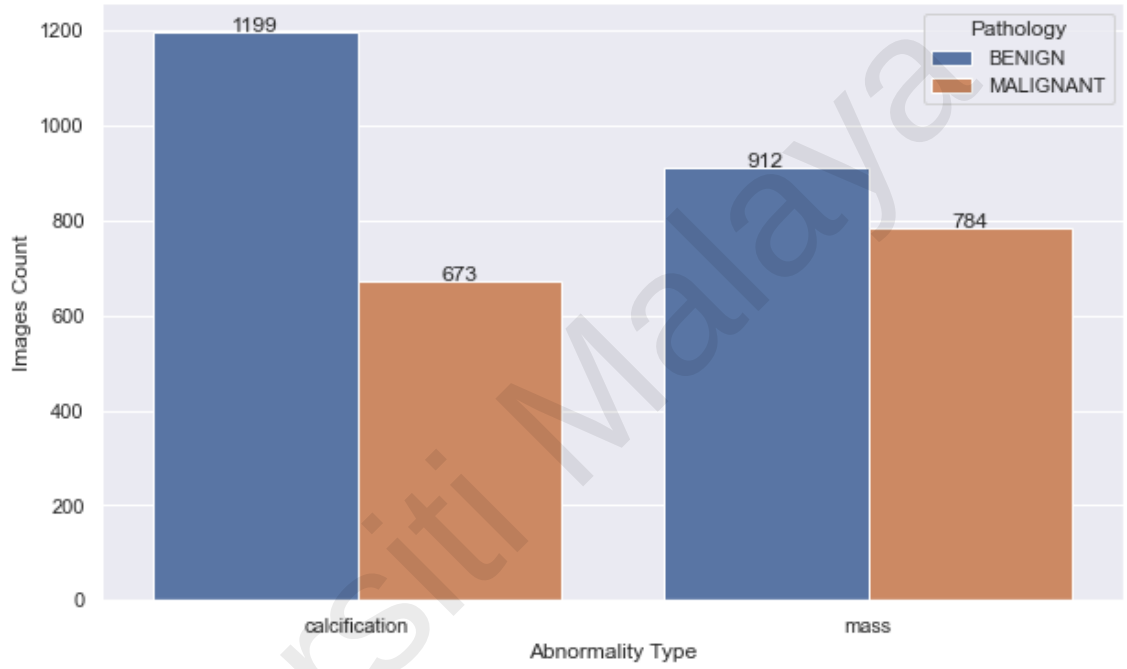


Figure 3.3: Count of pathology based on calcification and mass abnormalities

The .csv files of 4 types of datasets are used to analyse the data and their relationship. **Figure 3.3** shows the bar chart for the image count of calcification and mass abnormalities in CBIS-DDSM database. Based on the bar chart, majority images are benign compared to malignant.

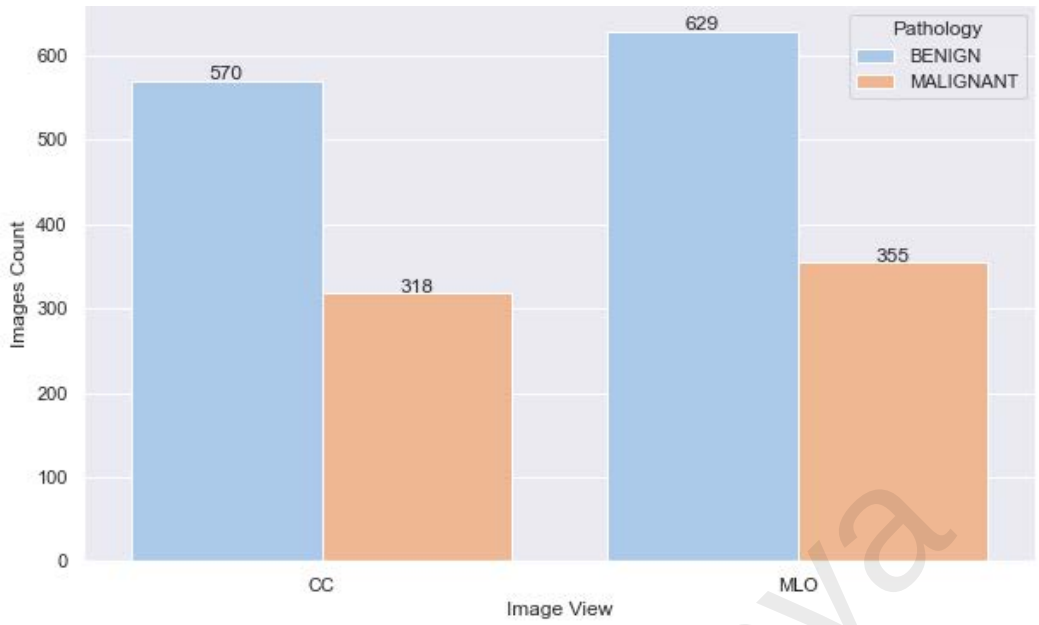


Figure 3.4: Count of pathology based on image view for calcification

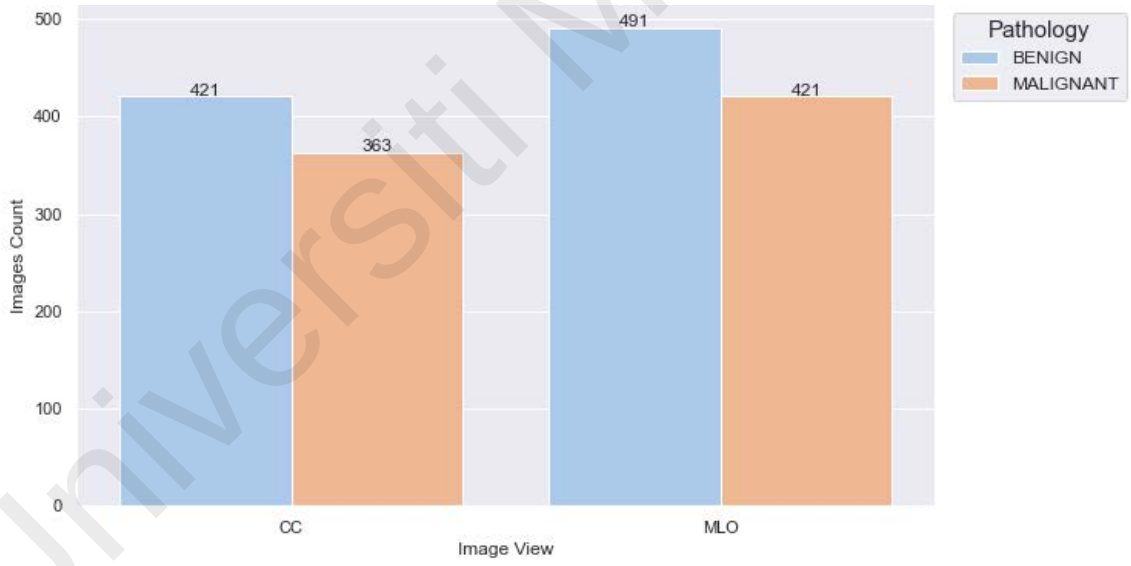


Figure 3.5: Count of pathology based on image view for mass

Based on **Figure 3.4**, the calcification dataset has more benign cases than malignant in both CC and MLO views. Meanwhile, as for the mass dataset, both benign and malignant has nearly equal proportion of image as shown in **Figure 3.5**. Summary of images both calcification and mass dataset as per **Table 3.2**.

Table 3.2: Summary of total images in CBIS-DDSM database

Abnormality	Pathology	CC	MLO	Sub-Total	Total
Calcification	**Benign**	570	629	1199	1872
	Malignant	318	355	673	
Mass	**Benign**	421	491	912	1696
	Malignant	363	421	784	
Total		1672	1896	3568	

Breast density is one of the assessments in mammogram report. Breast density is a correlation of fibrous and glandular tissue to fatty tissue in the breast. According to 2003 BI-RADS Atlas (Melnikow et al., 2016) in **Figure 3.6**, there are 4 categories for breast density which is type 1 = entirely fat in the breast (less than 25% glandular), type 2 = scattered areas of fibro glandular densities in the breast (estimated 25% - 50% glandular), type 3 = heterogeneously dense breast and may obscure small masses detection (estimated 51% - 75% glandular), type 4 = extremely dense breast and may reduce the mammography's sensitivity (above 75% glandular).

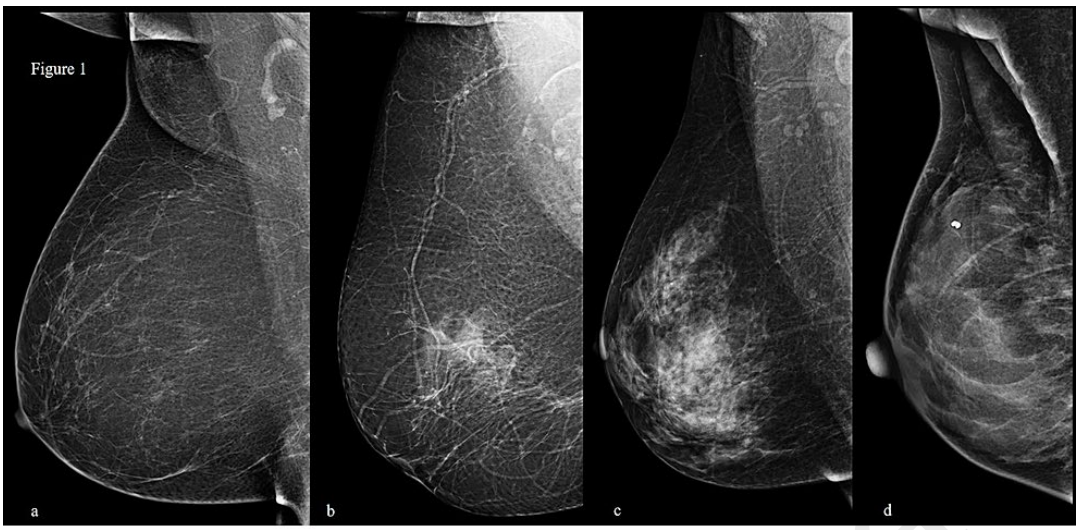


Figure 3.6: BI-RADS breast density (a) Type 1 – entirely fat in the breast. (b) Type 2 – scattered areas of fibro glandular densities. (c) Type 3 - heterogeneously dense breast. (d) Type 4 – extremely dense breast (Fatima, Mohsin, Rao, & Alvi, 2021)

The denser the breast, the harder to detect abnormality on a mammogram. Based on **Figure 3.7**, the bar chart shows that most abnormalities are in type 2 breast density. In the database, mostly benign detected in all of category breast density. Among all breast density category, type 4 is among the lowest in detecting the abnormality. It may be due to the extremely dense breast that the breast cancer is hard to detect by a mammogram and can mask a potential cancer. Moreover, dense breast is likely to develop cancer as compared to less dense breast.

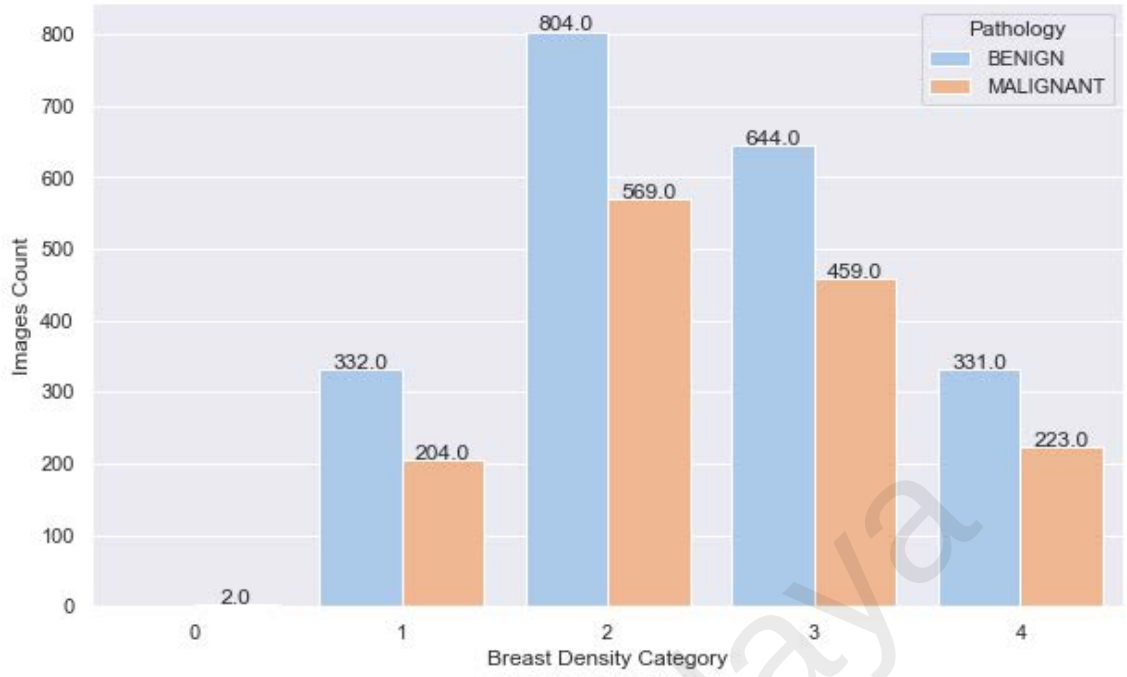
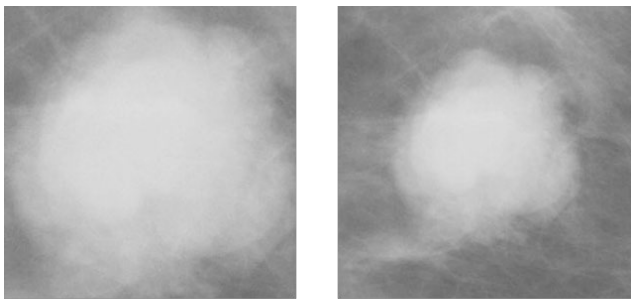


Figure 3.7: Count of pathology based on breast density

3.3.3 Data Preparation

In this study, the experiment conducted only using mass disease only. Mass abnormality which includes mass-test and mass-training images, was chosen as a dataset from CBIS-DDSM database. Mass images comprise of 912 benign and 784 malignant in CC and MLO view.

Instead of using mass in whole breast mammogram as input, this study uses breast mass patch as DCNN input images. This is due to the detail features of mass abnormality may be lost as the image of whole breast mammogram (original resolution around 3000 x 6000 pixels) is reduced as an input for DCNN. The smaller breast mass patch may become undetected restricting the classification model's ability. Furthermore, many breast mass features may be extracted which enhance the performance of the DCNN. The performance of DCNN may be improved as the breast mass patch can be extracted into multiple images due to data augmentation.



(a) (b)

Figure 3.8: (a) breast mass patch without surrounding area. (b) breast mass patch with surrounding area, which provide information for feature extraction.

The breast mass patch was extracted using image of whole breast mammogram and overlay with ROI mask image using Adobe Photoshop. The breast mass patch images were cropped to 256 x 256 pixels, including the area surrounding the breast mass patch as it may provide information that may useful for feature extraction. **Figure 3.8** shows breast mass patch without surrounding area and with the surrounding area. The images were converted from the original DICOM format into JPEG (Joint Photographic Experts Group) format. This process is essential due to graphic processing unit (GPU) memory's constraint because of the large memory size of DICOM format (40 MB) compared to JPEG format (20 kB).

3.3.4 Data Augmentation

Deep learning requires a large quantity of data to improve classification performance and prevent overfitting. Therefore, this method is an important element in this study to increase the number of breast mass patch by rotating, flipping and zoom. The process of augmented data is carried out manually by using Adobe Photoshop. The strategies of data augmentation to fulfill the necessity for vast amount of data are as follows: -

i. P1: original breast mass patch.
ii. P2: breast mass patch of P1 while removing surrounding area. This allows the neural network to learn mass's feature and border.
iii. P3: breast mass patch of P1 with rotation of 45º.
iv. P4: breast mass patch of P1 with rotation of 90º.
v. P5: breast mass patch of P1 with rotation of 135º.
vi. P6: breast mass patch of P1 with rotation of 180º.
vii. P7: breast mass patch of P1 with rotation of 225º.
viii. P8: breast mass patch of P1 with rotation of 270º.
ix. P9: breast mass patch of P1 with rotation of 315º.
x. P10: breast mass patch of P1 with horizontal flipping.
xi. P11: breast mass patch of P1 with vertical flipping.
xii. P12: Zoom out by double the size of P1 dimension to preserve the surrounding area data around the mass patch.
xiii. P13: ROI cropped image.

Figure 3.9 shows the result of data augmentation of original breast mass patch, P1. As a result, each original breast mass patch was expanded to 13 images. Since breast mass patch does not inherit orientation, their diagnostic was unaffected by the changes due to rotation and flipping, therefore these augmentations are justified. All images are fixed to 256 x 256 pixels.

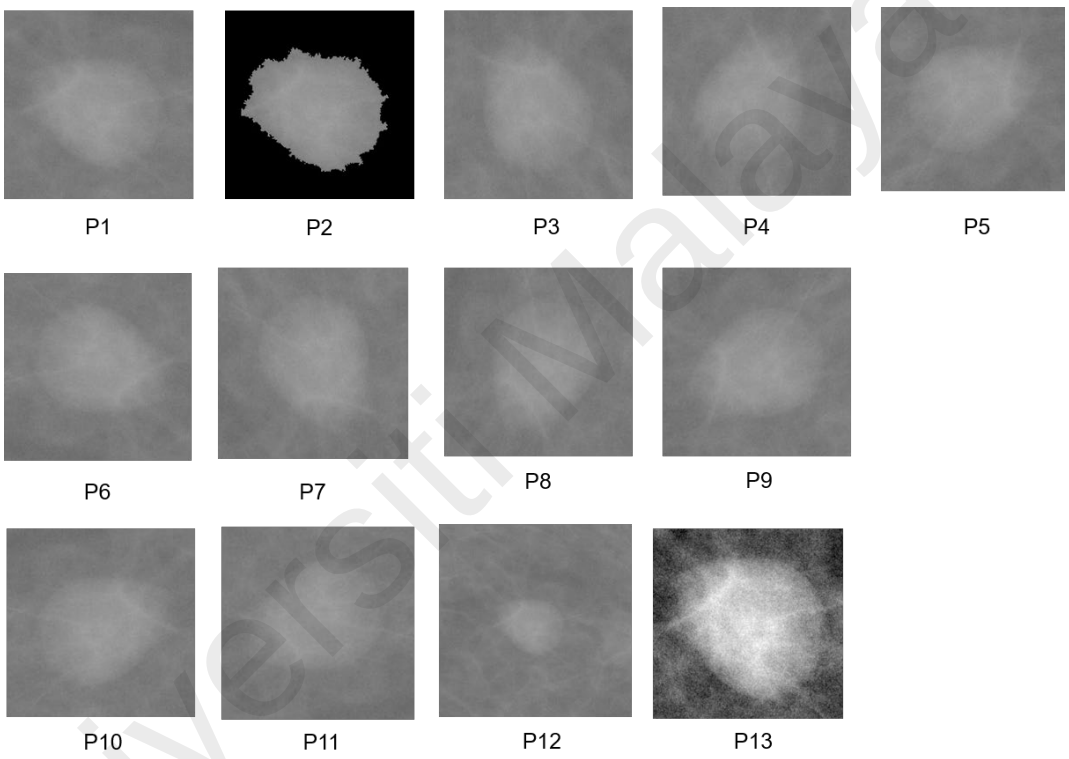


Figure 3.9: Example of malignant breast mass patch after data augmentation

3.3.5 Data Organizing and Structuring

In this study, the datasets are divided into 2 categories which is training and test dataset. Training dataset refer to the sample used to develop a learning model, whereas the test dataset used to validate performance of a learning model. Test dataset is separated from training dataset until learning model is completed.

In training dataset, a total set of 13,260 breast mass patches extracted from the data augmentation. The training dataset is randomly divided to three sections: 60% of training set, 20% of validation set and 20% of test set, with equal proportion of benign and malignant. The training set is used to train the model the hidden features in the data. The model continues to learn the features using the same training set into the network in each epoch. The validation set is used to validate the model's performance throughout the training process. At the end of each epoch, the model is assessed on validation set throughout the training. The purpose of validation set is to avoid overfitting, where the model unable to generalize and make accurate classification on the unknown data. The test set is a collection of data used to test the model after being trained at the end of each epoch, to evaluate model performance in term of accuracy and loss.

The files for training dataset shall also divide into three files: train file, validation file, and test file, where each file shall have two pathology files (benign and malignant) for classification. The file's structure and the images quantity as shown in **Table 3.3**. The validation set can also put in together with training set. The validation set will be automatically selected in NVIDIA DIGITS later. So, in this case, only 2 folders created which is training set and test set.

Table 3.3: File structure for training dataset

Files	Pathology	Sub-Total Image	Total Image	Percentage (%)
Train	**Benign**	3978	7956	60%
	Malignant	3978		
Validation	**Benign**	1326	2652	20%
	Malignant	1326		
Test	**Benign**	1326	2652	20%
	Malignant	1326		
Total		**13260**		**100%**

Test dataset is a different collection of data used to test the trained model. The test dataset provides the model final performance in term of accuracy, precision and etc. The test dataset used ROI cropped image from the remaining images in the CBIS-DDSM mass database, which not being used in the training dataset. **Table 3.4** shows the image quantity in test dataset.

Table 3.4: Test dataset image quantity

Files	Pathology	Total Image
Test Dataset	**Benign**	270
	Malignant	270
Total		540

3.4 Hardware Specification

The deep learning system was built on NVIDIA graphic processing unit (GPU), GeForce GTX1060 6GB GPU on Intel® Core™ i7-8750H processor running Ubuntu 20.04 LTS operating system of 16GB of RAM. The laptop run dual boot system of both Ubuntu 20.04 LTS and Windows 10, however, the deep learning system only works with Linux operating system.

3.5 Deep Learning System Software

The deep convolutional neural network (DCNN) was trained, validated and tested on version 6.1.1 NVIDIA Deep Learning GPU Training System (DIGITS) using Caffe framework. NVIDIA DIGITS is a web server that provide a convenient interface to train DCNN for image semantic segregation, object detection, and image classification tasks. The benefit of NVIDIA DIGITS is to optimize deep learning activities includes data management, design and train DCNN on multi-GPU system, real-time performance with enhanced visualization and select the best performing model for the deployment. NVIDIA DIGITS is open-source software available on GitHub.

In this study, NVIDIA DIGITS runs on Docker container, which mean that DIGITS can only run on a system with NVIDIA GPU (Graphic Processing Unit). DIGITS runs on a fully configured Ubuntu (version 20.04.3 LTS), NVIDIA GPU driver (version 470.86), CUDA software (version 11.4), cuDNN toolkit, Docker (version 20.10.11) and nvidia-docker (version 2.8.0). The container image can be pulled from NVIDIA GPU Cloud (NGC) registry.

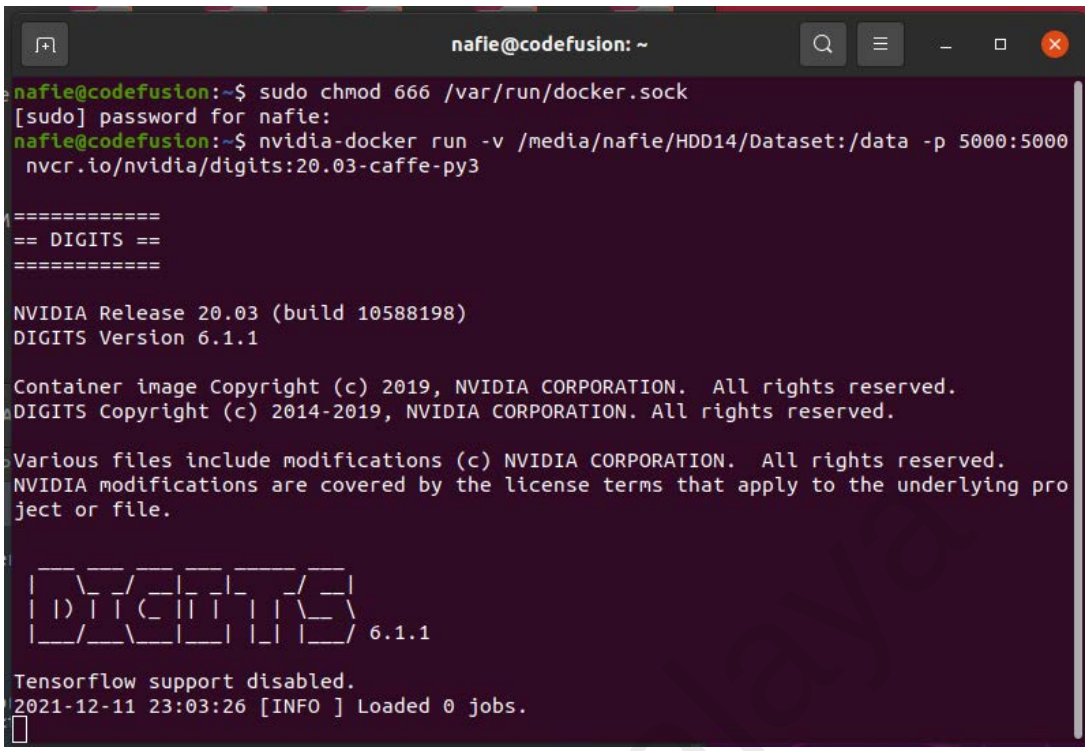


Figure 3.10: NVIDIA DIGITS launched through terminal

NVIDIA DIGITS also serves as an interface for common frameworks such as TensorFlow, Caffe and Torch. Caffe (Jia et al., 2014) is a deep learning framework built in C++ with Python interface that was designed at the University of California, Berkeley. Caffe framework was chosen for this study due to easy to learn and performs well for deep learning on image classification. NVIDIA DIGITS container together with Caffe framework (version 0.17.3) was pulled from NGC registry using version 20.03-caffe-py3 tag. The DIGITS container launched by mapping the folder that contains the training dataset that have training set and test set at Ubuntu terminal as shown in **Figure 3.10.** The dataset would become available at /data directory within the container.

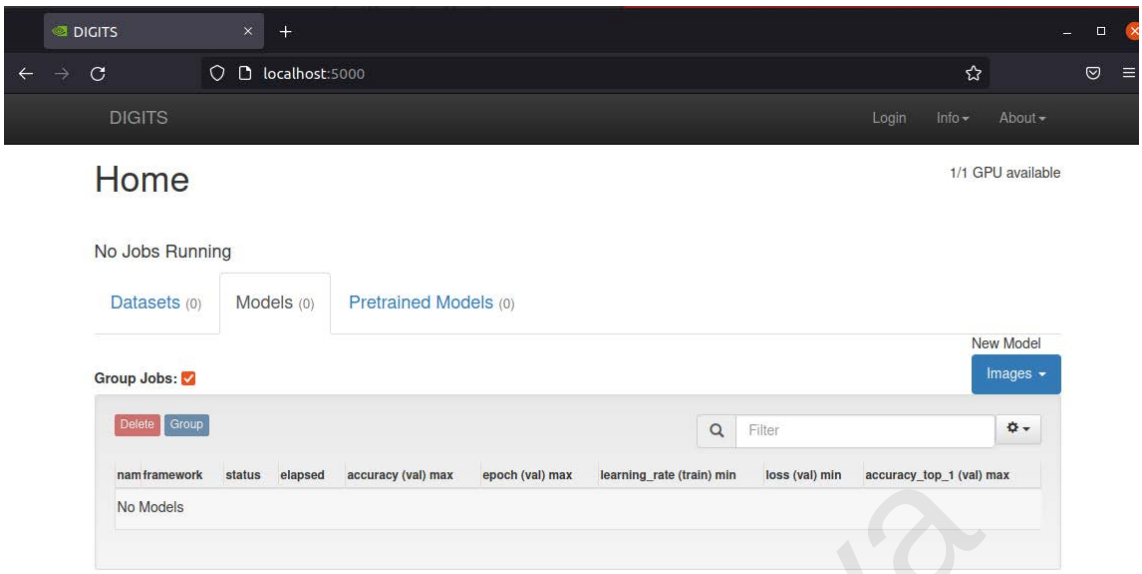


Figure 3.11: Home page of NVIDIA DIGITS

Once the container is running, DIGITS used web interface on localhost on port 5000 as shown in **Figure 3.11**. NVIDIA DIGITS includes well-known original DCNN architecture such as LeNet, AlexNet and GoogLeNet. Other than that, DIGITS also provide Model Store, which is a library of trained models that utilised as pre-trained weights using ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 dataset. In **Figure 3.**12, trained models can be imported into DIGITS from Model Store train the new data. The pre-trained DCNN architecture of AlexNet, GoogLeNet and VGG-16 were used in training process to execute classification task of benign or malignant, from breast masses patch.

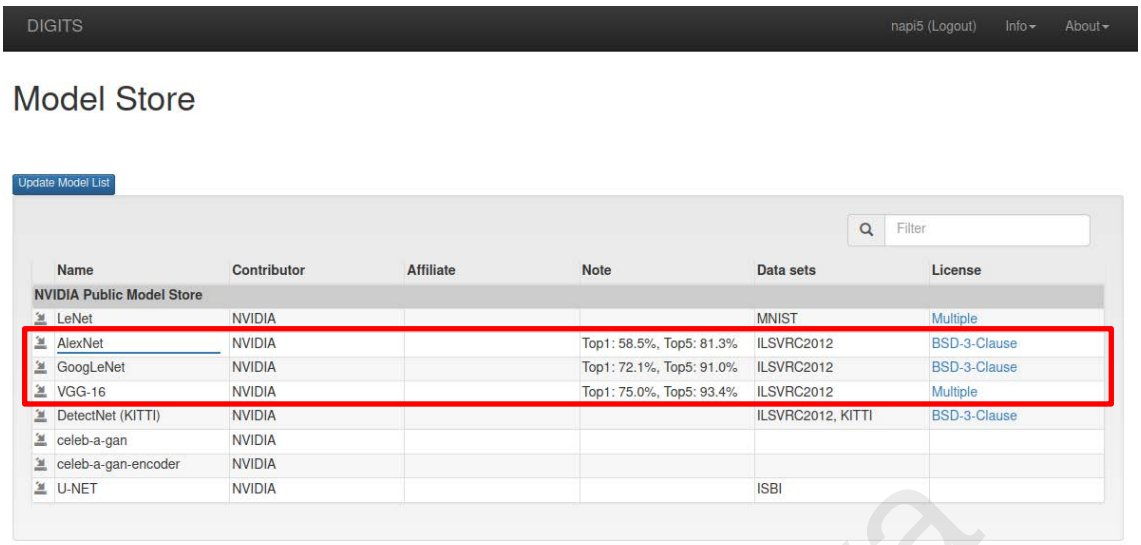


Figure 3.12: Model Store in NVIDIA DIGITS provide pre-trained model

3.6 Deep Convolutional Neural Network Architecture

3.6.1 AlexNet Architecture

AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) model was introduced by Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, and won the ImageNet Large-Scale Visual Recognition Challenge 2012. AlexNet accomplished top-5 error rate of 16.4% (83.6% accuracy) compared to second-best entry (non-CNN method) of 26.2%, and the first network to get error rate below 25%. AlexNet was pioneer that open the whole new research area of convolutional neural network (CNN) and made all entries of ImageNet competition using CNN for classification task.

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-
Fully Connected 1	-	-	-	-	4096	ReLU
Dropout 2	rate = 0.5	-	-	-	4096	-
Fully Connected 2	-	-	-	-	4096	ReLU
Fully Connected 3	-	-	-	-	1000	Softmax

Figure 3.13: AlexNet model architecture

AlexNet is comprised of 8 layers, which consists of 5 convolutional layers with max pooling and 3 fully connected layers as shown in **Figure 3.13**. Max pooling assists in the of overfitting by employs a max operation to pool sets of features, leaving the smaller number. Rectified Linear Unit (ReLU) is an activation function introduced in AlexNet, and used at each layer except the output layer, and by employing ReLU, it expedited the training process compared to tanh and sigmoid activation function. The model also avoid overfitting by made the used of dropout layer and data augmentation. In dropout layer, a neuron is dropped from the network with a probability of 0.5. So, each input is routed through a separate network architecture. As a result, the learnt weight is more robust and do not get overfitted easily. As for data augmentation, image have been created by using translations and horizontal mirroring, which increased the training dataset by 2048. AlexNet has a total of 62.3 million learnable parameters.

AlexNet received 256 x 256 pixels RGB as input. This implies that all images in the training dataset and test images need to be size 256 x 256 pixels. If the input image is grayscale, it needs to convert into RGB by reproducing the single channel into three-channel RGB image. Random crops of size 227 x 227 pixels were created from 256 x 256 pixels to feed the first layer of AlexNet.

3.6.2 GoogLeNet Architeccture

GoogLeNet model (Szegedy et al., 2015) model was developed by Google researchers and introduced at the ImageNet Large-Scale Visual Recognition Challenge 2014 to accomplish computer vision task including object detection and image classification. GoogLeNet achieved top-5 error of 6.67% (93.33% accuracy) in classification task, slightly less error than VGG-16 (runner up with top-5 error of 7.32%).

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Figure 3.14: GoogLeNet model architecture

GoogLeNet comprised of 22 layers DCNN model using a variation of 9 inception modules in the network as shown in **Figure 3.14**. By using variation of inception module, the GoogLeNet architecture has overcame the challenges of a large network where it is likely to overfit and reduce the problem of exploding or vanishing gradient. Inception module is a neural network architecture that used convolution layers with various filters to detect features at various scales. This network reduces computational process during training through dimensional reduction, thus drastically reduce the number of learnable parameters into 4 million. GoogLeNet received an image sized 224 x 224 pixels with RGB color channels as input.

In GoogLeNet, there is an additional component known as Auxiliary Classifier is used to regularize network and prevent it from overfitting and vanishing gradient. If vanishing gradient occurred, the network shall stop learning during training. Auxiliary Classifier was added in the inception 4(a) and inception 4(d) layers in the network, only used during training and removed during inference.

3.6.3 VGG-16 Architecture

VGG-16 (Simonyan & Zisserman, 2014) model was proposed by A. Zisserman and K. Simonyan from University of Oxford and also introduced at the ImageNet Large-Scale Visual Recognition Challenge 2014. VGG-16 accomplished top-5 error of 7.32% (92.68% accuracy) in classification task.

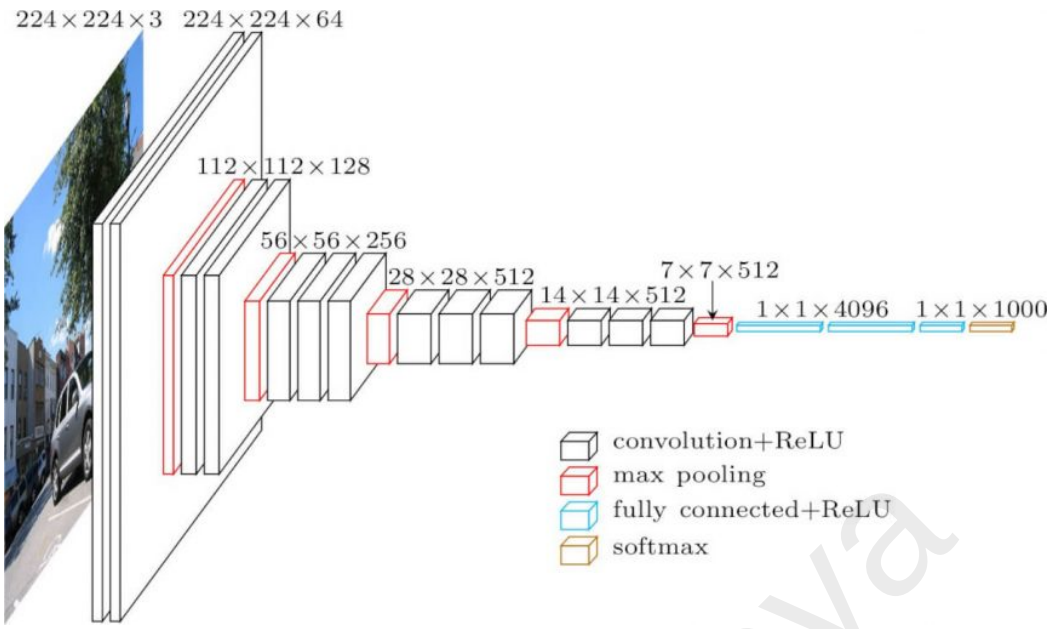


Figure 3.15: VGG-16 model architecture.

VGG-16 is a 16-layer VGG (Visual Geometry Group) model variant of 13 convolutional layers and 3 fully connected layers, as shown in **Figure 3.15**. VGG-16 model architecture improved compared with AlexNet model, which replaced large kernel-sized filters of 11 x 11 with stride 4 into multiple 3 x 3 kernel-sized filters with stride 1 uniformly. The concept of stacked convolutional layers of smaller receptive fields was a breakthrough innovation. VGG-16 is large network with around 138 million learnable parameters, but VGG-16 have a good network structure for learning since it is simple to apply. The main disadvantage is that due to large network, it takes more time to train the parameter and consume a lot of bandwidth or storage. VGG-16 architecture serves as foundation of ground-breaking object recognition models as the model won 1st place in object localization task to detect objects with 25.32% localization error. Currently, VGG-16 is the most popular option model in the field as image features extraction other than ResNet model. The model was feed RGB input image dimensions of 224 x 224 pixels.

3.6.4 Transfer Learning and Fine Tuning

Transfer learning refers to method that uses a previously trained (known as pre-trained) model as a starting point to learn new data. Fine-tuning a transfer learning usually the effective method to train a network with transfer learned feature compared to train a network from scratch with randomly initialized weights in smaller number of training dataset.

A pre-trained model is a saved neural network that was previously trained on huge amount database for image classification task such as ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). **Figure 3.16** shows the result of ILSVRC from 2010 to 2017 and top-5 error from each model.

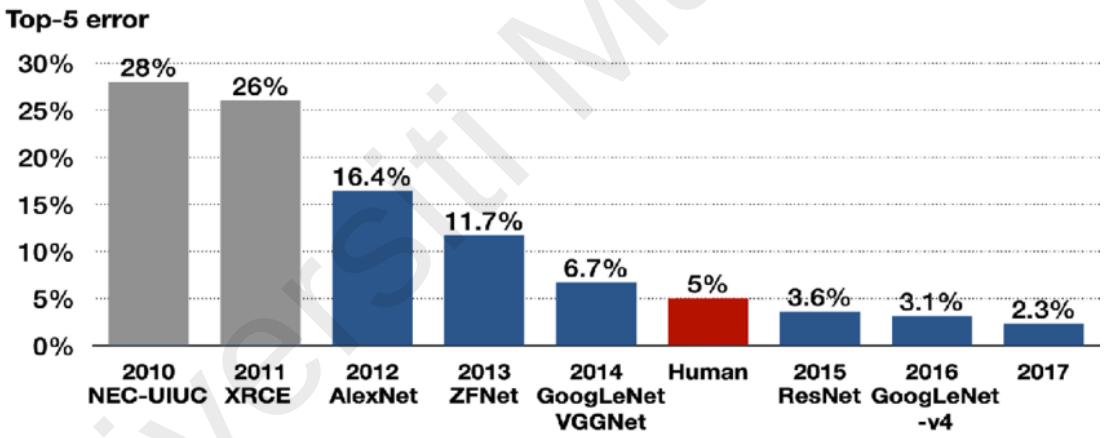


Figure 3.16: Top-5 error model winner of ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) from 2010 to 2017 where blue bar graph is model that use convolutional neural network (CNN) (Kang, Duong, & Park, 2020)

AlexNet, GoogLeNet and VGG-16 was employed as pre-trained model in this study to train the neural network for image classification task. These pre-trained data was trained using ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 as shown in **Figure 3.10**. ILSVRC 2012 contain 1.2 million training images, 50,000 validation images and 100,000 test images, organized into 1000 categories. The result of

pre-trained model using ILSVRC 2012 as shown in **Table 3.5**. Top-1 accuracy is the conventional accuracy, which requires that the model's highest probability answer perfectly match the predicted results. Meanwhile top-5 accuracy refers to the model's top 5 highest probability answer match with the predicted answer.

Table 3.5: Result top-1 and top-5 accuracy of pre-trained model using ILSVRC 2012 in NVIDIA DIGITS

Model	Top-1 accuracy	Top-5 accuracy
AlexNet	58.5%	81.3%
GoogLeNet	72.1%	91.0%
VGG-16	75.0%	93.4%

The method of transfer learning is to freeze all feature extraction stage (convolutional and max-pooling layers) that have ImageNet weights from being changed, and adjust the last classification neuron on classifier stage (fully connected layer) to match with new classification. The network needs to retrain which allowing the fully connected layer's weights to be changed. The advantage of transfer learning is to be able to take advantage of the previous network's feature extraction stage and just tune the final classifier to operate better with new dataset. Meanwhile, the method of fine-tuning is to be able to adjust and retrain both feature extraction stage and classifier stage.

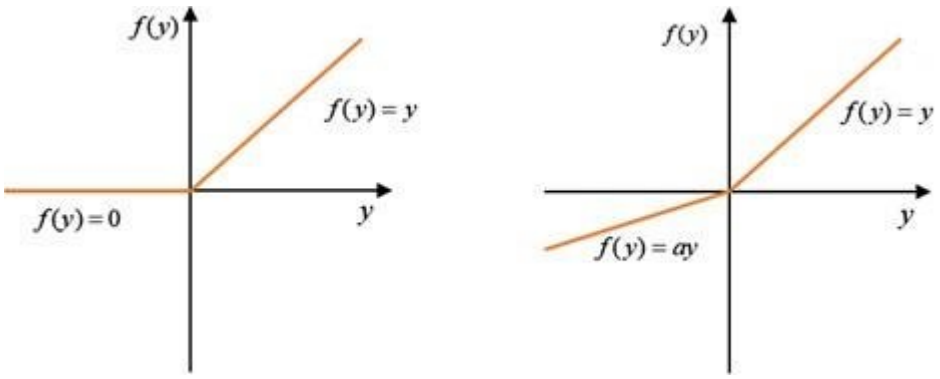


Figure 3.17: Activation function of ReLU (left) and PReLU (right) (Almaadeed, Elharrouss, Al-ma'adeed, Bouridane, & Beghdadi, 2019)

One of the methods to fine-tune the model is by changing the activation function parameter. Rectified linear unit (ReLU) is often used as activation function in DCNN, which defined as $y = \max(0, y)$ mathematically. ReLU identify only positive value and 0 for negative value. The benefit of ReLU is less time to train or run the model, not suffered from vanishing gradient problem (such as tanh and sigmoid) and faster to compute due to partially active network. The disadvantage of ReLU is a problem known as "dying ReLU" for being 0 for negative value. The negative neuron has no function in discriminating the data and is essentially worthless in the network. "Dying ReLU" prone to happen when learning rate is high or there is large negative bias. The alternative to prevent the problem is to have slight slope in negative value such as PReLU (Parametric ReLU) activation function. PReLU make the slope parameter to figure out itself for the neural network as $y = ay\ when\ y < 0$ mathematically shown in **Figure 3.17**. In this study, experiment will be done to evaluate the difference of using ReLU and PReLU during training process.

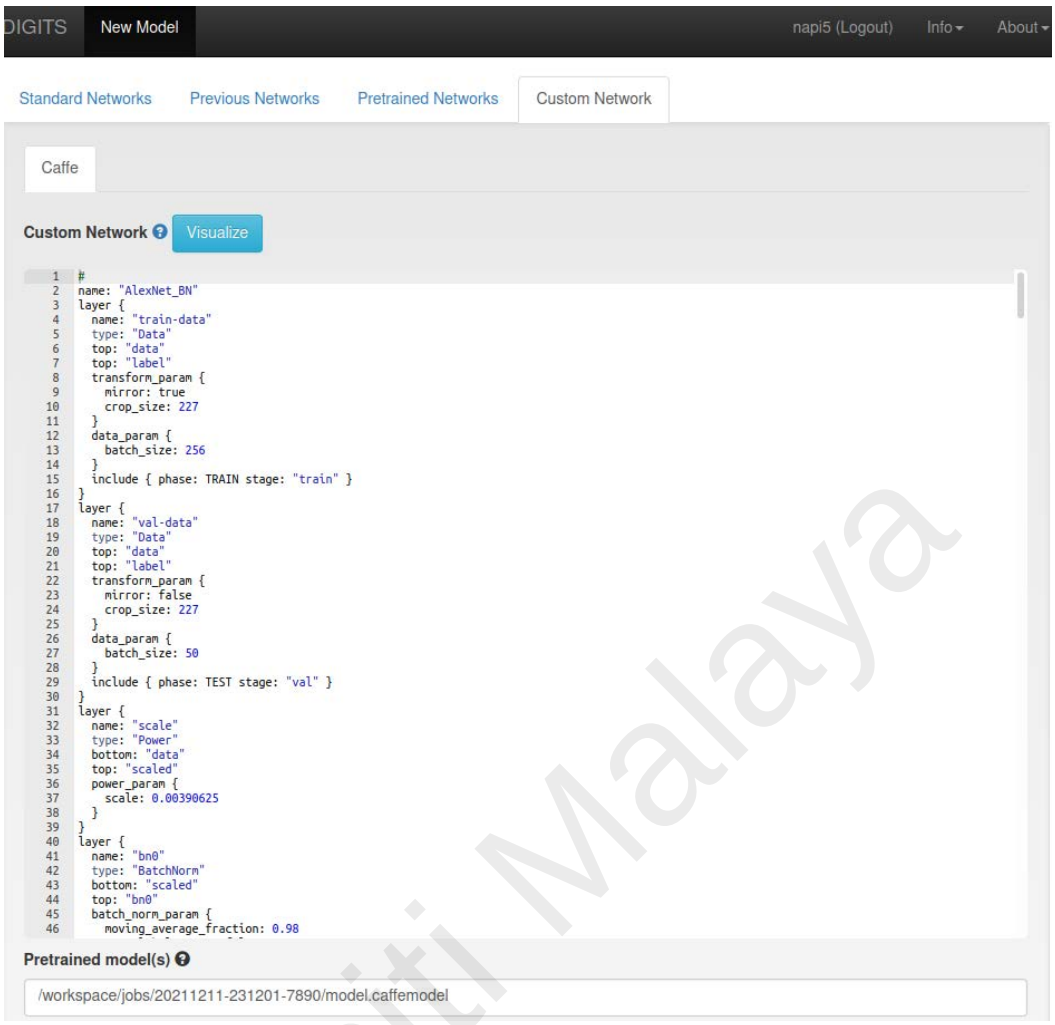


Figure 3.18: Interface to customize the model in NVIDIA DIGITS using Caffe framework

In NVIDIA DIGITS, the pre-trained model can be customized by selecting "Customize" at "Pretrained Networks" tab. It will direct it into "Custom Network" as shown in **Figure 3.18**. The protocol buffer definition file (prototxt) can adjust the layers and their parameter by using caffe framework library.

Pre-trained AlexNet, GoogLeNet and VGG-16 used the same base architecture which retaining the pre-trained layers but replaced the last fully connected layer from 1000 classes into 2 classes which corresponds to the classification of benign and malignant as shown in **Figure 3.19**. The layer's name needs to change because of the changes of parameter.

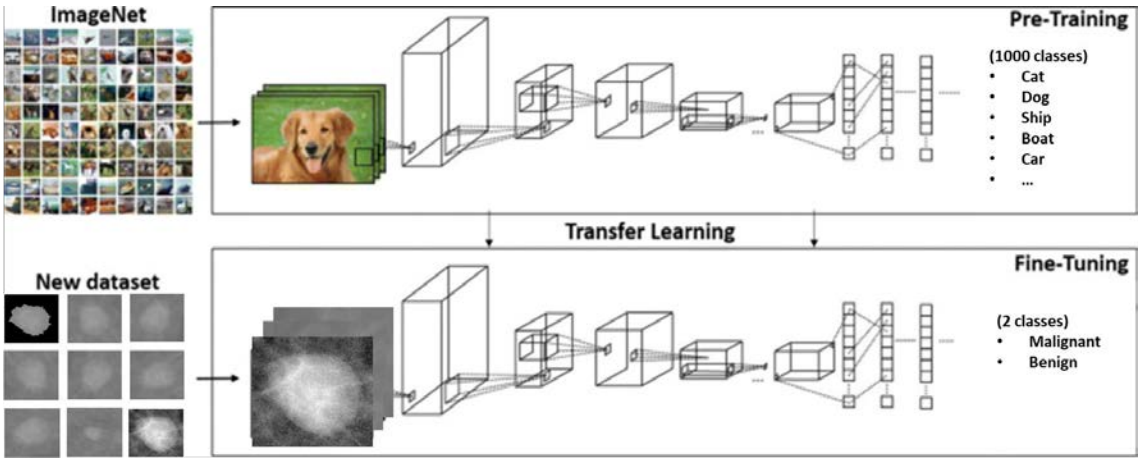


Figure 3.19: Transfer learning and fine tuning

Batch size is a hyperparameter that specifies the total amount training samples utilised in single batch before changing the existing internal model parameters. Higher the batch size, the higher memory needed for overall training procedure. As a result, the batch size was determined by the maximum data size could fit into 6 GB. If there was an error indicate not enough memory to process the training, it can be solved by reducing the images size or use smaller batch size.

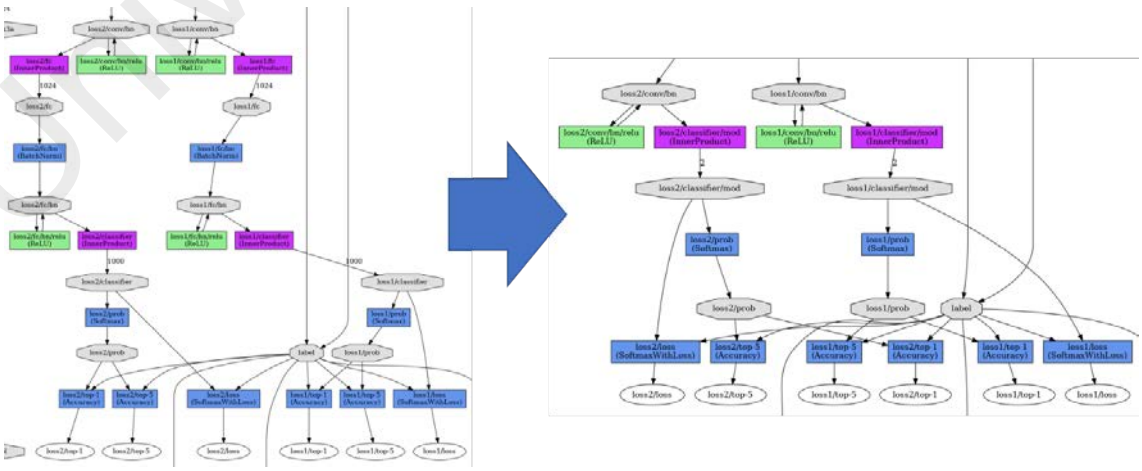


Figure 3.20: Removing auxiliary classifier from the pre-trained GoogLeNet model

Auxiliary classifier in pre-trained GoogLeNet needs to be removed since it impaired the overall training process. The method is by removing two 1024 fully connected layer which is layer loss1/fc and loss2/fc, together with BatchNorm and ReLU that followed as shown in **Figure 3.20**. The summary of customization of the model as shown in **Table 3.6**.

Table 3.6: Customization of pre-trained AlexNet, GoogLeNet and VGG-16

Model	Layer	Activation Function	Batch Size	Other
Pre-trained AlexNet	Last fully connected layer: change from 1000 classes into 2 classes	ReLU	Train: 256 Validation: 50	-
		PReLU	Train: 128 Validation: 64	-
Pre-trained GoogLeNet	3 nos. last fully connected layer: change from 1000 classes into 2 classes	ReLU & PReLU	Train: 32 Validation: 16	Remove auxiliary classifier; two 1024 fully connected layer which is layer loss1/fc and loss2/fc.
Pre-trained VGG-16	Last fully connected layer: change from 1000 classes into 2 classes	ReLU	Train: 16 Validation: 16	-
		PReLU	Train: 8 Validation: 8	-

3.7 Training the DCNN Model

3.7.1 Importing the dataset

To train the DCNN model, the first thing to do is to import the training dataset. The directory of dataset already been set up and describe in **Section 3.5**. At NVIDIA DIGITS home page, the training dataset can be imported by click the tab “Datasets” and create the “New Dataset Images” for “Classification” as shown in **Figure 3.21**.

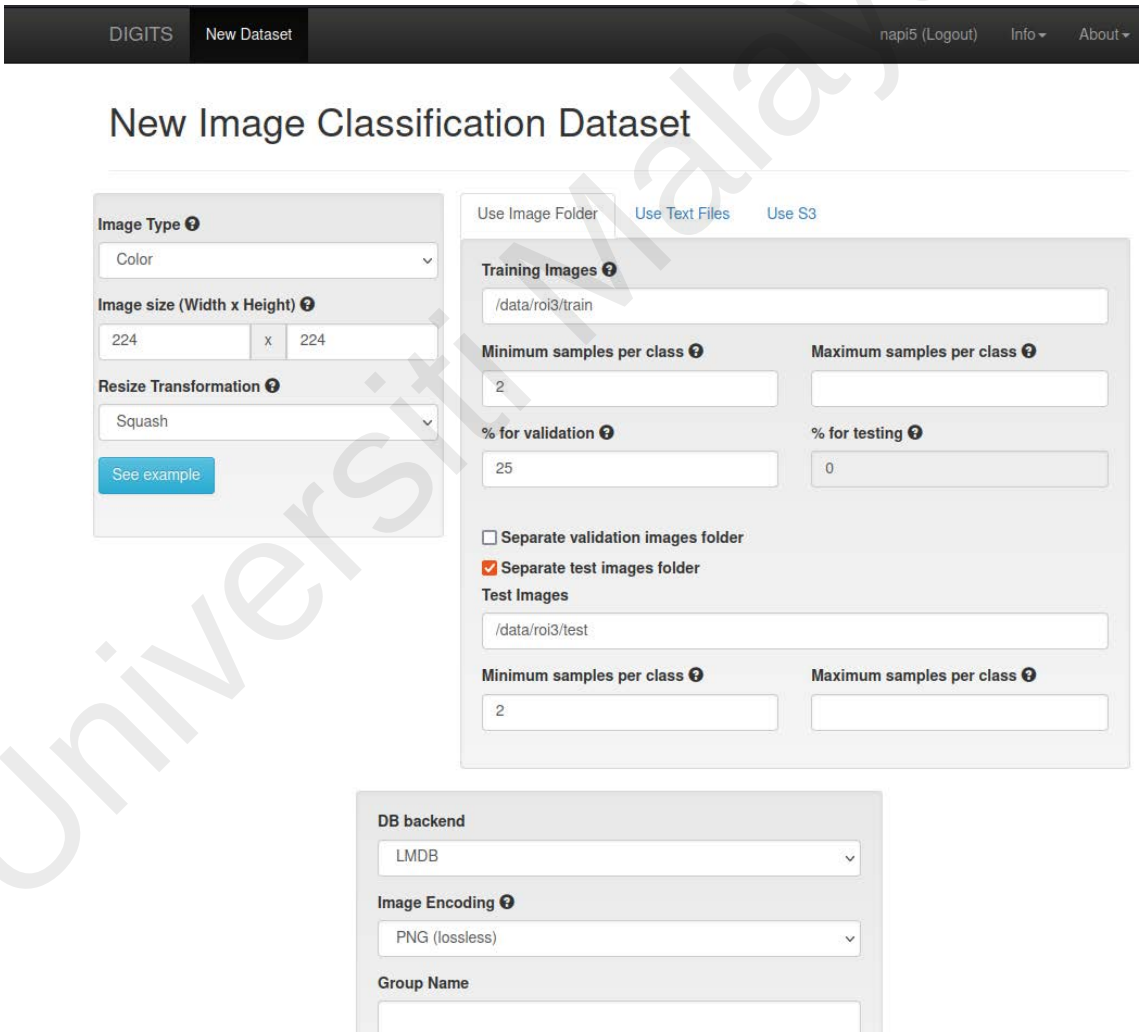


Figure 3.21: Interface to import training dataset in NVIDIA DIGITS

The image type needs to be color or RGB (3 channels) even though the training datasets is in grayscale (1 channel). This is due to the pre-trained model is already trained in RGB setting. Since the image size is 256 x 256 pixels, the images need to be resized by using “Squash” resize transformation and input image are fixed to 224 x 224 pixels as GoogLeNet and VGG-16 only accept that image sizes as input.

In training datasets, there are 2 folder which is train folder (80%) and test folder (20%) dataset. In order to configure training set and validation set in the train folder, 25% have been set for validation, which mean that from 80%, 60% as training set and 20% as validation set. Another alternative is tick “separate validation image folder” if have another folder for validation. As for test set, just tick “separate test images folder” for test folder. After configuration, NVIDIA DIGITS process the datasets and provide overview of the datasets including bar chart and the list of the images that import in NVIDIA DIGITS. The summary of configuration for importing dataset in NVIDIA DIGITS as **Table 3.7**.

As for test datasets, the folder that consists of 540 images are separated from the training datasets folder. Inside test folder already separate 2 classification images of benign (270 images) and malignant (270 images).

Table 3.7: Configuration for import training dataset in NVIDIA DIGITS

Parameter	Value
Image Configuration	
Image type	Color
Image size (width x height)	224 x 224
Resize transformation	Squash
Training and Validation Set Configuration	
Training images	/data/roi3/train <file path to training & validation set>
Minimum samples per class	2
% for validation	25
Test Set Configuration	
Test images	/data/roi3/test <file path to test set>
Minimum samples per class	2

3.7.2 Configure DCNN Training

To do this, click “Models” tab and create the “New Dataset Images” for “Classification”. This would display training parameter to train the network include the selection of dataset, number of epochs, type of solver, base learning rate, DCNN selection and etc. as shown in **Figure 3.22**.

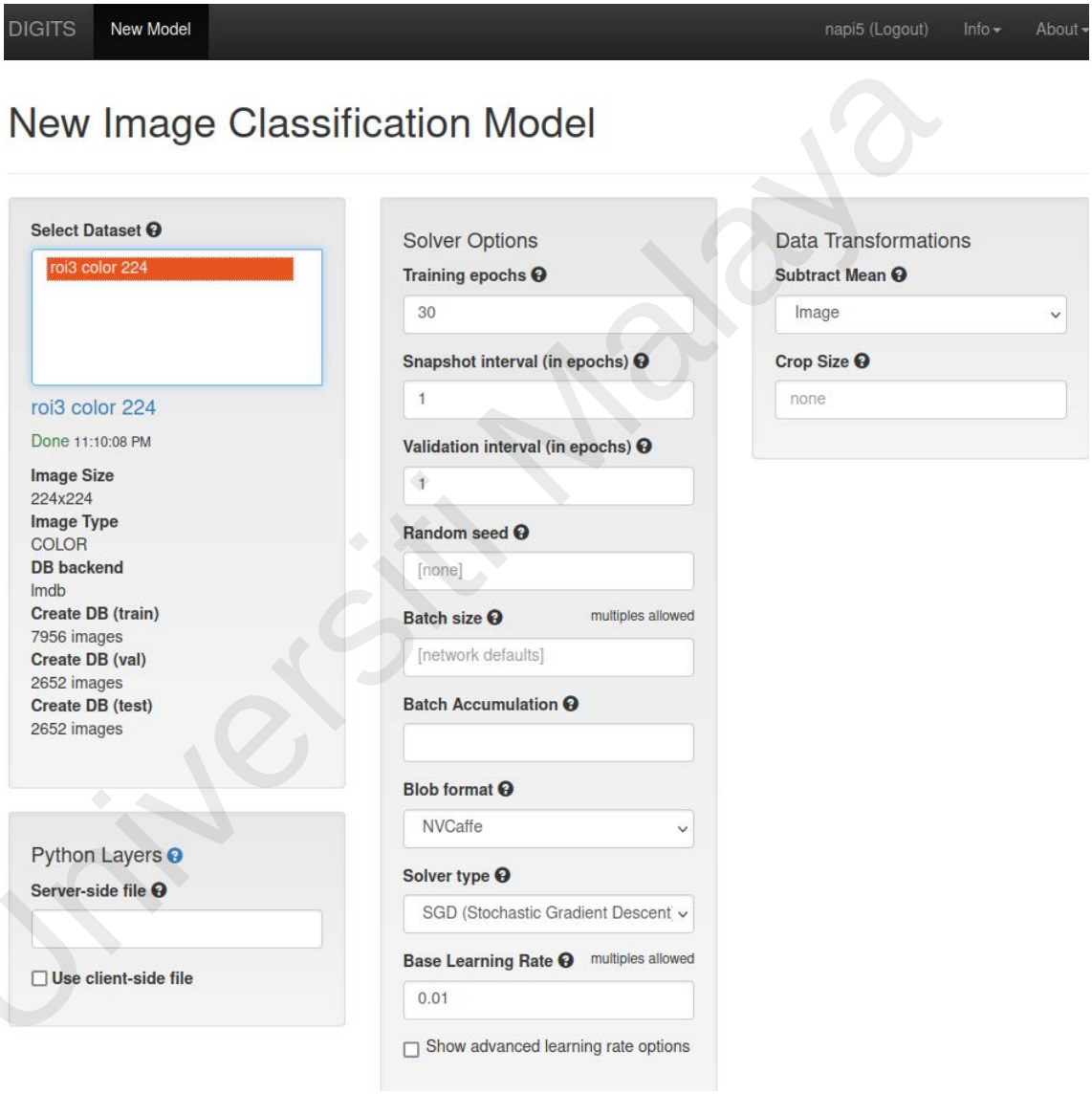


Figure 3.22: Interface for image classification training parameter in NVIDIA DIGITS

In this study, the training process conducted for 30 epochs. As for optimizer or solver, the model used SGD and Adam. SGD (Stochastic Gradient Descent) is most well-known

approach of optimizing neural networks, meanwhile Adam (Adaptive Moment Estimation) as the most favorable gradient descent optimization algorithms. So, the model utilised SGD and Adam as optimizer that able to give best result throughout training process. The base learning rate shall be set as 0.01, 0.001, 0.0001, 0.00001 throughout the study, and using a step-down learning policy. The summary of configuration for training parameter as **Table 3.8**.

Table 3.8: Configuration for image classification training parameter in NVIDIA DIGITS

Parameter	Value
Solver Options Configuration	
Training epochs	30
Solver type	Stochastic Gradient Descent (SGD) / Adaptive Moment Estimation (Adam)
Learning rate	0.01 / 0.001 / 0.0001 / 0.00001

In this study, DCNN model that are already fine-tune which is pre-trained AlexNet. GoogLeNet and VGG-16 are employed for mass classification. The network was trained on single NVIDIA graphic card (6 GB). During training phase, 13,260 breast mass patch were utilised to the model with 3978 benign and 3978 malignant in training set, 1326 benign and 1326 malignant in validation set, and 1326 benign and 1326 malignant in test set. Furthermore, the trained model will be assessing the classification performance using the test dataset of 540 ROI cropped mass images consists of 270 benign and 270 malignant.

3.8 Summary

The dataset was publicly available database called Curated Breast Imaging Subset of Digital Database for Screening Mammography, which is grayscale mammographic images in DICOM format and csv files that provide essential information about the database. By using Exploratory Data Analysis, the database has a total of 3568 whole breast mammograms images in MLO and CC views, which 1872 images is calcification and 1696 images is mass. The study focuses on breast mass patch from mass images which contain 912 benign and 784 malignant. The breast mass patch was cropped into the size of 256 x 256 pixels, and data augmentation of rotating, flipping and zoom has been carried out by using Adobe Photoshop. Total breast mass patch images after data augmentation for training dataset is 13,260 images which divided into 7956 training set images (60%), 2652 validation set images (20%) and 2652 test set images (20%). The test dataset provides the final trained model's performance by employing the remaining ROI cropped images which separate images from training datasets, have the total of 540 images with 270 images of benign and 270 images of malignant.

NVIDIA DIGITS version 6.1.1 using Caffe framework was used to trained, validated and tested the DCNN, which runs on NVIDIA GeForce GTX1060 6GB GPU. The training procedure for classification of benign and malignant from breast mass patch were performed by using pre-trained DCNN architecture of AlexNet, GoogLeNet and VGG-16. The pre-trained DCNN model was previously trained using ImageNet Large-Scale Visual Recognition Challenge 2012. The pre-trained DCNN model used transfer learning with fine-tuning method, which includes changing the activation function (ReLU and PReLU), replaced the last fully connected layer from 1000 classes into binary classes, configure the batch size and removing auxiliary classifier (for GoogLeNet model).

Before training process, the training dataset needs to be imported in the NVIDIA DIGITS. Parameters needs to be configured to run on pre-trained DCNN model such as RGB image type, image resized into 224 x 224 pixels and configuration of training dataset. DCNN training parameter also needs to be configured such as 30 epochs training process, optimizer of Stochastic Gradient Descent (SGD) or Adaptive Moment Estimation (Adam), range of learning base from 0.01 to 0.00001.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the findings from the methodology outlined in Chapter 3, as well as the performance analysis of deep convolutional neural network (DCNN) model. Section 4.2 compares the results of the model's accuracy and loss parameters in the DCNN model. The classification findings in Section 4.3 are based on the highest accuracy and low loss from each DCNN model, for performance analysis, which include accuracy, sensitivity, precision, specificity, F1-score, and area under the curve (AUC). The findings will be discussed in Section 4.4. Section 4.5 contains a summary of this chapter.

4.2 Training Model Results

The training process can begin after configuration is completed as discussed in Chapter 3. During the training process, as shown in **Figure 4.1**, NVIDIA DIGITS visualize network's performance by plotting the loss achieved on training and validation sets, as well as the accuracy gained on validation set throughout training epochs, to shows the effectiveness of network in learning the data. NVIDIA DIGITS have the ability to monitor network performance in real-time, allowing to control entire training process includes epoch's performance and detect faults during the process. The learning rate with step-down learning policy shows the learning rate decreases with time throughout training process.

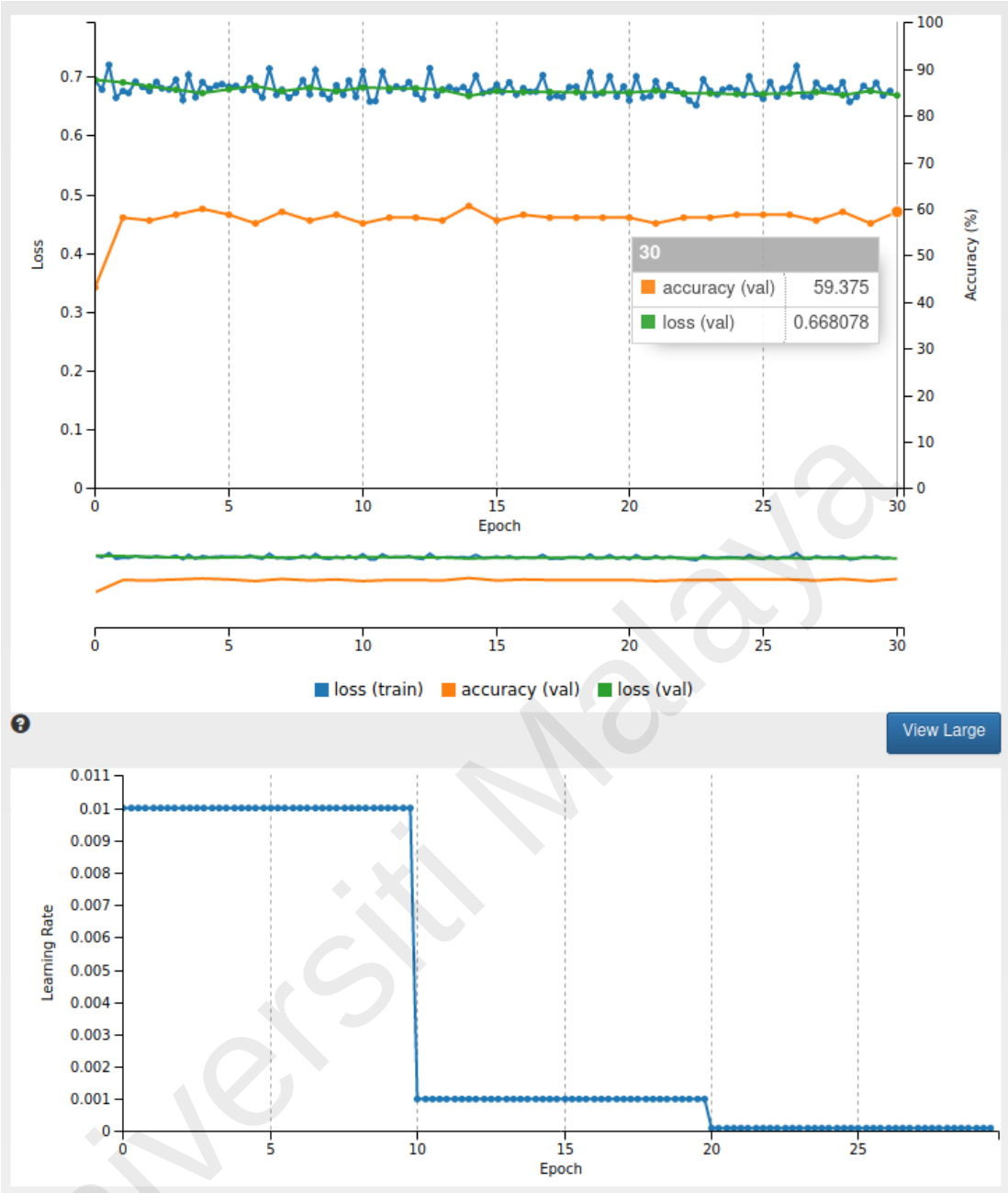


Figure 4.1: Example plot of loss, accuracy and learning rate with step-down learning policy.

The training model is assessed based on the outcomes, which include accuracy and loss. Accuracy is a mechanism to evaluate the classification model's performance in terms of percentage. Accuracy defined as number of prediction where the predicted value is equal to actual value. Meanwhile, the loss function considers the probability of how prediction value differs from the actual value, which provides a detailed view of the

model's efficiency. The loss is stated as the total number of errors committed in training and validation set, for each sample. Loss is widely used in the training process to achieved the best parameter values for the model, such as changing weights in neural networks, to lower the magnitude of the loss. The accuracy and loss often appear inversely proportional, but there is no mathematical relationship between them.

In this section, the accuracy and loss of the DCNN model is evaluated based on the NVIDIA DIGITS training process results with the parameter discussed in Chapter 3. The training model of each DCNN model is selected based on the results of high accuracy and low loss using validation set.

4.2.1 AlexNet Model

Based on **Table 4.1**, training model A1 shows the best result with 83.26% validation accuracy and 0.45 validation loss. The network performance plot of training model A1 network with 30 epochs as shown in **Figure 4.2**. The results are based on top-1 (val) for accuracy and loss (val) for loss at the plot.

Table 4.1: AlexNet Model Training Results

Training Model	Solver Type	Base Learning Rate	Activation Function	Accuracy	Loss	Time (minutes)
A1	**SGD**	**0.01**	**ReLU**	**83.26%**	**0.45**	**10**
A2	SGD	0.001	ReLU	77.41%	0.48	10
A3	SGD	0.0001	ReLU	67.96%	0.59	12
A4	SGD	0.00001	ReLU	58.19%	0.67	11
A5	Adam	0.01	ReLU	49.78%	0.69	11
A6	Adam	0.001	ReLU	80.89%	0.56	11
A7	Adam	0.0001	ReLU	82.78%	0.6	11
A8	Adam	0.00001	ReLU	75.33%	0.5	12
A9	SGD	0.01	PReLU	74.96%	0.5	12
A10	SGD	0.001	PReLU	78.39%	0.47	13
A11	SGD	0.0001	PReLU	70.01%	0.58	12
A12	SGD	0.00001	PReLU	62.57%	0.65	14
A13	Adam	0.01	PReLU	60.6%	0.68	11
A14	Adam	0.001	PReLU	75.19%	0.5	14
A15	Adam	0.0001	PReLU	81.03%	0.48	11
A16	Adam	0.00001	PReLU	76.04%	0.51	11

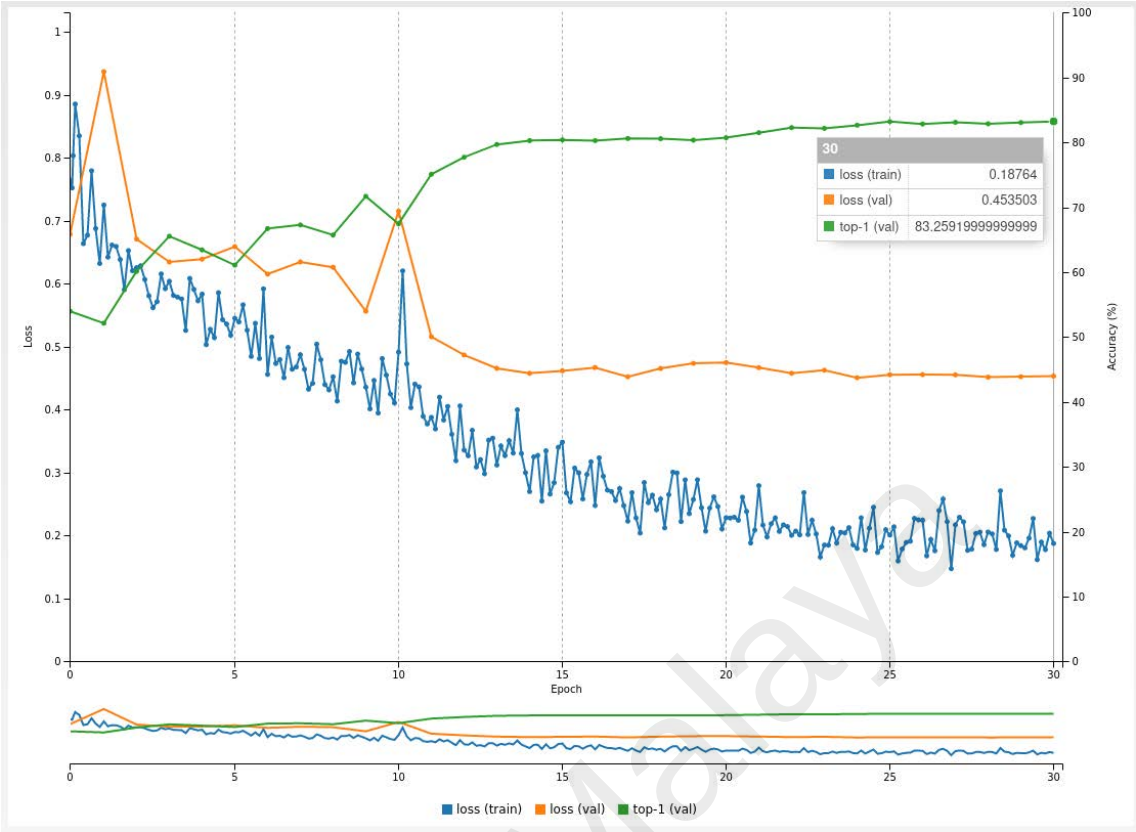


Figure 4.2: AlexNet Training Model A1 Network Performance

4.2.2 GoogLeNet Model

Based on **Table 4.2**, training model G7 shows the best result with 89.12% validation accuracy and 0.43 validation loss. The performance of training model G7 network with 30 epochs as shown in **Figure 4.3**. The results are based on loss3/top-1 (val) for accuracy and loss3/loss3 (val) for loss at the plot.

Table 4.2: GoogLeNet Model Training Results

Training Model	Solver Type	Base Learning Rate	Activation Function	Accuracy	Loss	Time (minutes)
G1	SGD	0.01	ReLU	75.67%	0.49	189
G2	SGD	0.001	ReLU	88.25%	0.41	193
G3	SGD	0.0001	ReLU	77.52%	0.51	167
G4	SGD	0.00001	ReLU	65.25%	0.63	56
G5	Adam	0.01	ReLU	65.96%	0.63	48
G6	Adam	0.001	ReLU	85.39%	0.75	52
G7	**Adam**	**0.0001**	**ReLU**	**89.12%**	**0.43**	**55**
G8	Adam	0.00001	ReLU	76.81%	0.64	56
G9	SGD	0.01	PReLU	84.56%	0.4	66
G10	SGD	0.001	PReLU	88.55%	0.36	67
G11	SGD	0.0001	PReLU	71.91%	0.57	69
G12	SGD	0.00001	PReLU	56.85%	0.68	67
G13	Adam	0.01	PReLU	66.68%	0.61	66
G14	Adam	0.001	PReLU	80.27%	0.45	66
G15	Adam	0.0001	PReLU	87.42%	0.6	66
G16	Adam	0.00001	PReLU	84.38%	0.39	65

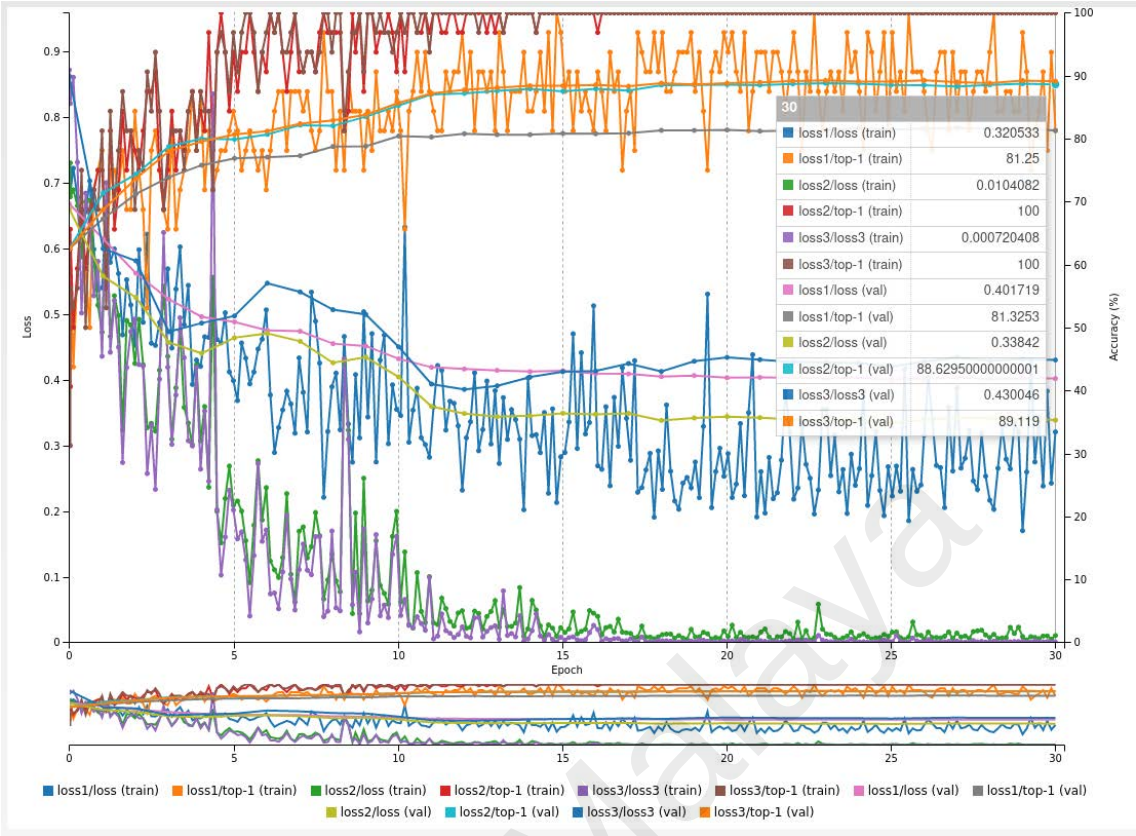


Figure 4.3: GoogLeNet Training Model G7 Network Performance

4.2.3 VGG-16 Model

Based on **Table 4.3**, training model V2 shows the best result with 86.22% validation accuracy and 0.64 validation loss. Even though training model V9, V10 and V11 show 100% of accuracy, the loss is high with 87.34, where the networks make big errors in some of the data. Some training models are aborted due to underfitting, since the plot shows that the accuracy is not improving and the loss is not decreasing. The performance of training model V2 network with 30 epochs as shown in **Figure 4.4**. The results are based on accuracy_top_1 (val) for accuracy and loss (val) for loss at the plot.

Table 4.3: VGG-16 Model Training Results

Training Model	Solver Type	Base Learning Rate	Activation Function	Accuracy	Loss	Time (minutes)
V1	SGD	0.01	ReLU	50%	0.69	Abort (underfit)
V2	**SGD**	**0.001**	**ReLU**	**86.22%**	**0.64**	**147**
V3	SGD	0.0001	ReLU	82.27%	0.44	153
V4	SGD	0.00001	ReLU	74.47%	0.53	148
V5	Adam	0.01	ReLU	50%	0.69	Abort (underfit)
V6	Adam	0.001	ReLU	49.96%	0.69	Abort (underfit)
V7	Adam	0.0001	ReLU	79.63%	0.63	183
V8	Adam	0.00001	ReLU	84.98%	0.78	211
V9	SGD	0.01	PReLU	100%	87.34	Abort (underfit)
V10	SGD	0.001	PReLU	100%	87.34	Abort (underfit)
V11	SGD	0.0001	PReLU	100%	87.34	Abort (underfit)
V12	SGD	0.00001	PReLU	69%	0.59	Abort (underfit)
V13	Adam	0.01	PReLU	54.03%	40.15	Abort (underfit)
V14	Adam	0.001	PReLU	49.62%	44	Abort (underfit)
V15	Adam	0.0001	PReLU	80.99%	1.37	195
V16	Adam	0.00001	PReLU	80.46%	0.49	203

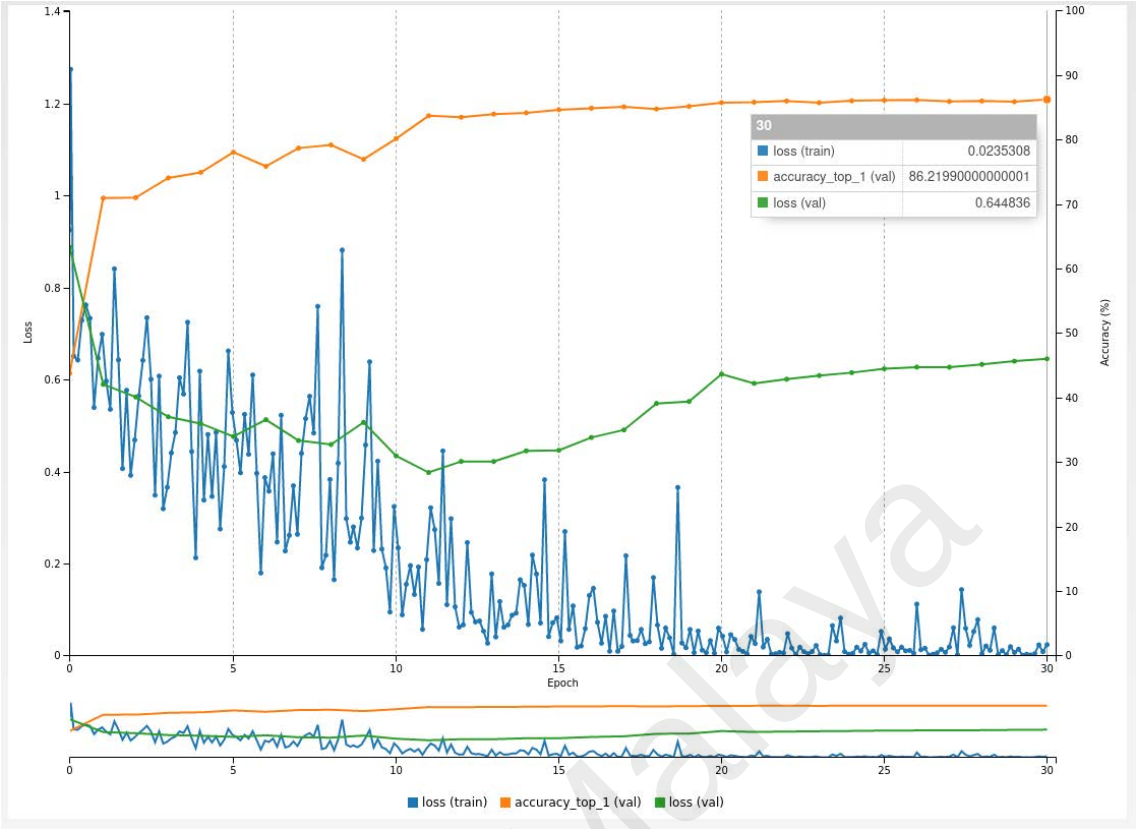


Figure 4.4: VGG-16 Training Model V2 Network Performance

4.2.4 Summary of Training Model Results

Table 4.4 shows the summary of training model results with high accuracy and low loss for each DCNN model.

Table 4.4: Summary of Training Model Results

Training Model	Solver Type	Base Learning Rate	Activation Function	Accuracy	Loss	Time (minutes)
AlexNet A1	SGD	0.01	ReLU	83.26%	0.45	10
GoogLeNet G7	Adam	0.0001	ReLU	89.12%	0.43	55
VGG-16 V2	SGD	0.001	ReLU	86.22%	0.64	147

4.2.5 Comparison of Training Model Results Using Pre-trained and Original Model

This section compares the performance of training model between using the pre-trained model and original model. The original model was train from scratch using the training dataset of 13,260 images only. The parameter was based on the best results obtained from pre-trained training model results. As for model, NVIDIA DIGITS only provide original model of AlexNet and GoogLeNet. **Table 4.5** shows the comparison of training model results using pre-trained model and original model. From the results, pre-trained model shows the increased of accuracy by 20% compared to original model.

Table 4.5: Comparison Training Model Results Using Pre-trained Model and Original Model

Model	Solver Type	Base Learning Rate	Activation Function	Pre-trained Model		Original Model	
				Accuracy	Loss	Accuracy	Loss
AlexNet	SGD	0.01	ReLU	**83.26%**	**0.45**	63.48%	0.63
GoogLeNet	Adam	0.0001	ReLU	**89.12%**	**0.43**	69.47%	0.59
VGG-16	SGD	0.001	ReLU	86.22%	0.64	n/a	n/a

4.3 Classification Results and Analysis

NVIDIA DIGITS provide the classification scores per sample in test dataset. In "Classify One Image" option, the classification score can give predictions and provide informative data as shown in **Figure 4.5**.

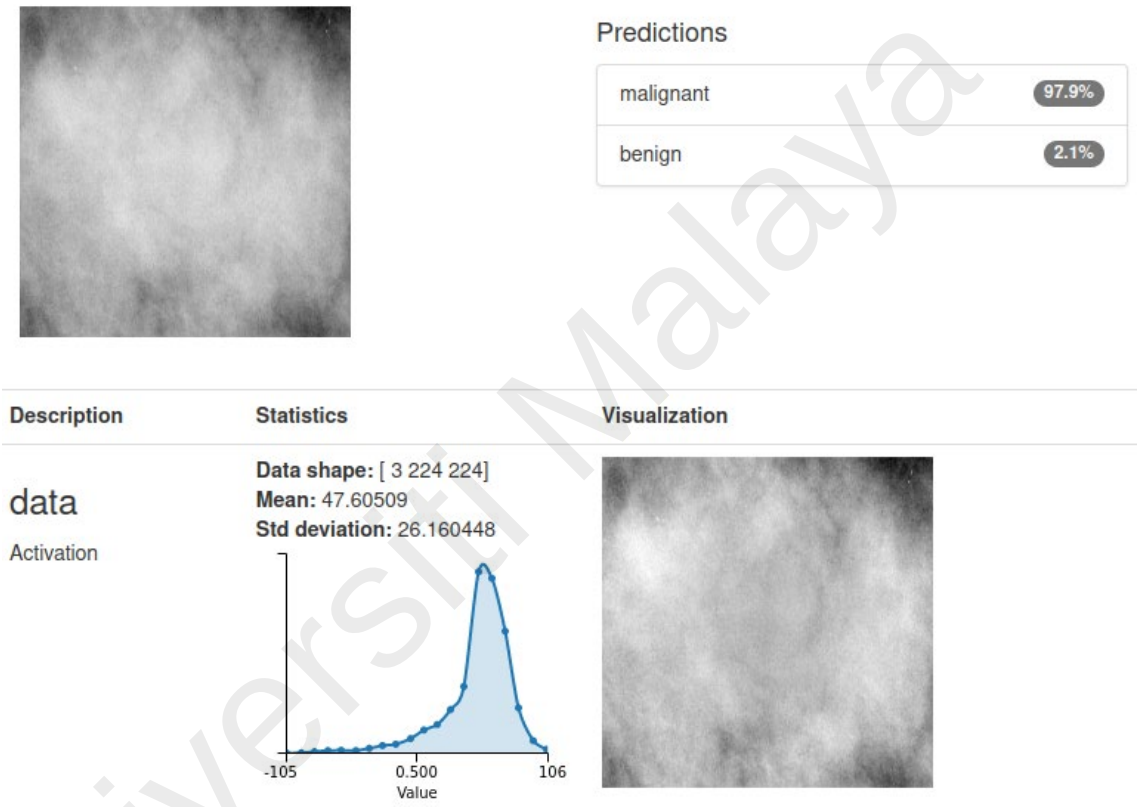


Figure 4.5: Classification result of "Classify One Image" using training model AlexNet A1

However, for further research in this study, NVIDIA DIGITS provide the classification scores samples in group called “Classify Many” options that allows text file list of test dataset path located on the host machine. The test.txt file was created by using NVIDIA DIGITS and obtained the classification scores for 540 mammographic image of breast mass patch samples in the list, classifying of benign or malignant in percentage to produce confusion matrix as shown in **Figure 4.6**.

Confusion matrix

	benign	malignant	Per-class accuracy
benign	193	77	71.48%
malignant	64	206	76.3%

All classifications

	Path	Ground truth	Top predictions			
1	/data/roi3 test/benign/P1439 LCC B AUG12.jpg	benign	**benign**	97.89%	**malignant**	2.11%
2	/data/roi3 test/benign/P1819 RCC B AUG12.jpg	benign	**benign**	93.01%	**malignant**	6.99%
3	/data/roi3 test/benign/P1685 RCC B AUG12.jpg	benign	**benign**	99.75%	**malignant**	0.25%
4	/data/roi3 test/benign/P1112 LMLO B AUG12.jpg	benign	**malignant**	100.0%	**benign**	0.0%
5	/data/roi3 test/benign/P1663 RMLO B AUG12.jpg	benign	**benign**	100.0%	**malignant**	0.0%
6	/data/roi3 test/benign/P1887 LMLO B AUG12.jpg	benign	**benign**	97.77%	**malignant**	2.23%
7	/data/roi3 test/benign/P1638 RCC B AUG12.jpg	benign	**benign**	100.0%	**malignant**	0.0%
8	/data/roi3 test/benign/P1152 RMLO1 B AUG12.jpg	benign	**benign**	99.95%	**malignant**	0.05%
9	/data/roi3 test/benign/P1826 RMLO B AUG12.jpg	benign	**benign**	57.85%	**malignant**	42.15%
10	/data/roi3 test/benign/P1608 RCC B AUG12.jpg	benign	**benign**	100.0%	**malignant**	0.0%

Figure 4.6: Classification result of “Classify Many” using training model GoogLeNet G7

The output of the classification score can be saved as HTML file and then converted into MATLAB could read it, to produce more detailed confusion matrix and receiver operating characteristic (ROC) curve.

Table 4.6: Breast Mass Patch Classification Confusion Matrix

Confusion Matrix		**Prediction**	
		Malignant	**Benign**
Actual	**Malignant**	True Positive, TP	False Negative, FN
	Benign	False Positive, FP	True Negative, TN

The confusion matrix as shown in **Table 4.6**, used to assess the binary classification performance tests, and consists of 2 x 2 table that contains four outcomes produced by classification, where True Positive (TP) represents the number of malignant masses correctly classified, False Positive (FP) represents the number of benign masses classified as malignant masses, True Negative (TN) represents the number of benign masses correctly classified and False Negative (FN) represents the number of malignant masses classified as benign masses. The confusion matrix to utilize the outcome of model prediction using test dataset.

Table 4.7, **Table 4.8** and **Table 4.9** describe confusion matrix of breast masses patch classification of AlexNet, GoogLeNet and VGG-16 model. Positive and negative prediction results in the table correspond to benign and malignant classification of these model using test dataset. According to the confusion matrix table, GoogLeNet model shows the highest accuracy of 73.9%, followed by VGG-16 (accuracy of 73.7%) and AlexNet (accuracy of 69.3%).

Table 4.7: Confusion matrix of AlexNet

AlexNet A1		Prediction		
		Malignant	Benign	
Actual	Malignant	181	89	67%
	Benign	77	193	71.5%
		70.2%	68.4%	69.3%

Table 4.8: Confusion matrix of GoogLeNet

GoogLeNet G7		Prediction		
		Malignant	Benign	
Actual	Malignant	206	64	76.3%
	Benign	77	193	71.5%
		72.8%	75.1%	73.9%

Table 4.9: Confusion matrix of VGG-16

VGG-16 V2		Prediction		
		Malignant	Benign	
Actual	Malignant	216	54	80%
	Benign	88	182	71.5%
		71.1%	77.1%	73.7%

A confusion matrix is essential to evaluate and analyze the training DCNN model performance, which includes metrics like F1-Score, specificity, precision, sensitivity/ recall and accuracy, which can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

$$Sensitivity\ or\ Recall\ = \frac{TP}{TP + FN} \tag{4.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4.4}$$

$$F1 - Score = \frac{2 * (Recall * Precision)}{Recall + Precision} = \frac{2TP}{2TP + FP + FN} \tag{4.5}$$

Accuracy is defined as the proportion of benign and malignant masses accurately classified to the total number of masses. Sensitivity or recall is defined as the proportion of malignant masses accurately classified to the total number of actual malignant masses. Precision is defined as the proportion of malignant masses correctly classified to the total number of predicted malignant masses. Specificity is defined as the proportion of benign masses correctly classified to the total number of actual benign masses. The F1-score is the harmonic mean (average) of recall and precision, considered as a performance indicator of classification.

Table 4.10: Results of Classification Performance

Training Model	Accuracy	Sensitivity/ Recall	Precision	Specificity	F1-Score
AlexNet A1	69.26%	67.04%	70.16%	**71.48%**	68.56%
GoogLeNet G7	**73.89%**	76.30%	**72.79%**	**71.48%**	74.50%
VGG-16 V2	73.70%	**80.00%**	71.05%	67.41%	**75.26%**

Table 4.10 compares the performance of the three DCNN training models based on measurements of accuracy, sensitivity/ recall, precision, specificity and F1-score. GoogLeNet model performed in accuracy rate of 73.89% and precision rate of 72.79%. However, sensitivity rate and F1-score of VGG-16 model was performed better than AlexNet and GoogLeNet model with 80% and 75.26% respectively. AlexNet and GoogLeNet model have the same specificity rate of 71.48%.

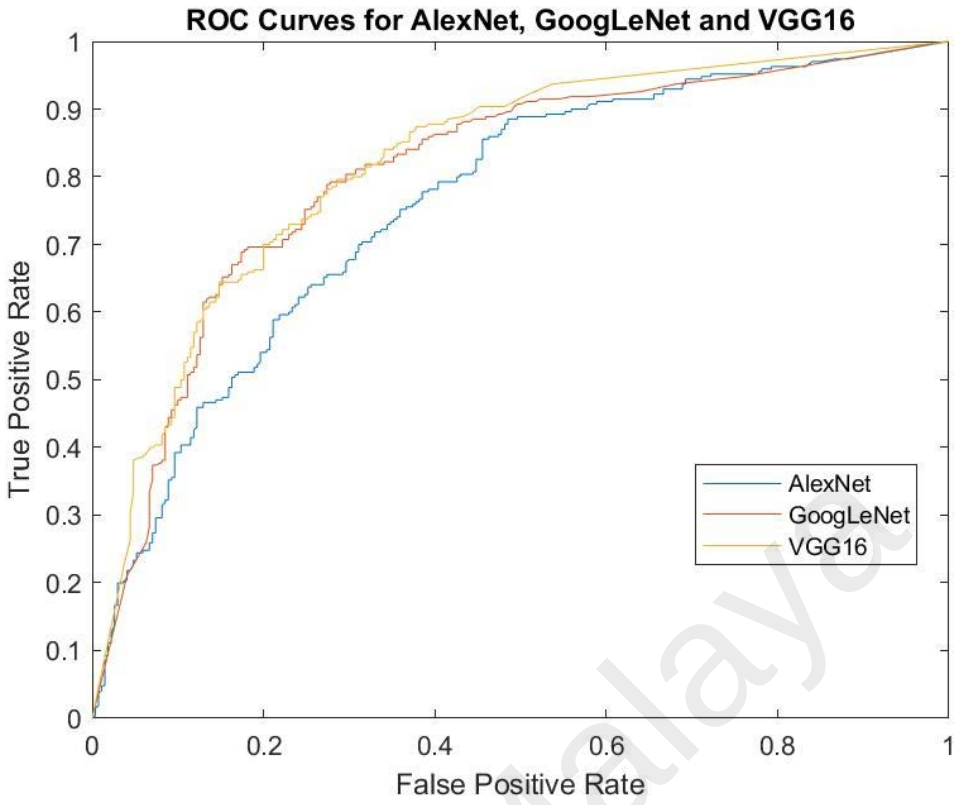


Figure 4.7: ROC Curves for breast mass patch of AlexNet, GoogLeNet and VGG-16

A receiver operating characteristic (ROC) curve is the second method to evaluate the performance tests for binary classification models. ROC curve are graphs of the true positive rate against the false positive rate for various classification test thresholds. Area Under the Curve (AUC) is a metric used to assess the effectiveness of binary classifier. The closer AUC value to 1, the better it is. **Figure 4.10** shows the graph of ROC curve and calculated AUC for verifying the binary classification's performance of breast masses patch based on SoftMax classifier of DCNN model. Among these DCNN models, VGG-16 performs the best (0.8207), trailed by GoogLeNet (0.8064) and AlexNet (0.7601). It shows that the overall ability of VGG-16 and GoogleNet superior over AlexNet in distinguishing benign or malignant breast mass patch.

4.4 Discussion

This study shows that fine-tune pre-trained DCNN model can efficiently classify the breast masses patch of benign and malignant from mammogram images. Based on the findings from the process, parameters such as solver (optimization), learning rate and activation function does affect the accuracy and loss rate of the network.

Deep learning neural networks are often trained using Stochastic Gradient Descent (SGD). SGD is an optimization technique that computes the error gradient for the current state of the model and uses samples from the training dataset, to update model's weights via backpropagation that scaled by the learning rate. Meanwhile, Adaptive Moment Estimation (Adam) is an extension of SGD that gained popularity in deep learning applications. It produces good results quickly and efficiently while using a large dataset and network. In **Table 4.4**, the optimal optimization for AlexNet is SGD, meanwhile, GoogLeNet is Adam. Even though the optimization for VGG-16 using SGD, Adam optimizer produce good results especially with a lower learning rate of 0.0001 and 0.00001 for VGG-16 model (refer **Table 4.3**).

Learning rate is a parameter that controls the weights in the neural network in relation to the loss gradient, meaning that learning rate control the speed rate of DCNN model to learn the problem. From the **Table 4.4**, the optimal learning rate for AlexNet is 0.01, GoogLeNet is 0.0001 and VGG-16 is 0.001. It is clear that if the model trained with either high or extremely low learning rate value, the model responded with low accuracy and high loss. A model with less layers such as AlexNet, preferred a higher learning rate, and model with more layers such as GoogLeNet and VGG-16 preferred lower learning rate. This is because a lower learning rate can overcome overfitting, a problem that usually occurs on a large network with many layers.

As for activation function, Xu, Wang, Chen, and Li (2015) reported that modified rectified linear unit (ReLU) such as randomized ReLU (RReLU), parametric ReLU (PReLU), and leaky ReLU, outperformed standard ReLU on image classification task. However, in this study, the most optimal activation function is ReLU compared to PReLU (refer **Table 4.4**). So, there is no significant difference using ReLU or PReLU because each activation function has its own characteristic and selection of suitable activation function determined by the network structure and the task that neural network is attempting to learn.

The amount of data used in the study is insufficient for training DCNN model. A small quantity of data cannot represent all sorts of masses and may results to overfitting. So, this study applied data augmentation and transfer learning method with fine-tune the pre-trained model of AlexNet, GoogLeNet and VGG-16 to overcome data limitation and improve the classification performance of breast masses patch in mammographic images as shown in **Table 4.5**. As a results, AUC for these models shows good result of around 0.8, with 0.8207 (VGG-16), 0.8064 (GoogLeNet) and 0.7601 (AlexNet). According to the findings of Xiaofei Zhang et al. (2017), utilizing a pre-trained model using large amount of data are more responsive in extracting features includes edges and shape compared with training the model from scratch, even with limited training dataset for medical-oriented application (Mehra, 2018; Tsochatzidis, Costaridou, & Pratikakis, 2019). Furthermore, data augmentation also provide variety to the limited dataset, improving discriminating ability and eliminating overfitting (Falconi, Perez, Aguilar, & Conci, 2020; Salama, Elbagoury, & Aly, 2020). This study concludes that training with transfer learning and data augmentation technique of DCNN model improves performance of breast masses patch classification, even with limited dataset.

Due to transfer learning and data augmentation technique, the results obtained from training process in **Figure 4.2**, **4.3** and **4.4**, the plots of AlexNet, GoogLeNet and VGG-16 can produce better results with few numbers of epochs such as 25 epochs instead of 30 epochs. The experiment done by Rahman et al. (2020) shows that the training process with 10 epochs is sufficient compared to train in a large number of epochs to produce significant results when using ResNet50 and Inception V3 to prevent the model from overfitting.

However, in **Figure 4.4**, the VGG-16 model tend to overfit the data after epoch 11 where the loss kept on increasing from 0.4 to 0.6 (epoch 30). Large capacity network such as VGG-16 tends to overfitting in small training dataset because of complexity due to the increased number of layers and non-linearities. One of the strategies to eliminate overfitting is to reduce effective capacity by freezing more layers of the networks. Another method is select and downloads training model with high accuracy and low loss, for example epoch 11 at VGG-16 model in NVIDIA DIGITS.

4.5 Summary

The results of pre-trained AlexNet, GoogLeNet and VGG-16 model to classify benign and malignant using breast masses patch mammographic images have been presented in this chapter. The results for training model are based on accuracy and loss from training process using NVIDIA DIGITS presented in Section 4.2. Training models with high accuracy and low loss are selected from each DCNN model of AlexNet (83.26% accuracy, 0.45 loss), GoogLeNet (89.12% accuracy, 0.43 loss) and VGG-16 (86.22% accuracy, 0.64 loss). In Section 4.3, the results of classification were obtained using test dataset to produce confusion matrix and receiver operating characteristic (ROC) curve. Confusion matrix contains four outcome (TP, FN, FP and TN) was analyzed, and the accuracy, sensitivity, precision, specificity and F1-score of the DCNN model were calculated. As an outcome, AlexNet and GoogLeNet have the same and highest specificity (71.48%); GoogLeNet have the highest accuracy (73.89%) and precision (72.79%); VGG-16 have the highest sensitivity (80%) and F1-score (75.26%). Meanwhile, area under the curve (AUC) obtained by ROC was used to evaluate the binary classification performance; as an outcome VGG-16 (0.8207) is the highest, trailed by GoogLeNet (0.8064) and AlexNet (0.7601).

CHAPTER 5: CONCLUSION

5.1 Conclusion

This study presented the breast's mass classification of benign or malignant, from mammogram images using deep convolutional neural networks (DCNN) architecture model. The purpose of this study is to evaluate and analyze the pre-trained DCNN model's performance consists of AlexNet, GoogLeNet and VGG-16, on the breast masses patch classification using the CBIS-DDSM database. This chapter also summarizes the method to optimize the classification performance and findings from the results obtained throughout the study. At the end of the chapter, a list of future works is also presented.

The dataset was obtained from public database called Curated Breast Imaging Subset of Digital Database for Screening Mammography. This dataset contains grayscale mammogram image in DICOM format consisting of 1696 mass images, which is a small number to be trained the DCNN model for classification task. The study demonstrated that the method of pre-processing data, data augmentation and transfer learning overcome the data size constraint and improved the performance of the classification, especially in acquiring the dataset of medical imaging.

The pre-processing data and data augmentation method have successfully increased the numbers of breast masses patch into 13,260 images training dataset and 540 images test dataset. Training dataset is consisting of 60%-20%-20%, training-validation-test data split which contains 7656 images training set, 2652 images validation set and 2652 images test set. Other than increasing the dataset, this method also enriches the apparent features and recognizes the similar shapes taken from different viewpoints.

On classification of benign or malignant breast mass, the evaluation results of pre-trained VGG-16, GoogLeNet and AlexNet model were discussed. DCNN models was previously trained using ImageNet Large-Scale Visual Recognition Challenge 2012 database. The transfer learning with fine-tuning method was employed in this study by making minor changes of altering the last fully connected layer parameter from 1000 classes into binary classes, which is the final output layer that meets the categorization of benign or malignant. The study demonstrated that making minor changes to the pre-trained DCNN models can produce high accuracy during the training process using the training dataset on NVIDIA DIGITS that runs on NVIDIA's GPU, GeForce GTX1060 6GB memory. The parameter includes learning rate and optimization can affect the performance of pre-trained DCNN model and the result of training model was obtained with high accuracy and low loss from each DCNN model which is AlexNet with 83.26% accuracy and 0.45 loss, GoogLeNet with 89.12% accuracy and 0.43 loss, VGG-16 with 86.22% accuracy and 0.64 loss.

The results of classification of each pre-trained DCNN model also obtained by using a test dataset to produce a confusion matrix and receiver operating characteristic (ROC) curve. Based on the confusion matrix, the accuracy, sensitivity/ recall, precision, specificity and F1-score can be calculated, and as a result, AlexNet and GoogLeNet have the highest specificity (71.48%); GoogLeNet have the highest accuracy (73.89%) and precision (72.79%); VGG-16 have the highest sensitivity/ recall (80%) and F1-score (75.26%). Area under the curve (AUC) obtained by ROC was used to evaluate the DCNN model's classifier performance, and as an outcome, VGG-16 (0.8207) has the highest AUC, trailed by GoogLeNet (0.8064) and AlexNet (0.7601).

In conclusion, the pre-trained DCNN model was successfully trained by using the optimization method of transfer learning and data augmentation, to classify benign or malignant from breast mass patch of mammogram images. Among three of pre-trained DCNN model, the overall ability of VGG-16 and GoogLeNet was superior over AlexNet in discriminated between benign or malignant. To concluded, it can be stated that this study has accomplished all of its objectives.

5.2 Suggestions for future works

The findings of this study suggested that the technique used in pre-trained DCNN model was performed well in classification of breast masses patch on the CBIS-DDSM database. Despite the prominent results, there are still room for improvement. The suggestions for methods to further improve the performance of classification are as follows: -

i. A large amount of training dataset that contains a sufficient number of diverse cases. If more training datasets obtained, the DCNN model's performance can improve and more precise in generalization capabilities. More data can be obtained from other public databases such as INbreast, Mammography Image Analysis Society (MIAS) and Breast Cancer Digital Repository (BCDR) database.
ii. Using much deeper pre-trained neural network includes VGG-19, ResNet50, DenseNet-121, etc. Deeper network allows it to learn the details of the image characteristic and extract features from the breast masses, improving the classification outcomes. A large amount of training dataset is required when using deeper networks to prevent overfitting.

iii. Using k-fold cross-validation technique throughout the training process of DCNN model. In k-fold cross-validation, the training dataset is randomly divided into equal-sized groups, referred to as k-fold values. For example, in 10-fold cross-validation, the dataset is randomly divided into ten groups, with one group serves as a test set and the other nine groups serve as a training set. The final classification results were obtained from averaging all randomized k-folds cross-validation results. Different k-folds value produces different classification results.

iv. This study training dataset was divided into 60%-20%-20% for training-validation-test set data split. Study by Mehra (2018) shows that the size allocated for training set has a major effect on the DCNN model's performance, where 90%-10% training-test data split was better performed than 80%-20% and 70%-30% training-test data split on their study of breast histopathological images.

v. Using deep learning - machine learning hybrid system where deep learning model (i.e., VGG-16 or ResNet50) used as feature extractor and machine learning classifier used to perform classification task on breast mass patch. The hybrid system can produce high performance of classification on breast masses compared when employing only machine learning or deep learning methods.

REFERENCES

Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief, 28*, 104863. doi:https://doi.org/10.1016/j.dib.2019.104863

Almaadeed, N., Elharrouss, O., Al-ma'adeed, S., Bouridane, A., & Beghdadi, A. (2019). *A Novel Approach for Robust Multi Human Action Detection and Recognition based on 3-Dimentional Convolutional Neural Networks*.

Alzubaidi, L., Al-Shamma, O., Fadhel, M. A., Farhan, L., Zhang, J., & Duan, Y. (2020). Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics, 9*(3), 445.

Anderson, W. F., Jatoi, I., Tse, J., & Rosenberg, P. S. (2010). Male breast cancer: a population-based comparison with female breast cancer. *Journal of Clinical Oncology, 28*(2), 232.

Ariji, Y., Yanashita, Y., Kutsuna, S., Muramatsu, C., Fukuda, M., Kise, Y., . . . Ariji, E. (2019). Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 128*(4), 424-430. doi:https://doi.org/10.1016/j.oooo.2019.05.014

Aydiner, A., İğci, A., & Soran, A. (2015). *Breast Disease: diagnosis and pathology*: Springer.

Azizah, A., Hashimah, B., Nirmal, K., Siti Zubaidah, A., Puteri, N., Nabihah, A., . . . Sharifah, S. (2019). Malaysia National cancer registry report (MNCR).

Bakkouri, I., & Afdel, K. (2017). *Breast tumor classification based on deep convolutional neural networks*. Paper presented at the 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP).

Breast Cancer. (2021). Retrieved from https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=Scope%20of%20the%20problem,the%20world's%20most%20prevalent%20cancer.

Breast cancer now most common form of cancer: WHO taking action. (2021). Retrieved from https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., . . . Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. *Physica Medica, 83*, 9-24.

Cristofanilli, M., Valero, V., Buzdar, A. U., Kau, S. W., Broglio, K. R., Gonzalez-Angulo, A. M., . . . Buchholz, T. A. (2007). Inflammatory breast cancer (IBC) and patterns of recurrence: understanding the biology of a unique disease. *Cancer:*

Interdisciplinary International Journal of the American Cancer Society, *110*(7), 1436-1444.

Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). Screening for breast cancer. *Jama*, *293*(10), 1245-1256.

Falconi, L. G., Perez, M., Aguilar, W. G., & Conci, A. (2020). Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. *Advances in Science, Technology and Engineering Systems*, *5*(2), 154-165.

Fatima, K., Mohsin, F., Rao, M. O., & Alvi, M. I. (2021). Mammographic Breast Density in Pakistani Women, Factors Affecting It, and Inter-Observer Variability in Assessment. *Cureus*, *13*(3).

Fornell, D. (2021). Use of Breast MRI Screening in Women with Dense Breasts. Retrieved from https://www.itnonline.com/article/use-breast-mri-screening-women-dense-breasts

Gilbert, F. J., & Pinker-Domenig, K. (2019). Diagnosis and Staging of Breast Cancer: When and How to Use Mammography, Tomosynthesis, Ultrasound, Contrast-Enhanced Mammography, and Magnetic Resonance Imaging. In J. Hodler, R. A. Kubik-Huch, & G. K. von Schulthess (Eds.), *Diseases of the Chest, Breast, Heart and Vessels 2019-2022: Diagnostic and Interventional Imaging* (pp. 155-166). Cham: Springer International Publishing.

Gøtzsche, P. C., & Jørgensen, K. J. (2013). Screening for breast cancer with mammography. *Cochrane database of systematic reviews*(6).

Hackshaw, A. K., & Paul, E. A. (2003). Breast self-examination and death from breast cancer: a meta-analysis. *British journal of cancer*, *88*(7), 1047-1053.

Hadadi, I., Rae, W., Clarke, J., McEntee, M., & Ekpo, E. (2021). Breast cancer detection: Comparison of digital mammography and digital breast tomosynthesis across non-dense and dense breasts. *Radiography*, *27*(4), 1027-1032. doi:https://doi.org/10.1016/j.radi.2021.04.002

Hoff, S. R., Abrahamsen, A.-L., Samset, J. H., Vigeland, E., Klepp, O., & Hofvind, S. (2012). Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography—results from a retrospective review. *Radiology*, *264*(2), 378-386.

Houssami, N., & Skaane, P. (2013). Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *The Breast*, *22*(2), 101-108.

Hurria, A., Leung, D., Trainor, K., Norton, L., & Hudis, C. (2003). Screening chest imaging studies are not effective in the follow-up of breast cancer patients. *The Journal of oncology management: the official journal of the American College of Oncology Administrators*, *12*(5), 13-15.

Hylton, N. (2005). Magnetic resonance imaging of the breast: opportunities to improve breast cancer management. *Journal of Clinical Oncology*, *23*(8), 1678-1684.

James, J., Teo, M., Ramachandran, V., Law, M., Stoney, D., & Cheng, M. (2019). A critical review of the chest CT scans performed to detect asymptomatic synchronous metastasis in new and recurrent breast cancers. *World journal of surgical oncology, 17*(1), 1-7.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., . . . Darrell, T. (2014). *Caffe: Convolutional Architecture for Fast Feature Embedding*. Paper presented at the Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA. https://doi.org/10.1145/2647868.2654889

Kang, D.-Y., Duong, H., & Park, J.-C. (2020). Application of Deep Learning in Dentistry and Implantology. *The Korean Academy of Oral and Maxillofacial Implantology, 24*, 148-181. doi:10.32542/implantology.202015

Kelly, K. M., Dean, J., Comulada, W. S., & Lee, S.-J. (2010). Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology, 20*(3), 734-742.

Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H.-J., & Kim, N. (2019). Deep Learning in Medical Imaging. *Neurospine, 16*(4), 657-668. doi:10.14245/ns.1938396.198

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*, 1097-1105.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., . . . Schild, H. H. (2005). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of Clinical Oncology, 23*(33), 8469-8476.

Kuwada, C., Ariji, Y., Fukuda, M., Kise, Y., Fujita, H., Katsumata, A., & Ariji, E. (2020). Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 130*(4), 464-469. doi:https://doi.org/10.1016/j.oooo.2020.04.813

Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., . . . Feig, S. (2010). Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *Journal of the American college of radiology, 7*(1), 18-27.

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data, 4*(1), 170177. doi:10.1038/sdata.2017.177

Lévy, D., & Jain, A. (2016). Breast mass classification from mammograms using deep convolutional neural networks. *arXiv preprint arXiv:1612.00542*.

Lewin, J. M., D'Orsi, C. J., Hendrick, R. E., Moss, L. J., Isaacs, P. K., Karellas, A., & Cutter, G. R. (2002). Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *American Journal of Roentgenology, 179*(3), 671-677.

Li, J., Wu, Y., Shen, N., Zhang, J., Chen, E., Sun, J., . . . Zhang, Y. (2020). A fully automatic computer-aided diagnosis system for hepatocellular carcinoma using convolutional neural networks. *Biocybernetics and Biomedical Engineering, 40*(1), 238-248. doi:https://doi.org/10.1016/j.bbe.2019.05.008

Lim, G. C. C., Rampal, S., & Yahaya, H. (2008). *Cancer Incidence in Peninsular Malaysia, 2003-2005: The Third Report of the National Cancer Registry, Malaysia*: National Cancer Registry.

Liu, X., & Tang, J. (2014). Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method. *IEEE Systems Journal, 8*(3), 910-920. doi:10.1109/JSYST.2013.2286539

McDonald, E. S., Oustimov, A., Weinstein, S. P., Synnestvedt, M. B., Schnall, M., & Conant, E. F. (2016). Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. *JAMA oncology, 2*(6), 737-743.

Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express, 4*(4), 247-254.

Melnikow, J., Fenton, J. J., Whitlock, E. P., Miglioretti, D. L., Weyrich, M. S., Thompson, J. H., & Shah, K. (2016). Supplemental screening for breast cancer in women with dense breasts: a systematic review for the US Preventive Services Task Force. *Annals of internal medicine, 164*(4), 268-278.

Mordang, J. J., Gubern-Mérida, A., Bria, A., Tortorella, F., Mann, R. M., Broeders, M. J. M., . . . Karssemeijer, N. (2018). The importance of early detection of calcifications associated with breast cancer in screening. *Breast Cancer Research and Treatment, 167*(2), 451-458. doi:10.1007/s10549-017-4527-7

Ottesen, G. L., Graversen, H. P., Blichert-Toft, M., Christensen, I. J., & Andersen, J. A. (2000). Carcinoma in situ of the female breast. 10 year follow-up results of a prospective nationwide study. *Breast Cancer Research and Treatment, 62*(3), 197-210.

Pacal, I., Karaboga, D., Basturk, A., Akay, B., & Nalbantoglu, U. (2020). A comprehensive review of deep learning in colon cancer. *Computers in Biology and Medicine, 126*, 104003. doi:https://doi.org/10.1016/j.compbiomed.2020.104003

Piperkova, E., Raphael, B., Altinyay, M. E., Castellon, I., Libes, R., Sandella, N., . . . Abdel-Dayem, H. (2007). Impact of PET/CT in comparison with same day contrast enhanced CT in breast cancer management. *Clinical nuclear medicine, 32*(6), 429-434.

Rahman, A. S. A., Belhaouari, S. B., Bouzerdoum, A., Baali, H., Alam, T., & Eldaraa, A. M. (2020). *Breast Mass Tumor Classification using Deep Learning.* Paper presented at the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT).

Rampun, A., Scotney, B. W., Morrow, P. J., & Wang, H. (2018). *Breast mass classification in mammograms using ensemble convolutional neural networks.* Paper presented at the 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom).

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision, 115*(3), 211-252. doi:10.1007/s11263-015-0816-y

Salama, W. M., Elbagoury, A. M., & Aly, M. H. (2020). Novel breast cancer classification framework based on deep learning. *IET Image Processing, 14*(13), 3254-3259.

Shawky, M., Ali, Z. A. E., Hashem, D. H., & Houseni, M. (2020). Role of positron-emission tomography/computed tomography (PET/CT) in breast cancer. *Egyptian Journal of Radiology and Nuclear Medicine, 51*(1), 125. doi:10.1186/s43055-020-00244-9

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Skaane, P. (2016). Breast cancer screening with digital breast tomosynthesis. In *Digital Breast Tomosynthesis* (pp. 11-28): Springer.

Sturesdotter, L., Sandsveden, M., Johnson, K., Larsson, A.-M., Zackrisson, S., & Sartor, H. (2020). Mammographic tumour appearance is related to clinicopathological factors and surrogate molecular breast cancer subtype. *Scientific Reports, 10*(1), 20814. doi:10.1038/s41598-020-77053-7

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians, 71*(3), 209-249.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Tian, Y., & Fu, S. (2020). A descriptive framework for the field of deep learning applications in medical images. *Knowledge-Based Systems, 210*, 106445. doi:https://doi.org/10.1016/j.knosys.2020.106445

Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep learning for breast cancer diagnosis from mammograms—a comparative study. *Journal of Imaging, 5*(3), 37.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

Zhang, X., Sasaki, T., Suzuki, S., Takane, Y., Kawasumi, Y., Ishibashiz, T., . . . Yoshizawa, M. (2017). *Classification of mammographic masses by deep learning.* Paper presented at the 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE).

Zhang, X., Zhang, Y., Han, E. Y., Jacobs, N., Han, Q., Wang, X., & Liu, J. (2017). *Whole mammogram image classification with convolutional neural networks.* Paper presented at the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

Zhang, X., Zhang, Y., Han, E. Y., Jacobs, N., Han, Q., Wang, X., & Liu, J. (2018). Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. *IEEE transactions on nanobioscience, 17*(3), 237-242.

Zhou, H., Zaninovich, Y., & Gregory, C. (2017). *Mammogram classification using convolutional neural networks.* Paper presented at the International conference on technology trends.