MICROCALCIFICATION DETECTION IN MAMMOGRAPHY FOR EARLY BREAST CANCER DIAGNOSIS USING DEEP LEARNING TECHNIQUE

LEONG YEW SUM

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

MICROCALCIFICATION DETECTION IN MAMMOGRAPHY FOR EARLY BREAST CANCER DIAGNOSIS USING DEEP LEARNING TECHNIQUE

LEONG YEW SUM

RESEARCH PROJECT SUBMITTED TO THE FACULTY OF ENGINEERING UNIVERSITY OF MALAYA, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BIOMEDICAL ENGINEERING

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Leong Yew Sum

Matric No: S2039447

Name of Degree: Master of Biomedical Engineering

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

Microcalcification Detection in Mammography for Early Breast Cancer Diagnosis using Deep Learning Technique

Field of Study: Machine Learning

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

Subscribed and solemnly declared before,

Witness's Signature Date

Name:

Designation

MICROCALCIFICATION DETECTION IN MAMMOGRAPHY FOR EARLY BREAST CANCER DIAGNOSIS USING DEEP LEARNING TECHNIQUE

ABSTRACT

Breast Cancer is one of the common cancers in women and may cause lives to be lost if they were misdiagnosed and left untreated. Existence of breast microcalcifications are common in breast cancer patients and they are an effective indicator of early breast cancer. This project will incorporate the use of machine learning in segmenting breast mammogram images with calcifications of either benign or malignant cases for early breast cancer diagnosis. ROI images of breast microcalcification will be utilized to train several pretrained models from fastai library in Google Colaboratory platform using supervised learning with a ratio of 0.80 for training dataset and 0.20 for validation dataset. Image processing of ROI images were conducted to remove possible artifacts and noises in order to enhance the quality of the images before training. The pretrained models that were included in this study are Resnet34, Resnet50, VGG16 and Alexnet. Different hyperparameters such as epoch, batch size etc were tuned in order to obtain the best possible result in this study. Confusion matrices were utilized in order to measure the output parameters of the models for comparison in terms of performance. The result from this study shows that Resnet50 achieves the highest accuracy with a value of 97.58%, followed by Resnet34 of 97.35%, VGG16 of 96.97% and finally Alexnet of 83.06%.

Keywords: breast microcalcification, deep learning, automated segmentation, image processing

PENETAPAN MIKROKALIFIKASI DALAM MAMMOGRAFI UNTUK DIAGNOSIS KANSER PAYUDARA AWAL DENGAN MENGGUNAKAN TEKNIK PEMBELAJARAN DALAM

ABSTRAK

Kanser Payudara adalah salah satu jenis barah yang lazimnya terjadi pada wanita dan boleh menyebabkan nyawa hilang sekiranya salah didiagnosis dan tidak dirawat. Kehadiran mikrokalsifikasi payudara adalah perkara yang lazimnya dijumpai pada pesakit barah payudara dan ia merupakan petunjuk berkesan untuk barah payudara awal. Projek ini akan menggabungkan penggunaan pembelajaran mesin dalam klasifasikasi gambar mamogram payudara kepada kes jinak atau malignan untuk diagnosis kanser payudara awal. Imej mikrokalsifikasi payudara akan digunakan untuk melatih beberapa model pra-latihan dari perpustakaan *fastai* di platform *Google Colaboratory* melalui pembelajaran yang diselia, dengan menggunakan nisbah 0.80 untuk set data latihan dan 0.20 untuk set data pengesahan. Pemprosesan gambar gambar ROI telah dilaksanakan untuk menghilangkan kemungkinan artifak dan mengurangkan bunyi imej untuk meningkatkan kualiti gambar sebelum latihan. Model pra-latihan yang termasuk dalam kajian ini adalah Resnet34, Resnet50, VGG16 dan Alexnet. Hiperparameter yang berbeza seperti *epoch*, ukuran kumpulan dan lain-lain telah ditala untuk mendapatkan hasil yang terbaik dalam kajian ini. *Confusion Matrix* telah digunakan untuk mengukur parameter output model untuk perbandingan dari segi prestasi. Hasil dari kajian ini telah menunjukkan bahawa Resnet50 mencapai ketepatan tertinggi dengan nilai 97.58%, diikuti oleh Resnet34 dengan nilai 97.35%, VGG16 dengan nilai 96.97% dan akhirnya Alexnet dengan nilai 83.06%.

Kata Kunci: mikrokalsifikasi payudara, pembelajaran dalam, automatik segmentasi, imej pemprosesan

ACKNOWLEDGEMENTS

To make this project a success, many people have bestowed their blessings and genuine support on me. I would like to take this opportunity to express my gratitude to everyone who has contributed to the success of this project. First and foremost, I would like to express my gratitude to the Almighty God for blessing me over the length of this project, which would not have been possible without His grace. Moving on, I would like to thank my parents, who raised me with love and support to this point in my life. My deepest gratitude goes out to my parents for their mental and financial support of my studies. Without their support, I wouldn't have gone as far as I have. Furthermore, I would like to express my gratitude to my supervisor, Dr Khairunnisa Binti Hasikin, for helping me throughout the study of this research. She has provided me with a great deal of useful guidance that immediately benefits my academic endeavours. Finally, I would like to express my gratitude to my peers for their enthusiasm and support for my project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AI	:	Artificial Intelligence
BC	:	Breast Cancer
CAD	:	Computer - aided Diagnosis
CC	:	Craniocaudal
CIBS-DDSM	:	Curated Breast Imaging Subset of DDSM
CNN	:	Convolutional Neural Networks
Conv	:	Convolutional
CT	:	Computed Tomography
DDSM	:	Digital Database Screening Mammography
DL	:	Deep Learning
ERBB2	:	Epidermal Growth Factor 2
FC	:	Fully Connected
LR	:	Learning Rate
MCC	:	Matthew Correlation Coefficient
MLO	:	Mediolateral Oblique
MIAS	:	Mammographic Image Analysis Society
MRI	:	Magnetic Resonance Imaging
NAC	:	Nipple Areola - complex
PET	:	Positron-emission Tomography
RAG	:	Region Adjacency Graph
ReLU	:	Rectified Linear Activation Function
ROI	:	Region of Interest
SPECT	:	Single-photon Emission Computed Tomography
SVM	:	Support Vector Machines
TCIA	:	The Cancer Imaging Archive
TL	:	Transfer Learning
WHO	:	World Health Organization

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background Study

According to the World Health Organization (WHO) (2021), in the year 2020, breast cancer has affected 2.3 million people worldwide, with 685, 000 fatalities, making it the most common disease in the world. When it comes to lowering the mortality rate, accurate diagnosis and assessment of breast cancer in its early stages is critical. Therefore, an effective method of breast cancer screening is necessary.

As reported by Kashif et al. (2020), detecting and diagnosing a breast lesion purely based on mammography results is challenging and heavily rely on the radiologist's skill. Newton (2019) has reported that the false-negative rate of mammography is approximately 8-10 percent, according to statistics from the Breast Cancer Detection Demonstration Project (BCDDP). Based on the recent study of Batchu et al (2021), false negative cases are most likely to occur in women aged 50–89 who have had prior benign biopsies. False negative cases are an alarming issue as it might cause the patient to miss the best possible treatment time. Advancements in artificial intelligence (AI) have been used to build software in assisting radiologists during clinical diagnosis, to improve accuracy and minimizing the rates of false positives and false negatives. Therefore, automatic data analysis and smart data with high computation has played an important role in the healthcare industry to deal with human-errors.

Artificial intelligence (AI) allows a machine to make judgments based on the information provided to it (Lytras & Visvizi, 2018). According to Barot (2020), Deep Learning (DL) is a subset of Machine Learning, which is a subset of AI as well. DL algorithms, which may also be referred to as Deep Convolutional Neural Networks (D-CNN), have quickly become the preferred approach for evaluating medical images (Litjens et al., 2017). LeCun, Bengio, & Hinton (2015) agreed that DL allowed computational models to learn multiple degrees of abstraction for data representations which significantly improved the state-of-the-art in image processing. Deep learning may be used to develop and test algorithms that help with prediction, pattern recognition, and classification.

According to the study of Spruit & Lytras (2018), installation of a sophisticated computer programme in healthcare necessitates a multi-pronged strategy as it often involves political, economical and social issues. Safdarian & Hedyezadeh (2019) has asserted that the application of deep learning in image processing and pattern recognition in the diagnosis and classification of breast cancer from mammogram images is able to help clinicians by lowering human error and improving detection time.

1.2 Problem Statement

Mammograms can be challenging to read, resulting in incorrect findings as the breast tissue's density varies greatly among women. Dense breasts are more difficult to analyse in a picture, limiting the capacity of the mammography to properly identify cancer (Zhao et. al., 2015). Mammograms often failed to detect the exact size and location of the lesion due to thick glands and overlapping structures.

In terms of detection, diagnosis, and treatment, many nations lack the human resources and technological capabilities to deliver timely care to breast cancer patients (Beeravolu et al., 2021). Although various ML techniques have recently been utilised to build computer-aided diagnosis (CAD) systems to improve breast cancer diagnosis in medical pictures, they are mostly based on hand-crafted features and the systems have frequently yielded false positive and false negative results (Alkhaleefah & Wu, 2018) . Thus these approaches are seen to be laborious, time-consuming, and necessitate the use of specialists in the area, particularly for feature extraction and selection activities.

Calcifications may appear in the breast. Specific calcification patterns might signify breast cancer or precancerous alterations in breast tissue, therefore it is very important to detect breast calcifications earlier. Calcification poses a problem in terms of perception as well as interpretation. Small calcification clusters are easy to overlook and misinterpret. The rate of microcalcification detected at screening is determined by the population's age and screening frequency. As a result, establishing a baseline of expected values for evaluation and the incidence of cancer detection from calcifications is challenging (Wilkinson,Thomas, & Sharma, 2017).

1.3 Aims and Objectives

This study attempts to allow early detection of breast cancer using deep learning algorithms to automatically classify microcalcification

Specific objectives of the research project is as following:-

1. To perform preprocessing operations for the collected mammogram images prior to the use with deep learning algorithms.
2. To employ the transfer learning technique of CNN to build a breast cancer image classifier.
3. To compare classification performance of machine learning based models in distinguishing between benign and malignant cases of breast cancer.

1.4 Scope of Research Project

The scope of this research focuses on the development of an automated breast microcalcification detection in mammography for early breast cancer diagnosis. This research will be utilizing CIBS-DDSM dataset from Cancer Imaging Archive (TCIA). The image dataset consists of digital mammography ROI images of breast microcalcification in grayscale. The images were downloaded using NBIA Data Retriever and are further resized to a resolution of 224 x 224 by using DICOM.

This research has included preprocessing of mammography images for artifacts and noise removal in order to enhance the quality of the image before feeding them to the pretrained models. As such, Otsu Segmentation method and MorphologicalEx Transformation method under the research of Xi, Shu & Goubran (2018) will be utilized to perform artifact removal. For noise removal, adaptive median filter, median filter, and mean filter will be tested to perform noise removal on the raw ROI images and the filter with lowest mean square error (MSE) and highest Peak Signal to Noise Ratio (PSNR) will be selected as the best final filter in preprocessing the images for image enhancement.

Pretrained models were obtained from fastai library in Google Colaboratory platform to facilitate transfer learning, which includes Resnet34, Resnet50, VGG16, and Alexnet. Different hyper parameters such as epoch, batch size and learning rate will be tuned in order to achieve the best possible result of the model. Overall, this paper proposes the use of residual CNN - ResNet-50, which has 50 layers deep, to classify breast mammogram images to benign or malignant cases to assist diagnosis of breast cancer. Different models were included in this research to compare their performance.

Upon training the model, a confusion matrix will be used in order to compute and compare the performance of each model to deduce the best model for breast image classification. The proposed model will be developed in Google Colaboratory platform with Intel Core i7-4710 HQ CPU @3.5 GHz, 1 TB SSD Memory and 4 GB RAM environment.

1.5 Structure of Research Project

The structure of this research consists of five main chapters and the outline is structured as follows:

- Chapter I: Introduction

  Chapter I delivers the background study of this research. The problem statement and the research objectives were included in this chapter.

- Chapter II : Literature review

  Chapter II gives thorough review on breast cancer as well as the different types of calcification that might present in patients. This chapter also highlights the type of preprocessing that is conducted on breast mammography images before feeding them into deep learning models. In addition to that, this chapter also discusses the type of supervised training and gives an overview on the architecture of deep learning.

- Chapter III : Methodology

  Chapter III proposes this research's work, including the materials required for research, the database used, image processing techniques, as well as the type of CNN models for tests. Upon obtaining the output of the models, a confusion matrix will be used in order to measure the output parameters of the model such as accuracy, specificity, F1-score and more.

- Chapter IV: Result and Discussion

  Chapter IV shows the output of the model in terms of different parameters and provides detailed discussion based on the result obtained. The performance of the model built was compared to the state-of-art models.

- Chapter V: Conclusion and Future Work

  Chapter V indicates a brief summary of this research's work and proposes future improvement or exploration possible on the studied topic.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In the study of Siegel, Miller, & Jemal (2015), breast cancer (BC) is defined as the most prevalent kind of cancer among women. BC can be caused by a variety of variables, such as hormonal imbalances, or issues related to reproductive organs (Amrane et al., 2018). Sun et al. (2017) states that early detection of the disease can result in a favourable prognosis and a higher percentage of survival.

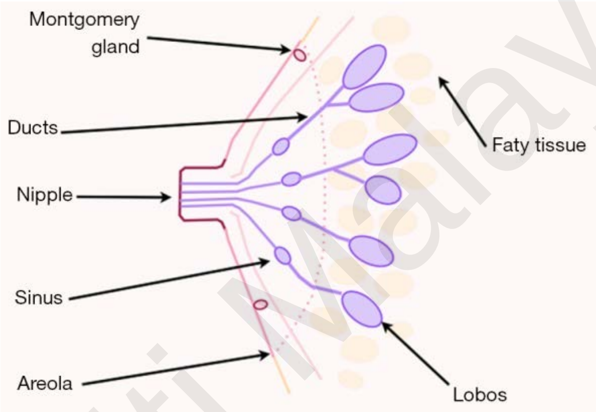


Figure 2.1: Schematic view of nipple areola-complex (Zucca-Matthes, Urban, & Vallejo, 2016).

The interior structure of a breast picture is shown in Figure 2.1. A woman's breast is made of lobules, ducts , nipples, and fatty tissues (Zucca-Matthes, Urban, & Vallejo, 2016). Usually, epithelial tumours develop inside the lobules and ducts of the breast nipple areola complex (NAC), eventually becoming cancer.

Imaging of the breast can be conducted in two basic methods - invasive or non-invasive. Examples of non-invasive imaging techniques are mammography magnetic resonance imaging (MRI), positron-emission tomography (PET) and computed tomography (CT) (Jafari et. al., 2018). Mammography is the gold standard technique for detecting breast cancer patients due to high sensitivity and specificity, low cost, and good tolerability (Wellings, Vassiliades & Abdalla, 2016). This imaging has still remained the most effective technique for general population screening up to this day (Tsochatzidis, Costaridou & Pratikakis, 2019).

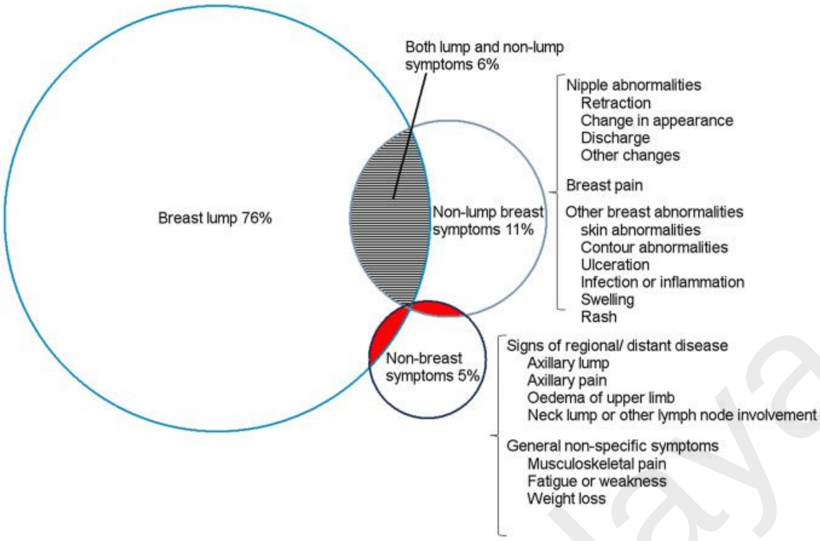


Figure 2.2: Venn diagram on symptoms of breast cancer (Koo et al., 2017).

Breast lump is the most common presenting symptom among women with breast cancer (refer Figure 2.2). Studies of Redaniel et al. (2015) showed that breast lump has relatively high predictive value for malignancy. Breast tumours may contain calcification and they appear as white dots on mammogram images.

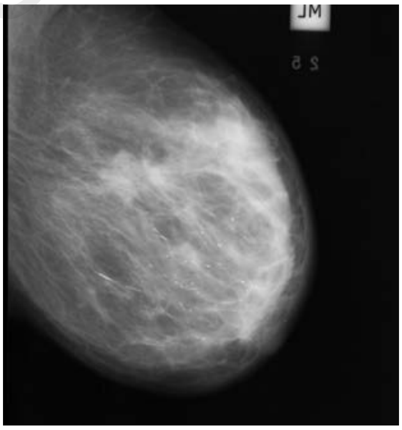


Figure 2.3: Breast mammogram images with calcifications

Calcifications on the breast are frequent, especially beyond the age of 50. Specific calcification patterns might signify breast cancer or precancerous alterations in breast tissue. Breast calcifications can appear as macrocalcifications or microcalcifications on a mammogram (Mayo Clinic, 2021):

- Macrocalcifications. Large white spots or dashes that are usually cancer-free and don't require any more testing or follow-up.
- Microcalcifications. Tiny, white flecks that resemble salt granules. Typically not malignant but specific patterns can be an indication of cancer in its early stages.

Nahid & Kong (2017) stated that calcifications under benign (non-cancerous) tumors are non-life-threatening. They will rarely progress to malignant (cancerous). Benign tumours are separated from other cells by the immune system known as "sac" and may be simply eliminated from the body. On the other hand, malignant tissue nuclei are often significantly larger than normal tissue nuclei. They begin as an uncontrolled cell growth that spreads quickly and invades adjacent tissue. Cheriyedath (2021) believes that calcifications are a reliable early indicator of breast cancer and may provide insight into the disease's severity. The size, pattern, density, and location of breast microcalcifications can provide insight into whether the tumour is benign or malignant.

The current sensitivity of screening for malignant calcifications is quite poor (Mordang et al., 2018). Even with visible calcifications, the majority of lesions are not recalled immediately, but identified as interval cancer in subsequent screening. Interval cancers are primary breast cancers that are discovered in women after a negative screening exam, defined as no recall suggestion or a negative aftercall screening, before or within a period of two years after a second screening (van Bommel et al., 2017). Therefore, methods that allow accurate detection of individuals with malignant calcifications cases without raising false positives must be developed to allow earlier treatments for breast cancer patients. Examples of studies that involve classification of microcalcification of breast into malignant and benign cases are illustrated in Table 2.1.

Table 2.1: Models of breast image classifier for microcalcification detection

Reference	Base Model	Type of image	Database	Accuracy
Wang et al. (2016)	SVN	Histopathology	Private*	85.8%
	KNN			84.3%
	Linear discriminant analysis (LDA)			74.0%
	Stacked autoencoder (SAE)			89.7%
Heenaye-Mamode et al. (2021)	Resnet50	Mammography	CBIS-DDSM and UPMC	88.0%
Hakim, Prajitno, & Soejoko (2021)	Resnet50	Mammography	INBreast	90.3 %
Li et al. (2021)	CancerNet (modified VGG16)	Mammography	PINUM (Private)	90.0%
			DDSM	87.0%
Khamparia et al. (2021)	Modified VGG	Mammography	DDSM	94.3%
Xiao et al. (2021)	2D Resnet34 with anisotropic 3D Resnet	Digital Breast Tomosynthesis (DBT) images	Private DBT	76%
Cai et al. (2019)	Alexnet	Mammography	Private*	88.6%
Hekal, Elnakib & Moustafa (2021)	Modified Alexnet	Mammography	CBIS-DDSM from TCIA	84.0%
	Modified Resnet50			91.0%
Tsochatzidis, Costaridou, & Pratikakis (2019)	Alexnet	Mammography	CIBS-DDSM from TCIA	75.3%
	VGG16			71.6%
	Resnet50			74.9%

Private = SunYat-sen University Cancer Center (Guangzhou, China) (SYUCC) and Nanhai Affiliated Hospital of Southern Medical University (NAHSMU) (Foshan, China)*

2.2 Breast Image Classifier

Conventional images of mammograms are highly affected by low contrast and unclear boundaries between surrounding normal tissues (Nayak et al., 2019). The introduction of digital mammography images has made deep learning approaches for cancer diagnosis possible in recent years (Abdelhafiz et al., 2019). Deep neural networks have advanced to the point where they can automatically learn from enormous picture data sets and detect abnormalities in mammograms such as mass lesions (Nelson et al., 2016). Generally, a breast image classifier consists of four stages (refer Figure 2.4).

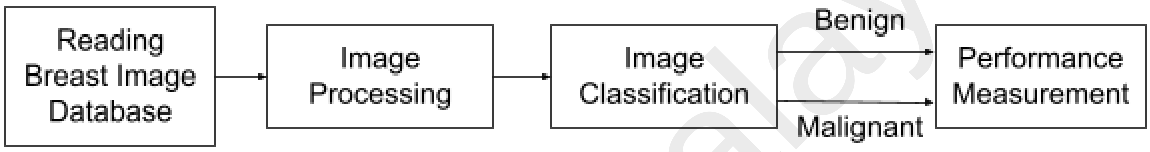


Figure 2.4: Flow chart of breast image classifier

The steps of a breast image classification model is as following:

- Step 1: Reading images of a selected breast image database
- Step 2: Performing selection of features and image enhancement on mammogram images
- Step 3: Feeding information into a Breast Classifier Model to classify benign or malignant case
- Step 4: Measuring the performance of the model to determine accuracy

2.3 Breast Image Database

A well-established picture database is important to make a trustworthy conclusion regarding the diagnosis of cancer. Various organisations have created picture databases that scholars may use to perform additional study. Table 2.2 lists a handful of the databases that are available, along with their parameters.

Table 2.2: Example of dataset (Nahid & Kong, 2017)

Database	Number of images	Database size (GB)	Image capture technique	Total patients
MIAS	322	2.3	MG	161
CBIS-DDSm	4067	70.5	MG	237
ISPY1	386,528	76.2	MR, SEG	237
Breast-MRI-NACT-Pilot	99,058	19.5	MRI	64
QIN-Breast	100835	11.286	PET/CT, MR	67
Mouse-Mammary	23487	8.6	MRI	32
TCGA-BRCA	230167	88.1	MR, MG	139
QIN Breast DCE-MRI	76328	15.8	CT	10
BREAST-DIAGNOSIS	105050	60.8	MRI/PET/CT	88
RIDER Breast MRI	1500	.401	MR	5
Inbreast	419		MG	115

Digital mammogram images are obtainable under Mammographic Image Analysis Society (MIAS), Curated Breast Imaging Subset of DDSM (CIBS-DDSM), and INBreast databases. Recently, InBreast databases are not made available to the public anymore. Nahid & Kong (2017) mentioned that researchers have primarily utilized MIAS and DDSM databases for study of breast image classification.

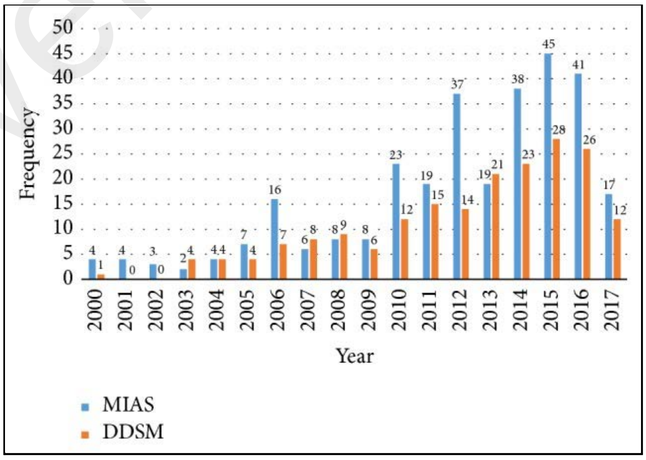


Figure 2.5: Usage of MIAS and DDSM dataset year 2000 to year 2017 (Nahid & Kong, 2017)

The timeline for the usage of DDSM and MIAS dataset from the span of 2000 to 2017 is illustrated in Figure 2.5. The usage of MIAS dataset is higher than DDSM, probably because MIAS dataset is equipped with .csv files which contain features computed for each cell nucleus, such as area, texture, parameter and etc to build machine learning algorithms (Goel, 2018). A comparison of DDSM and MIAS dataset is tabulated in Table 2.3.

Table 2.3:Comparison DDSM and MIAS data (Smith, 2016; Suckling et al., 2015)

DDSM	Aspect	MIAS
Larger	Amount of Data	Smaller
Available in Image	Region of Interest (ROI)	Available in Coordinates
Available	Mediolateral oblique (MLO) view	Not Available
Available	Craniocaudal (CC) view	Not Available
Available	Microcalcification Images	Not Available
Higher	Resolution of Images	Lower

The resolution for the DDSM database is much higher as compared to the resolution of mammogram images in the MIAS database. In addition to that, the amount of images available for DDSM is significantly higher as compared to MIAS. For instance, DDSM contains 4067 images while MIAS only contains 322 images. CBIS-DDSM dataset, which contains 6775 studies, is a subset of the DDSM dataset. This subgroup, on the other hand, is chosen by a mammography specialist. As a result, it's widely accepted as an updated and standardised version of DDSM.

Generally, it is relatively important to prepare an abundant set of images for machine learning to train and increase the accuracy of the model. After considering the pros and cons, the dataset for CBIS-DDSM is utilized in this project. According to the data provided in TCIA webpage, the digital mammogram images are in DICOM format. ROI images of the breast calcification are made available with proper labelling of benign and malignant cases.

2.4 Image Processing for Breast Cancer Mammograms

2.4.1 Removal of Artifacts

Removing artifacts in mammography images is important as they might contribute unnecessary learning features for the transfer learning of developed CNN. Examples of artifacts that might be present at mammography images are patient's details, numbering labels, etc.

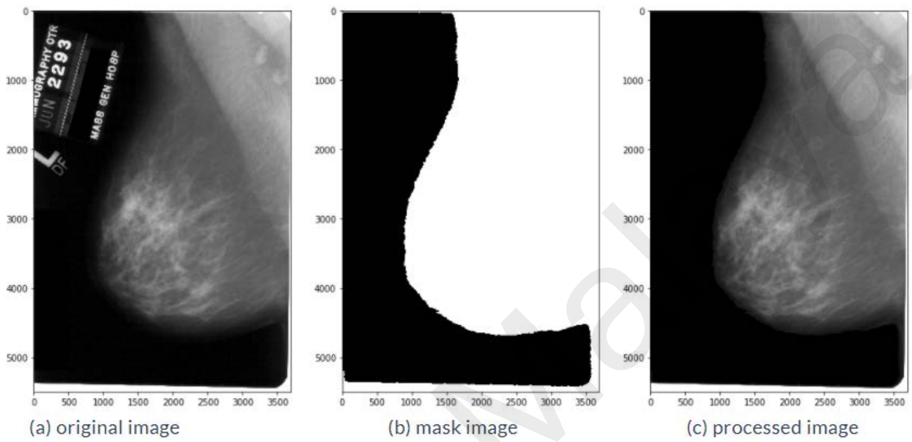


Figure 2.6: Process of artifact removal (Xi, Shu & Goubran, 2018)

Xi, Shu & Goubran (2018) has demonstrated the removal of artifacts by applying *Otsu Segmentation Method* and *MorphologicalEx Method* (refer Figure 2.6). The author has demonstrated the use of both methods in identifying artifacts and enclosing the small region within white artifacts. A mask image was created by selection of 'largest object', which is the breast region. The original image and the mask image were compared bit-by-bit using cv2.bitwise_and(), where the regions which appear black on the mask are applied to the original image.

Otsu Segmentation Method works on grayscale images and involves the use of a global thresholding or local thresholding to classify pixel values (Thanh et al., 2019; Suradi, Abdullah & Isa, 2021). For instance, we denote a mammogram image as a function of G(x,y) and intensity value of I {I = 0,1, 2, …I-1}. The variance of these two variables can be computed by using Equation 2.1.

$$\sigma_m^2 = \Theta_1^{(th)} \cdot \sigma_1^2(th)_{+\Theta_2}(th) \cdot \sigma_2^2(th) \qquad \text{(Equation 2.1)}$$

Whereby,

$$\Theta_1(th) = \sum_{i=1}^{th} P(i) \qquad \text{(Equation 2.2)}$$

$$\Theta_2(th) = \sum_{i=th+1} P(i) \qquad \text{(Equation 2.3)}$$

Threshold value th, which determines the class probability of pixels, is denoted as Θ_1 and Θ_2, and the mean of the class is calculated as $u_1$and as u_2 in Equation 2.4 and 2.5 below. The threshold value that is predetermined earlier $Th\ (k) = k, 0 < the < Ibe$ will be utilized to divide the original mammogram image into two segments according to the intensity, which are $[0, th]$ and $[th + 1, I]$.

to the intensity, which are $[0, th]$ and $[th + 1, I]$.

$$u_1(th) = \sum_{i=1}^{th} \frac{iP_{(i)}}{\Theta_1(th)} \qquad \text{(Equation 2.4)}$$

$$u_2(th) = \sum_{i=th} \frac{iP_{(i)}}{\Theta_2(th)} \qquad \text{(Equation 2.5)}$$

The value of interclass variance and global mean-variance can then be computed by using Equation 2.6 and 2.7 respectively.

$$\sigma_1^2(th) = \sum_{i=1}^{th} [1 - u_i(th)]^2 \frac{P_{(i)}}{\Theta_1(th)} \qquad \text{(Equation 2.6)}$$

$$\sigma_2^2(th) = \sum_{i=th+1}^{I} [1 - u_i(th)]^2 \frac{P_{(i)}}{\Theta_2(th)} \qquad \text{(Equation 2.7)}$$

The optimum threshold is chosen to achieve the best performance in distinguishing the target class from the background class which is mostly utilised in mammography image binarization. Before executing the breast cancer detection segmentation and feature extraction procedure, this thresholding approach is employed as a pre-processing technique (Bhandari, Maurya & Meena, 2018; Khaimar, Thepade & Gite, 2021).

On the other hand, simple logical operations on local groupings of pixels are defined as morphological operators. Two main morphological operations are dilation and erosion, which are shown in Equation 2.8 and 2.9 respectively (Nixon & Aguado, 2012). Many functions, such as opening and closing, are derived from these operators. When a picture is opened, it undergoes erosion and then dilation, and when it is closed, it undergoes dilation and then erosion (Kaur, Virmani & Thakur, 2019).

$$X \ominus B = \left\{ x | B^{1}_{x} \subset X \right\} \qquad \text{(Equation 2.8)}$$

$$X \oplus B = \left\{ x | B^{2}_{x} \subset X \right\} \qquad \text{(Equation 2.9)}$$

Rolling Ball Algorithm is also one of the common algorithms used in artifact removal. Basile et al. (2019) have explained that *Rolling Ball Algorithm* helps to emphasise the key areas inside the breast. Scikit-Image (2021) has further added that this algorithm calculates the background intensity of an image for them to be subtracted as a whole.

Beeravolu et. al (2021) has achieved removal of artifacts by applying *Rolling Ball Algorithm* with combination of *Fuzzy Thresholding* and *Morphological Transformation*. *Fuzzy's Thresholding* applies the idea of acquiring suitable threshold values from the artifacts of an image and label them as a group. The artifacts that appear white in the image can later be degraded and minimized by using *Erosion*, and the minimized degraded image are later expanded again using *Dilation* under *Morphological Transformation*. Lastly, the image produced under *Rolling Ball Algorithm* and *Morphological Transformation* image are merged using bitwise AND to eliminate the artifacts.

2.4.2 Image Enhancement and Noise Reduction

Raw mammography images are frequently accompanied by sounds or undesired signals, causing the image to be degraded in some cases. Filtering, also known as denoising, is an important technique for improving image processing. Odat, Otair & Shehadeh (2015) have highlighted that the basic goal of filtering is to remove undesirable noise from a picture. The primary procedures that may be performed in filtering are noise reduction, edge detection, sharpening, and smoothing. In order to eliminate the noises, here are several image processing techniques that have been adopted. A collective of filters applied in breast image is tabulated in Table 2.4.

Table 2.4 Type of filters for image enhancement and noise reduction in mammography (Ramani, Vanitha, & Valarmathy, 2013; George & Dhas, 2017)

Filter	Description
Mean filter or average filter	Replaces each pixel with the average value of the intensities in the neighbourhood.
Median filtering	Replace a median value of intensity at the input, centre, and output of the scanning window. Able to keep the sharpness of image edges while removing noise.
Adaptive median filter	Discriminates pixels in the scanning window and a filtering technique is applied to corrupted pixels in the window. Able to smoothen the non-repulsive noise without blurring edges.
Wiener filter	Works by building an optimal estimate of the original image by enforcing a minimum mean square error constraint between estimated and original image
Gaussian Filter	Able to minimize the low and high signals from distortion

Abdelhafiz et. al. (2019) have highlighted that the commonly used filters include median filter, adaptive median filter, and mean filter. Research of Massodi,Safdarian,& Kalantar (2015) have shown that median filter works by 'sliding a window' over the image to scan all pixels of the picture and utilises local image processing algorithms to replace a median value of intensity at the input, center, and output of the window to smoothen salt-and-pepper noise.

According to the study of Mehta & Aggarwal (2014), adaptive filtering is a better filtering approach compared to median filtering, which applies filtering just to the image's damaged pixels while leaving the uncorrupted pixels alone. During filtering, the adaptive filtering technique is utilised to minimise the amount of noisy pixels. One of the studies has found that the quality of the output image by adaptive median filter is much superior as compared to other filters (Ramani, Vanitha, & Valarmathy, 2013).

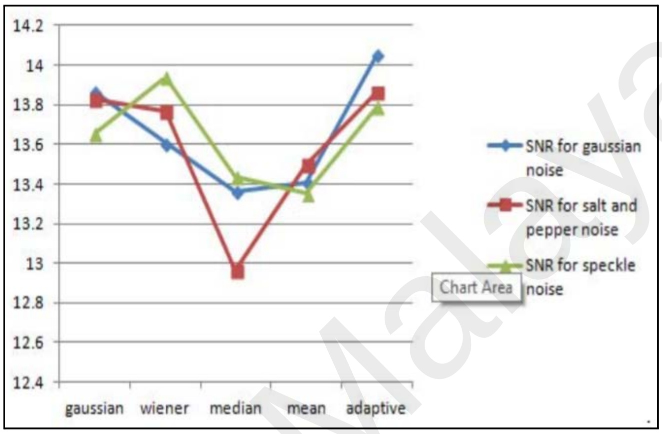


Figure 2.7: Performance of filters for SNR (George & Dhas, 2017)

In the study of George & Dhas (2017), all mean filter, median filter, and adaptive median filter are able to perform well in eliminating salt-and-pepper noise. However, adaptive median filters are able to perform more noise removal. Figure 2.7 has highlighted that the adaptive median filter is proven to be the most effective filter in removing the overall signal-to-noise ratio.

To compare picture compression quality, the mean square error (MSE) and peak signal-to-noise ratio (PSNR) are employed (MathWorks, 2021). The MSE is the squared average of the "errors" between the actual image and the noisy image. The error is the difference between the original image's values and the degraded image's values. Hence, the lower the MSE value, the better is the filter. PSNR value is closely linked with MSE as it is computed based on MSE values. The formula for PSNR is described in Equation 2.10, where MAX_f is the maximum signal value that exists in our original “known to be good” image (National Instruments, 2016).

$$PSNR = 20log_{10}(\frac{MAX_f}{\sqrt{MSE}}) \quad \text{(Equation 2.10)}$$

The value of PSNR tends to be higher if the reconstruction method is superior and the better the damaged picture has been reconstructed to match the original image. In this research, adaptive median filter, median filter and mean filter will be utilized and their performance in terms of PSNR and MSE will be calculated to determine the best filter that will be utilized in this research.

2.4.3 Removal of Pectoral Muscle

Removal of the pectoral muscle is able to effectively reduce the strain of the neural network, which shortens the computing time effectively as it eliminates sections of the picture that aren't needed, resulting in smaller images. There are several methods of removing the pectoral muscle of the breast.

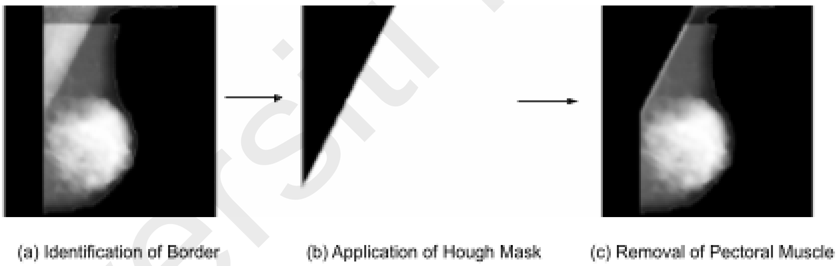


Figure 2.8: Process of removing pectoral muscle (Vagssa et. al., 2020)

In the study of Vagssa et. al. (2020), removal of pectoral muscle was successfully performed by utilizing *Hough Transformation* (refer Figure 2.8). *Hough Line* was used to contain ROI, which is defined by the upper left quarter of the preprocessed image. Moving on, *Canny Filter* was applied to highlight the contours of ROI, and a *Hough Mask* is created based on the lines of contour. Lastly, the *Hough Mask* is applied to ROI and binary subtraction is performed to remove the muscle. The author highlights that the removal of pectoral muscles may not be of a total success as it may be difficult to apply to images that have ROI near the identified pectoral muscles.

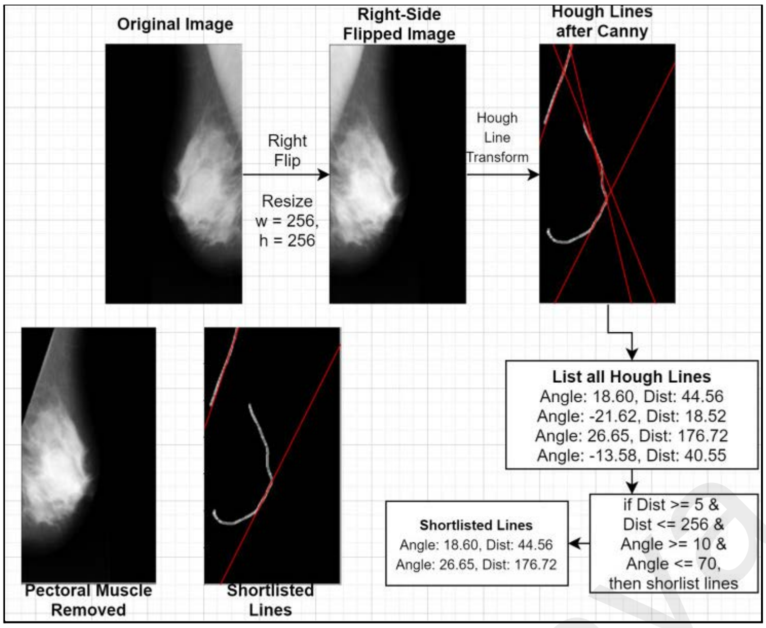


Figure 2.9: Process of removing pectoral muscle (Beeravolu et. al. 2021)

In another study, the usage of Hough line and Canny Edge Method to remove the pectoral muscle was observed (Beeravolu et. al, 2021). Canny edge technique is used to extract the breast's edges, the Hough transform is utilised to calculate the distance and angle to the muscle border from the image's centre. Figure 2. 9 shows the process of pectoral muscle removal.

The research of Alam & Islam (2014) has successfully demonstrated the pectoral muscle removal by utilizing *K-Means Clustering Approach.* Firstly, thresholding technique is used to scan all the intensity of the pixels to identify sudden big intensity near the edge location. This big intensity is then marked. To get a high contrast image with fine features, the *Contrast Equalisation Technique* is used. Using *MorphologyEx* and the *K-Means Clustering* method, the pectoral muscle is then excluded from the image, and a straight line is predicted based on the edge point. Later, the area, or pectoral muscle, is removed based on the anticipated line.

2.4.4 Segmentation

Image segmentation is the process of dividing a digital image into numerous parts in digital image processing. Image segmentation is a technique of locating the region of interest (ROI), such as breast calcification, for machine learning.

The aim of segmentation is to split pictures with various characteristic tissues into areas that can be interpreted semantically (Shareef, 2014). A precise and consistent approach for the segmentation of breast tumours is essential in the detection and quantification of breast cancer (Shrivastava & Bharti,2019).

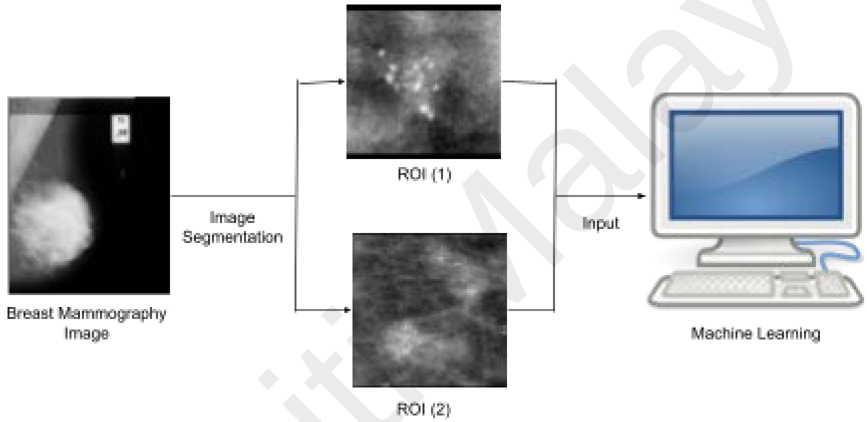


Figure 2.10: Workflow of image processing

Generally, there are several methods that can be conducted in order to perform image segmentation (refer Table 2.5). In the study of Kashif et. al. (2020), *Region-Based Segmentation* and *Edge-Based Segmentation* are two of the most common methods in breast image segmentation. Es-salhi et. al. (2016) stated that *Region-Based Segmentation* works by grouping pixels of identical intensity level into homogenous regions of interest.

Contour based methods or Edge-Based Segmentation has some of the common techniques that most researchers adopt during image processing. Gayathri & Raajan (2016) have stated that the idea of *Contour Based* or *Edge-Based Segmentation* is based on image discontinuities in grey levels, colour, or texture to separate discrete regions.

Clustering Methods can be further divided into several sub categories. The use of *Clustering Segmentation* is to perform grouping, in k clusters, for pixels having the same properties, according to the study of Muhammad et. al. (2020).

Based on the study of Huang, Luo & Zhang (2017), *Thresholding Segmentation* is able to identify ROI by using information from histogram, in order to segmentate objects from background. Thresholding produces a binary picture, which lowers data complexity and makes detection and classification easier (Dey et. al., 2018).

Justaniah & Alhothali (2021) describes *Model-Based Segmentation* as deformable models of either parametric or non-parametric class. This technique involves the process of assigning labels to pixels or voxels by comparing the picture input to a prior known object model.

Last but not least, Artificial Neural Network (ANN) Based Segmentation uses AI to automatically analyse and recognise picture elements such as objects, faces, text, handwritten writing, and so on (Prasad, 2021).

Table 2.5 Segmentation techniques implemented in breast image classification

Segmentation Technique	Category	Reference
Region-Based Segmentation	Region growing	(Es-salhi et. al., 2016)
	Split and Merge method	
	Watershed transform	
Contour-Based Segmentation	Roberts Edge Detection	(Muthukrishnan & Radha, 2011)
	Sobel Edge Detection	
	Prewitt Edge Detection	
	Kirsch Edge Detection	
	Robinson Edge Detection	
	Marr-Hildreth Edge Detection	
	Laplacian of Gaussian (LoG) Edge Detection	
	Canny Edge Detection	
Clustering Segmentation	Single Linear Iterative Clustering	(Muhammad et. al., 2020)
	K-Means	
	Fuzzy C-Mean Clustering	
	Expectation Maximization	
	Hierarchical Clustering	
	Mean-Shift Clustering	
	Normalized Cuts	
	Neutro-Connectedness	
Thresholding Segmentation	Global Thresholding	(Huang, Luo & Zhang, 2017; Dey et. al., 2018)
	Local Thresholding	
	Adaptive Thresholding	
Model-Based Segmentation	Markov Random Field	(Behera et. al., 2017;)
ANN-Based Segmentation	Feed Forward	(Elprocus, 2021)
	Feed Backward	
	Classification-Prediction	

Generally, the procedures involved processing the raw mammography images is as following:

1. Removal of Artifacts
2. Image Enhancement and Noise Removal
3. Removal of Pectoral Muscle
4. Segmentation of Image

Based on the selected dataset for this research (CIBS-DDSM), the ROI for breast calcified area for benign and malignant cases is provided. Hence, the process for removal of pectoral muscles and breast segmentation were not included in this study. This study will primarily focus on artifacts removal and image enhancement of the ROI mammography images before machine learning.

2.5 Machine Learning

2.5.1 Understanding Machine Learning

Carleo et. al. (2019) stated that machine learning can be understood as a large category of algorithms and modelling tools used for a wide range of data processing tasks, which recently has gained traction in many scientific areas. The machine learning technique may be further divided into Deep Network (DNN) or conventional (without DNN) (Refer Table 2.6). There are three methods on how a machine learns (Kotsiantis, Zaharakis & Pintelas, 2006):

1. Supervised
2. Unsupervised
3. Semi Supervised.

Table 2.6: Learning technique (Nahid & Kong, 2017)

Learning technique	**Algorithm**		
Supervised	Conventional	(a)Logic based	(1) ID3, (2) C4.5, (3) bagging,
			(4) random trees, (5) Random Forest,
			(6) boosting, (7) advanced boosting,
			(8) Extreme Boosting (XGBoosting).
		(b) Bayesian	(1) Naive Bayes
			(2) Bayesian Network
		(c) Conventional Neural Network	
		(d) Support Vector Machine	
	DNN-based	(a) Convolutional Neural Network (CNN),	
		(b) Deep Belief Network (DBN),	
		(c) Generative Adversarial Network (GAN).	
Unsupervised	Conventional	(a) *k*-Means Clustering	
		(b) Self-Organizing Map (SOP)	
		(c) Fuzzy *C*-Means Clustering (FCM)	
	DNN-based	(a) Deep Belief Network (DBN)	
Semi-supervised	Conventional	(a) Self-training	
		(b) Graph Based	
		(c) S3V3	
		(d) Multiview	
		(e) Generative model	

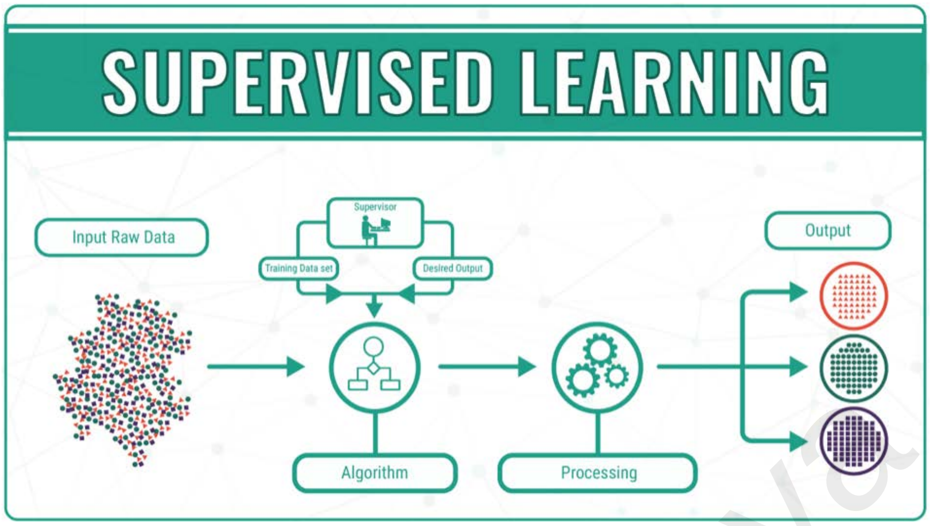


Figure 2.11: Workflow of supervised learning (Mezic, 2021)

Figure 2.11 illustrates the workflow for supervised learning. According to the study of Riese & Keller (2020), input output data pairs that are categorized can be used to train a supervised model. In supervised learning, the input dataset is segregated into training and testing datasets. The training dataset consists of an output variable that has to be predicted or categorised. All discovered features from the training dataset are applied to the testing dataset for prediction or classification.

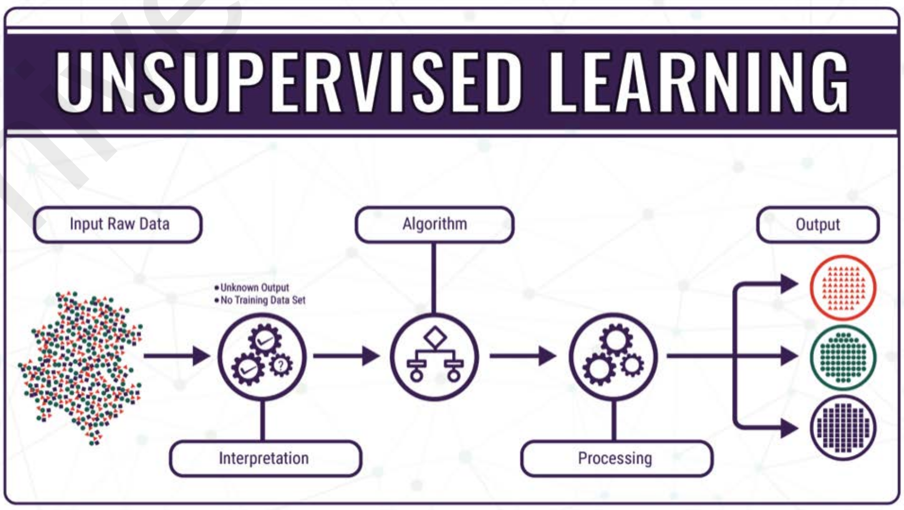


Figure 2.12: Workflow of unsupervised learning (Mezic, 2021)

Figure 2.12 illustrates the workflow for unsupervised learning. Unsupervised learning is a method of learning that employs data that has not been classified or labelled and allows the algorithm to run without the assistance of such info (Goel, 2018). Unsupervised learning does not have any accurate response or standards. When it comes to identifying and displaying the data's unique features, algorithms are left to their own to decide. Only a few characteristics are extracted from the data using unsupervised learning techniques. When new data is introduced, it uses previously learned characteristics to recognise the data's category.

Based on Bi et. al. (2019), the authors have stated that semi-supervised machine learning combines the benefits of supervised and unsupervised machine learning methods. Semi-supervised learning fits models using both labelled and unlabeled data. The model must learn and make predictions on new examples using a small number of labelled examples and a large number of unlabeled examples.

2.5.2 Deep Learning CNN - A Supervised Learning Technique

Recently, DL techniques, especially CNN, have gained lots of attention to CAD for mammography as they help to overcome CAD systems' limitations (Hedjazi et al., 2017). Thanks to advances in computational technologies, the introduction of digital mammography images have further improved the early detection of breast cancer using DL methods (Lee et al., 2017). Generally, DL is a subset of ML who is also a subset of AI (refer Figure 2.13).

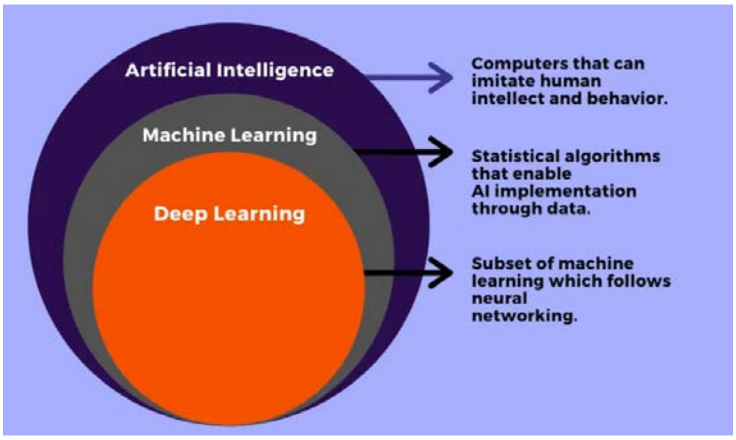


Figure 2.13: Relationship of AI, ML and DL (Barot, 2020)

Based on the explanation of Abdelhafiz et al. (2019), DL is a subset of machine learning. The term "deep" usually indicates the number of hidden layers in neural networks. For instance, ResNet 152 has a depth of 152 layers which is 8 × deeper than VGGNet 19. Papandrianos et al. (2020) reported that their architecture has fully connected layers, with each neuron connecting every other neuron to the next layer.

The use of CNNs in imaging processing is common. According to the study of Kooi et. al. (2016), CNNs is able to achieve higher detection accuracy compared to CAD models by delivering quantitative analysis of suspicious lesions. Research of Weiss, Khoshgoftaar & Wang (2016) have stated that training a deep CNN model with a limited number of medical data might decrease its accuracy, but this issue can be addressed by applying transfer learning (TL) and augmentation techniques, which will be further discussed in the Section 2.5.4 and Section 2.5.5.

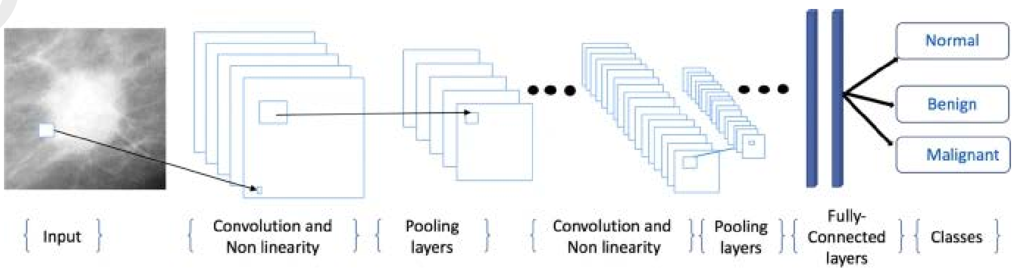


Figure 2.14 CNN architecture (Abdelhafiz et al., 2019)

The architecture of CNNs consists of four main layers, which are convolutional layer (Conv), nonlinear layer (e.g. ReLU), pooling layer (e.g. Max-pooling), and fully connected (FC) layer, which consists of a loss function such as Softmax or Support vector machines (SVM). The class output may be single (such as benign class, malignant class or etc) or a probability of classes. Based on the research of Saha (2019), the operation of convolution is based on the width (W1), height (H1) and number of channels (D1) of the input image. The model learns distinctive low-level and high-level features of the input image. For coloured images, D1 equals 3 (Red, Green and Blue), while for grayscale images, D1 equals 2 (Black and White). Breast mammogram images are categorized as grayscale images.

Yamashita et. al. (2018) have further added that the convolutional layer is fundamental as it forms activation maps for learning. This layer aids the algorithm to categorize new images based on the extracted features from previously recognized patterns. For each map, a non-linear activation function, such as rectified linear activation function (ReLU) is applied. According to Baeldung (2020), the use of ReLU prevents the computation required to run the neural network from growing exponentially. After applying ReLU, pooling is applied on the spatial dimension of the map, which is relatively critical in down-sampling the image and eliminating noise at the same time. Pooling operates with a specified threshold, keeping any pixel value that is higher than the threshold and discarding any value that is below. Next, is the fully connected layer. In the research of Papandrianos et al. (2020), this layer converts the previous output in vector format, and the algorithm labels the image accordingly.

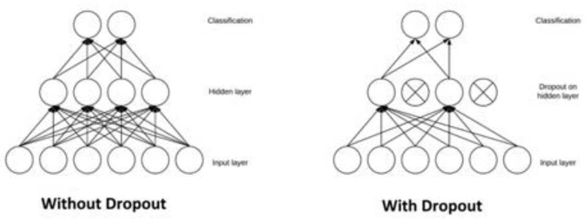


Figure 2.15: Adding dropout layer in CNN (Baeldung, 2020)

A Dropout layer is another common feature of CNNs. The Dropout layer is a mask that nullifies some neurons' contributions to the following layer while leaving all others unchanged. A Dropout layer can be applied to the input vector, however, it can also be applied to a hidden layer, nullifying certain hidden neurons. Dropout layers are crucial in CNN training because they prevent overfitting of data (Baeldung, 2020). Table 2.7 tabulates the configurations for some of the commonly available DL-CNN. There are various types of models for CNN available to the internet. By increasing the depth of the network, deep CNNs may learn more features (Al-Haija & Adebanjo, 2020). However, when the network depth grows, vanishing gradients and degradation becomes an issue. In 2015, Microsoft Research introduced Deep Residual Learning for Image Recognition (ResNet) (Dietz, 2017). ResNet has directly enriched the field of AI in developing DNN (Shorten, 2019). He et al. (2016) have added that this framework results in a much simpler network optimization. ResNet has even won first place on the task of ImageNet detection, ImageNet localization, and COCO competition.

Table 2.7 Configuration of DL-CNN (Abdelhafiz et. al., 2019)

	AlexNet	**ZF-Net**	**GoogLeNet**	**VGG-Net**	**ResNet**
Year	2012	2013	2014	2014	2015
Image Resolution	227 ×227	227 ×227	224 ×224	224 ×224	2244 ×224
Number of layers	8	8	22	19	152
Number of Conv-Pool layers	5	5	21	16	151
Number of FC layers	3	3	1	3	1
Fully connected layer size	4096,4096, 1000	4096,4096, 1000	1000	4096,4096 ,1000	1000
Filter Sizes	3, 5, 11	3, 5, 11	1,3,5,7	3	1,3,7
Number of Filters	96 - 384	96 - 384	64 - 384	64 - 512	64 - 2048

Table 2.7, Continued

	AlexNet	ZF-Net	GoogLeNet	VGG-Net	ResNet
Strides	1, 4	1, 4	1, 2	1	1, 2
Data Augmentation	+	+	+	+	+
Dropout	+	+	+	+	+
Batch Normalization	-	-	-	-	+
Number of GPU	2 GTX	1 GTX	A few high-end	4 Nvidia	
580 GPUs	580 GPUs	GPUs	Titan Black GPUs	Titan Black GPUs	8 GPUs
Training Time	5:6 days	12 days	1 week	2:3 weeks	2:3 weeks
Top-5 error	16.40%	11.2%	6.70%	7.30%	3.57%

2.5.3 ResNet

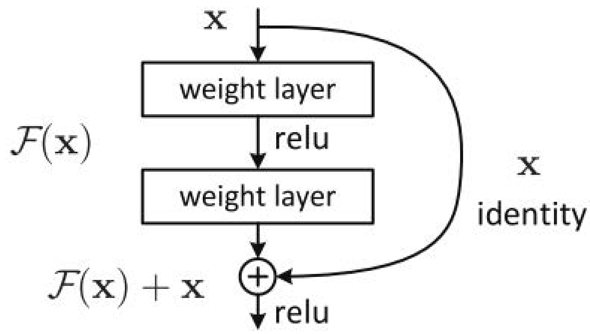


Figure 2.16: Residual Network Building Block

The ResNet model has many variants, of which the latest is ResNet152. ResNet introduces a "Residual block" that comes with a "skip connection", by which the output from the previous layer is added to the layer ahead (refer Figure 2.16). According to the study of Maladkar (2018), the precision declines and the error rate increases as the number of layers grows. But this issue can be tackled by implementation of identity mapping. The equation for identity mapping can be represented in Equation 2.11.

$$y = F(x, \{Wi\}) + x \qquad \text{(Equation 2.11)}$$

The input and output vectors of the layers of interest are denoted as x and y. F(x, {Wi}) is the function that denotes the residual mapping to learn. In the case where there are two layers (as in Figure 2.16) and ReLu is activated, the function is replaced with F = W2σ(W1x), where σ represents ReLU.

Although the study of Al-Haija & Adebanjo (2020) stated that degradation increases as network depth grows, results from the study of He et al. (2016) demonstrate that ResNet, even with far more layers than typical CNN, is able to maintain stability. According to the study of Purva (2020), another important aspect to notice is that the ResNet creators believe the more layers that are being stacked, the better the model will perform, which is basically identical to VGG16. However, instead of just stalking the layers on top of each other, ResNet modifies the underlying mapping. ResNet50 is also one of the most popular models available, with a top-5 error rate of about 5%. Figure 2.17 shows some of the ResNet variants.

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

Figure 2.17 :Variants of ResNet (Purva, 2020)

In this study, ResNet-50 (residual CNN with 50 layers deep) is the proposed model to classify calcified ROI images in order to aid early breast cancer diagnosis. However, other models such as VGG16, Alexnet, and Resnet34 will also be included in order to compare their performance in different parameters such as accuracy and error value.

2.5.4 Transfer Learning Module

Study of Bengio, Courville & Vincent (2013) state that feature learning (FL) is a collection of approaches in machine learning that allows a system to find the representations needed for feature detection, prediction, or classification from a preprocessed dataset automatically.

According to Ganin & Lempitsky (2015), in deep learning, feature learning may be achieved by either building a full convolutional neural network (CNN) from scratch or modifying a pretrained CNN in the classification/prediction for a new image set. The latter approach is known as transfer learning. Figure 2.18 depicts the concept of both approaches. Transfer learning uses the convolutional basis (green module in the picture) and re-train the classifier according to the dataset, as shown in the Figure 2.18 (pink module).

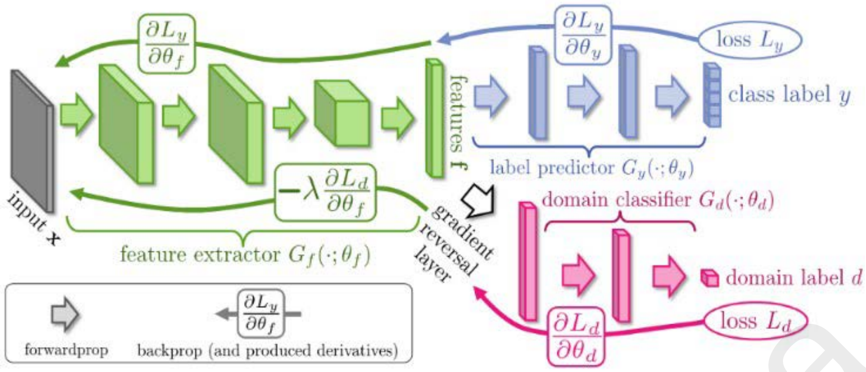


Figure 2.18: Illustration of no-transfer learning vs transfer learning CNN (Ganin & Lempitsky, 2015).

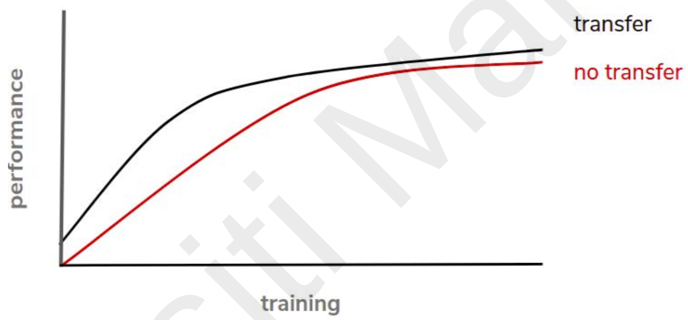


Figure 2.19: Training performance of no-transfer learning CNN vs transfer learning CNN (Lopes, 2018).

Transfer learning is a technique used often in Deep Learning (DL) applications to enable users to use a previously trained network to perform new prediction or classification tasks. The learning parameters of the employed pretrained network with randomly initialised weights must be fine-tuned to fit the new learning tasks. Transfer learning is usually faster and easier than learning or training the network from the scratch (Lopes, 2018). Figure 2.20 compares the training performance of CNNs that use transfer learning with CNNs that do not use transfer learning.

Table 2. Performance of deep neural networks using the from-scratch training scenario.

CNN	DDSM-400		CBIS-DDSM	
	AUC	ACC	AUC	ACC
AlexNet	0.657	0.610	0.716	0.656
VGG-16	0.621	0.590	0.702	0.580
VGG-19	0.644	0.588	0.707	0.581
ResNet-50	0.595	0.548	0.637	0.627
ResNet-101	0.637	0.588	0.641	0.662
ResNet-152	0.596	0.543	0.609	0.647
GoogLeNet	0.580	0.569	0.590	0.598
Inception-BN (v2)	0.652	0.590	0.577	0.654

Table 3. Performance of convolutional neural networks (CNNs) initialized on pre-trained weights (fine-tuning).

CNN	DDSM-400		CBIS-DDSM	
	AUC	ACC	AUC	ACC
AlexNet	0.805	0.733	0.802	0.753
VGG-16	0.844	0.748	0.781	0.716
VGG-19	0.835	0.738	0.783	0.736
ResNet-50	0.856	0.743	0.804	0.749
ResNet-101	0.859	0.785	0.791	0.753
ResNet-152	0.786	0.630	0.793	0.755
GoogLeNet	0.830	0.758	0.767	0.720
Inception-BN (v2)	0.850	0.780	0.774	0.747

Figure 2.20: Comparison between training NN from scratch and pretrained weights (Tsochatzidis, Costaridou, & Pratikakis, 2019)

In the study of Tsochatzidis, Costaridou, & Pratikakis (2019), a comparison between pre-trained and untrained models in terms of their performance in classifying breast mass lesions into benign or malignant cases has been performed. The authors have stated that a training machine from pre-trained weights such as Imagenet is able to produce much higher accuracy.

In this research, transfer learning method is utilized to retrain the readily made available model. The pretrained models such as Alexnet, Resnet, and VGG are downloaded from fastai library in Google Colaboratory platform. Fine tuning of the model's hyperparameters will be conducted to further enhance the performance of the model.

2.5.5 Hyper-Parameters

In the study of Wu, Perin & Picek (2020), the authors stated that all configuration variables outside of the model, such as the number of hidden layers in a neural network, are called hyperparameters. Hyperparameters are chosen manually prior to CNN training.

According to Brownlee (2019), one of the most significant hyper-parameters that impacts the performance of CNNs is the learning rate (LR). A stochastic gradient descent optimizer such as Adam, RMS Prop, Adagrad, and more are usually utilized to train the model. The learning rate determines how much network parameters are changed to reduce the network's loss function.

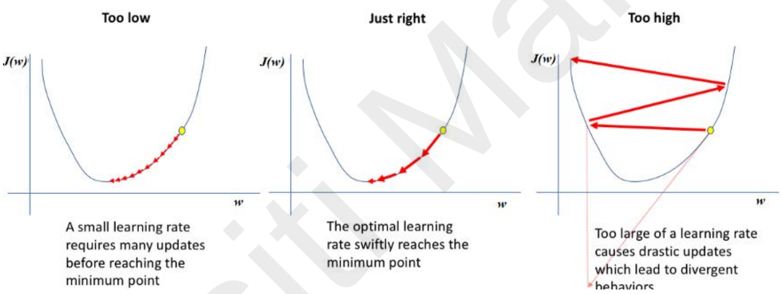


Figure 2.21: Learning rate (Nabi, 2019)

Based on Nabi (2019), training will be more reliable if the learning rate is low, but it will take a long time since steps towards the loss function's minimum are small. Training may not converge or even diverge if the learning rate is excessive. Weight fluctuations might be so large that the optimizer skips the minimum and exacerbates the loss.

More abstract shapes may frequently be built in terms of less abstract shapes recorded in earlier layers, thus deeper structures (larger number of hidden layers) can lead to better abstract representations. Adding more layers to the model will help it extract more features. Increasing the number of hidden layers improves the accuracy of huge data sets. However, there is a limit on how many levels may be added. Instead of extracting features, the network might be overfitted, which might lead to false positives.

Adding layers to a CNN needlessly increases the number of parameters, but may reduce the accuracy of the test data for a smaller data set. Deep architectures are notoriously difficult to train properly. The choice of a smaller or bigger network cannot be predicted mathematically. On the basis of the dataset, a trade-off between accuracy and deep networks must be made using a trial-and-error technique as well as some expertise and practice.

Tuning the batch size and the number of epochs is also important in achieving higher accuracy. Steward (2019) explained that the batch size determines how many samples will be sent across the network. For instance, if there are 2000 images and the batch size is set to be 200, the network takes the first 200 images on the first training and the next 200 on the second training. The number of epochs is a hyperparameter that determines how often the learning algorithm runs across the whole training dataset.

2.6 Performance Measuring Parameter

When it comes to evaluating the performance of the model, a confusion matrix is commonly utilized in deep learning. Brownlee (2020) has stated that the confusion matrix is a 2D table used to visually represent classification results, where (i, j)th position of the confusion table indicates the number of times that the ith object is classified as the jth object. The matrix is able to indicate the number of times the objects are correctly or wrongly classified.

	Hypothesized class	
True class	True positive (A)	False negative (B)
	False positive (C)	True negative (D)

Figure 2.22: Confusion matrix

Figure 2.22 depicts a graphical illustration of a Confusion Matrix. There are four main parameters that are presented in a confusion matrix, which are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) (Markham, 2021).

Table 2.8: Parameters of Confusion Matrix (Markham, 2021)

Parameters	Description
True Positive (TP)	Cases that are predicted positive by the machine, and they are truly positive.
True Negative (TN)	Cases that are predicted negative by the machine, and they are truly negative.
False Positive (FP)	Known as "Type I" error. Cases that are predicted positive by the machine, but in fact they are negative.
False Negative (FN)	Known as "Type II" error. Cases that are predicted negative by the machine, but in fact are positive.

According to Alkhaleefah et al. (2020), based on the parameters that are provided by the confusion matrix, there are several new parameters that can be deduced.

Table 2.9: Additional parameters deduced by confusion matrix (Alkhaleefah et al.,2020)

Parameters	Equation Involved	Description
Recall	$\frac{TP}{TP + FN}$	Recall deals with the division of corrected positive images to all corrected images of both categories.
Precision	$\frac{TP}{TP + FP}$	Precision points out the ratio of corrected images of positive class to all images that were predicted for the same category.
Specificity	$\frac{TN}{TN + FP}$	The truly negative rate. In this paper, TN corresponds to Benign cases. It computes the proportion of actual negative cases that are predicted as negative cases.
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Represents the number of correctly predicted cases over all cases.
F-1 score	$\frac{2 \times Recall}{2 \times Recall + FP + FN}$	F1-score is a harmonic mean between recall and precision.

Matthew Correlation Coefficient (MCC) is also one of the additional parameters that can be deduced from the details provided in the confusion matrix. MCC measures the performance of the parameters in the confusion matrix. The classifier produces a more accurate classifier if the MCC values trend more towards +1, and the reverse situation occurs if the MCC values trend more towards 1. Equation of MCC can be found in Equation 2.12.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \qquad \text{(Equation 2.12)}$$

According to Papandrianos et. al. (2020), a confusion matrix is commonly used in machine-learning applications because it provides an effective approach to observe the performance of an algorithm, such as binary image classification. It is a helpful assessment metric for calculating the classifier's errors and offers the exact number of images categorised with the incorrect label. This study aims to deliver an optimal confusion matrix of several CNN architectures with the least classification errors.

2.7 Summary

Detecting malignant cases of microcalcification on breast mammography images accurately is important as it allows the patient to seek treatment as soon as possible. To improve the accuracy of detection, it is important to preprocess the images in order to increase its quality prior training. Models that were built for breast image classification of benign and malignant cases all have the same goal, which is to achieve the best possible accuracy in detection. To achieve this, techniques such as data augmentation, transfer learning, and tuning of hyperparameters were all utilized.

In order to determine which model performs best, a confusion matrix is normally applied on the model to test their accuracy. Over the years, there have been several models that were developed in order to classify breast calcified images. Usually, the model will be based on either histopathological images or segmented mammography images.

This study incorporates the use of transfer learning in order to classify breast microcalcification into benign and malignant cases by using CIBS-DDSM database. The pretrained models that will be put into test involve VGG16, Resnet34, Resnet50, and Alexnet. The procedure for utilizing and training these models will be further discussed in Chapter 3.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This section of the paper explains how a binary classifier is being used to distinguish between benign and malignant cases of calcified breast mammography images. The proposed deep learning model will be trained and assessed using Google Colab's platform with OpenCV library of programming functions.

The chapter begins with a brief summary of the proposed work that will be conducted in this research in section 3.2. The workflow of the proposed work is depicted in Figure 3.1. Data acquisition is performed whereby the breast mammography images with microcalcification are downloaded from the TCIA database (refer section 3.4). Microcalcified images of the breast mammography were categorized into benign and malignant cases based on the information given in the *.csv files from TCIA. Moving on, the images that were downloaded were addressed with preprocessing techniques such as artifacts removal and noise removal (refer section 3.4.2). Later, the images were fed into a pretrained network for training and different data augmentation techniques as well as tuning of hyperparameters were included to obtain the best possible result from the models trained.

This chapter ends with section 3.6 which details the type of tests that will be performed.

3.2 Proposed work

First and foremost, the dataset will be downloaded and rescaled into a specific size. Later, the images will be preprocessed to remove noise and artifacts. Moving on, they will be split into a training set and validation set, which each consist of benign and malignant subfolders containing respective ROI images. To categorise the pictures into benign and malignant cases, a CNN model is utilised as a baseline. Transfer learning is used instead of training CNN from scratch. As such, different CNN models pre-trained with torchvision from fastai library will be transferred to conduct the classification. In order to get the best possible result, hyperparameters such as the number of extra layers, learning rate, batch size, and epochs will be tuned. Finally, the confusion matrix will be utilized to assess the performance of the model to get the best possible accuracy. Figure 3.1 shows the proposed workflow of the model.

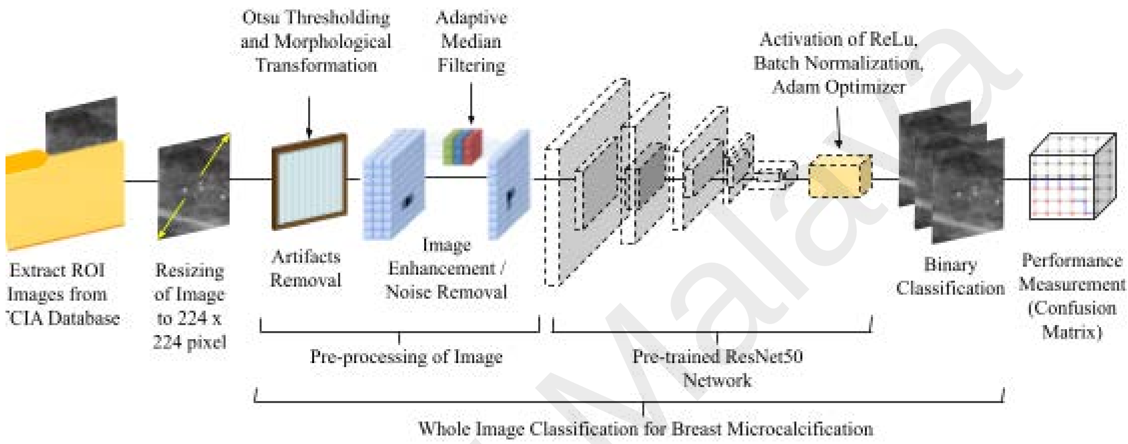


Figure 3.1: Workflow of proposed design

3.3 Materials

The following are the materials needed for the work of this research:

1. Laptop - Intel Core i7-4710 HQ, 3.5 GHz, 1 TB SSD, 4 GB RAM
2. Google Colaboratory Platform (Python OpenCV language and Fastai Library)
3. Breast Image dataset CIBS-DDSM from TCIA

3.4 Preparing Dataset

3.4.1 Dataset

The CIBS-DDSM dataset of the images for this research will be obtained from Cancer Imaging Archive (TCIA) (Smith, 2016). CIBS-DDSM dataset will be utilized in this research because it has higher resolution, larger amount of images available, and availability of calcified breast ROI images. The dataset consists of 1077 benign and 577 malignant ROI images in various sizes in DICOM format. To download radiological pictures from the TCIA Radiology Portal, NBIA Data Retriever needs to be installed. The image will be fed into DICOM to be saved as *.jpeg format with size of 224 x 224.

In order to improve the accuracy for deep learning, the total number of images for benign and malignant cases were multiplied. The ROI images of benign and malignant cases are rotated and further added into the main dataset. The following Table 3.1 illustrates the dataset distribution for the research.

Table 3.1: Dataset distribution

Image	Calcified Benign ROI	Calcified Malignant ROI
Original Image	1077	577
Rotated at 90 degrees	1077	577
Rotated at 180 degrees	1077	577
Rotated at 270 degrees	1077	577
Total Number of Images	4958	1653

3.4.2 Preprocessing

Prior to training CNNs, the images will be preprocessed to remove the artifacts and improve the contrast by removing noise. Otsu Segmentation Method which is presented by Xi, Shu & Goubran (2018) will be utilized to remove the artifacts that may be present at the image. There are three types of filters included in this test, which are adaptive median filter, median filter and mean filter. The filters are utilized in order to remove image noise. The best filter will be selected based on the highest PSNR and lowest MSE value.

3.5 Proposed Model

3.5.1 Deep CNNs Architecture

This research utilizes supervised learning Deep CNNs Architecture. A pretrained model from torchvision will be utilized from fastai library that is made available at the Google Colab platform.

Prior training, the data will be divided into training and testing sets at random. The *valid_pct* function splits the dataset into training and testing sets at a particular ratio of 0.80 testing and 0.20 validation.

Data augmentation technique is used on the training set to avoid over-fitting. *get_transforms* function is used to increase the volume of the dataset by artificially producing new training data from the current data. The parameters for data augmentation utilized for this research are depicted in Table 3.2.

Table 3.2: Parameters of data augmentation

Parameter	Function	Description
Flipping	do_flip () flip_vert ()	Randomly flips the images at the vertical and horizontal axis
Zooming	max_zoom()	Randomly zooms the image at a certain scale
Rotating	max_rotate()	Randomly rotates the images at a certain degree
Lighting	max_lighting () p_lighting ()	If not None, a random lightning and contrast change controlled by max_lighting is applied with probability p_lighting

fastai libraries from Google Colab were utilized in order to include the pretrained models from torchvision for transfer learning. The pretrained network was downloaded from the fastai library using create_cnn().

```
learn = create_cnn(data, models.resnet50, metrics=[error_rate,accuracy], model_dir = '/content/working')

/usr/local/lib/python3.7/dist-packages/fastai/vision/learner.py:109: UserWarning: `create_cnn` is deprecate
  warn("`create_cnn` is deprecated and is now named `cnn_learner`.")
Downloading: "https://download.pytorch.org/models/resnet50-0676ba61.pth" to /root/.cache/torch/hub/checkpoi
100%                                        97.8M/97.8M [00:00<00:00, 114MB/s]
```

Figure 3.2: Example of importing Resnet50 model from fastai library

Resnet50, Resnet34, VGG16 and Alexnet trained networks were utilized in this research. Figure 3.2 shows an example of importing the Resnet50 model from fastai library using create_cnn().

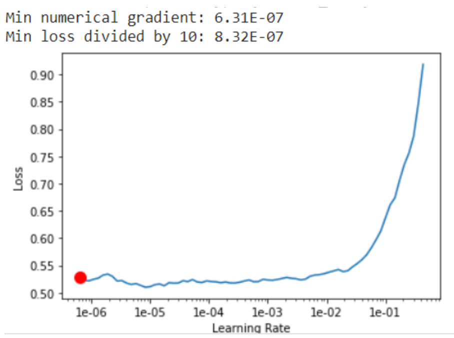


Figure 3.3: Learning curve plotted using learn.recorder.plot(). Red dot shows the minimum gradient of the learning curve.

After importing the models, the first layer of the model was trained by using learn.fit_one_cycle(). Later, the learning rate for the model was determined with the aid of learn.lr_find() and learn.recorder.plot(), which illustrates the learning curve of the model after training the first layer and suggests the lowest gradient of the learning curve in red dot (refer Figure 3.3). The learning rate is determined manually on the steepest negative gradient just before the loss starts to increase.

```
learn.unfreeze()
learn.fit_one_cycle(5, max_lr=slice(2e-6,1e-5))
```

epoch	train_loss	valid_loss	error_rate	accuracy	time
0	0.516951	0.482689	0.270045	0.729955	02:04
1	0.507811	0.477767	0.272315	0.727685	02:01
2	0.497787	0.469763	0.258699	0.741301	02:01
3	0.489055	0.465715	0.255673	0.744327	02:01
4	0.492780	0.468119	0.264750	0.735250	02:01

Figure 3.4: Example of code section for learn.fit_one_cycle()

Moving on, all layers of the model were unfreezed using learn.unfreeze() to allow more parameters to be trainable. The model undergoes training again with learn.fit_one_cycle(), but with restraination on a cyclic learning rate using max_lr(). The first parameter for learn.fit_one_cycle() indicates the epoch to be executed in the run (refer Figure 3.4).

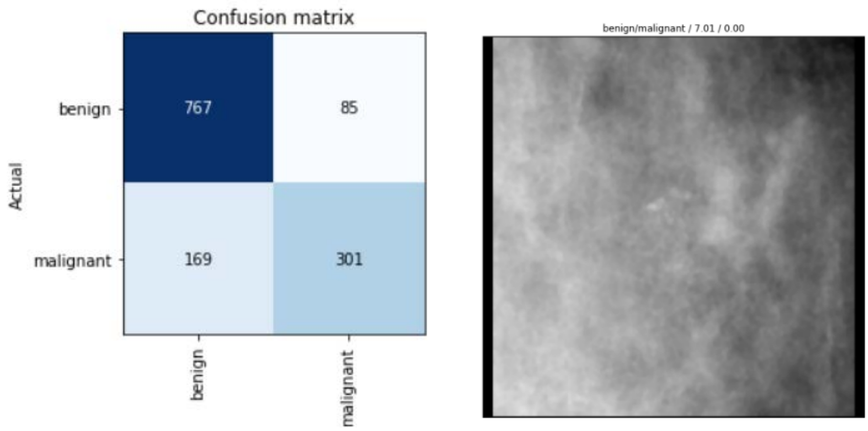


Figure 3.5: Example of confusion matrix (left) and top loss image (right)

Upon running the number of epochs determined, the confusion matrix of the model on the validation set was plotted using interp.plot_confusion_matrix(). The top losses of images during training were also plotted using interp.plot_top_losses() with labels of "Prediction/Actual/Loss/Probability". By the end of the training, the value for training loss, validation loss, error rate and accuracy were recorded.

Before training the CNNs, hyper-parameters are chosen manually. The purpose of tuning the hyperparameters is to identify the best possible parameters that are able to deliver the best possible accuracy on binary classification. Table 3.3 describes the different hyperparameters that will be tuned in this research.

Table 3.3: Tuning of hyperparameters

Parameters	Description
Number of layers	Various layers such as convolution, pooling, or dropout layers may be added in order to enhance the model to achieve a better accuracy. For instance, ResNet 34 will have 34 layers deep while ResNet 50 will have 50 layers deep.
Learning Rate	Training will be more reliable if the learning rate is low, but this takes time. Learning rate that is too high will cause the training to diverge.
Batch Size	Refers to the amount of training examples used in one iteration.
Epoch	Determines how often the learning algorithm runs across the whole training dataset.

According to the pre-trained CNN model from fastai library, the optimization algorithm using Adam is included to enhance the effectiveness of the model. In addition to that, ReLu is activated to prevent the computation required to run the neural network from growing exponentially. Batch Normalization is activated to enable each layer of the network to conduct learning more independently by re-centering and re-scaling the layers' inputs to improve the speed and stability of the network.

3.5.2 Performance Measurement

When it comes to evaluating the performance of the model, a confusion matrix is utilized. The values obtained from the confusion matrix will be further analyzed to compute additional parameters such as Recall, Precision, Specificity, Accuracy, F-1 score and MCC, as explained in Chapter 2 (section 2.6).

3.6 Summary

ResNet50 is the main proposed model that will be used in this research to classify calcified breast images into benign and malignant cases. Different layers of ResNet (e.g. ResNet 38) will also be included in the test for comparison in terms of accuracy. We will also be looking at the output from VGG-16 and Alexnet to compare with the output of ResNet 50. Table 3.4 illustrates the proposed tests that will be performed in this research. The output of the tests will be further discussed in Chapter 4. Chapter 4 will also incorporate comparison of existing work with the result from this study.

Table 3.4: Proposed tests

CNN Model	Batch Size	Learning Rate	Epoch	Image Size	ReLU	Batch Normalization
VGG16	32	8e-6,1e-4	15	224 x 224	Active	Active
VGG16	64	8e-6,1e-4	15	224 x 224	Active	Active
VGG16	64	8e-6,1e-4	30	224 x 224	Active	Active
VGG16	32	2e-6,1e-3	15	224 x 224	Active	Active
VGG16	64	2e-6,1e-3	15	224 x 224	Active	Active
VGG16	64	2e-6,1e-3	30	224 x 224	Active	Active
Resnet34	32	8e-6,1e-4	15	224 x 224	Active	Active
Resnet34	64	8e-6,1e-4	15	224 x 224	Active	Active
Resnet34	64	8e-6,1e-4	30	224 x 224	Active	Active
Resnet34	32	2e-6,1e-3	15	224 x 224	Active	Active
Resnet34	64	2e-6,1e-3	15	224 x 224	Active	Active
Resnet34	64	2e-6,1e-3	30	224 x 224	Active	Active
Alexnet	32	8e-6,1e-4	15	224 x 224	Active	Active
Alexnet	64	8e-6,1e-4	15	224 x 224	Active	Active
Alexnet	64	8e-6,1e-4	30	224 x 224	Active	Active
Alexnet	32	2e-6,1e-3	15	224 x 224	Active	Active
Alexnet	64	2e-6,1e-3	15	224 x 224	Active	Active
Alexnet	64	2e-6,1e-3	30	224 x 224	Active	Active
Resnet50	32	8e-6,1e-4	15	224 x 224	Active	Active
Resnet50	64	8e-6,1e-4	15	224 x 224	Active	Active
Resnet50	64	8e-6,1e-4	30	224 x 224	Active	Active
Resnet50	32	2e-6,1e-3	15	224 x 224	Active	Active
Resnet50	64	2e-6,1e-3	15	224 x 224	Active	Active
Resnet50	64	2e-6,1e-3	30	224 x 224	Active	Active

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

In chapter 3, the proposed work to be conducted in this research has been presented. In this chapter, the result of the proposed work will be presented and the result will be compared with some of the existing techniques as presented in Chapter 2.

The details of data acquisition and resizing of images will be described in Section 4.2. Section 4.3 will discuss how the images have undergone artifacts removal while section 4.4 will provide the result from the application of adaptive median filter, median filter as well as mean filter. The result of the trained models will be compared in terms of accuracy, error rate, training loss and validation loss (refer Table 4.2 in Section 4.5). In order to measure the performance of the trained models, a confusion matrix was utilized and their values are used to perform calculations on additional measuring parameters such as F1-score and MCC values. The summary of this chapter will be presented in Section 4.6.

4.2 Dataset

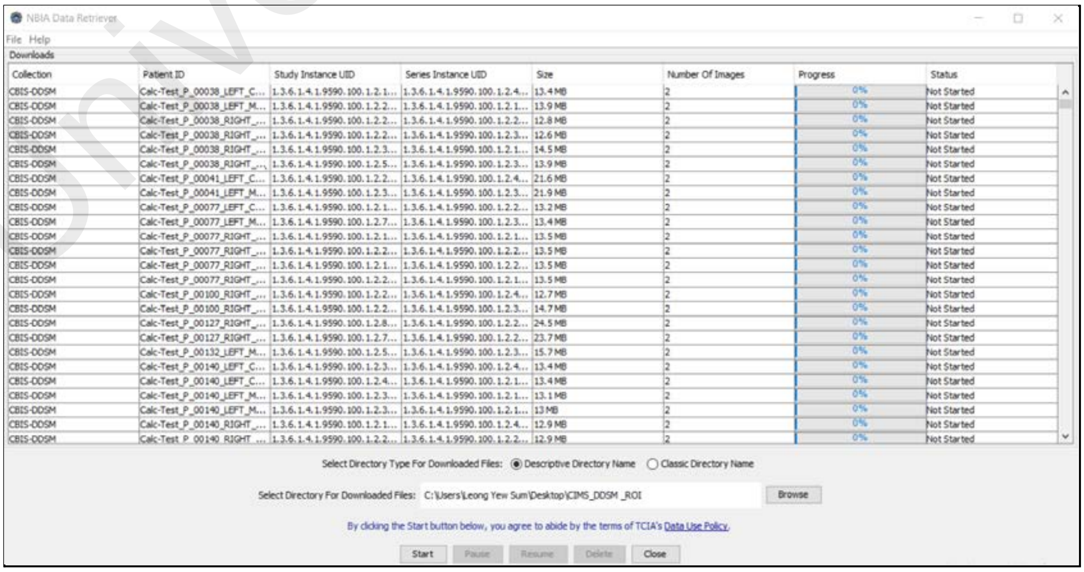


Figure 4.1: *.tcia files selected

A total of 1078 calcified benign ROI and 577 calcified malignant ROI images (26.1 GB) were selected from TCIA Radiology Portal. NBIA Data Retriever software was utilized in order to download all the breast mammogram images that were selected (refer Figure 4.1).

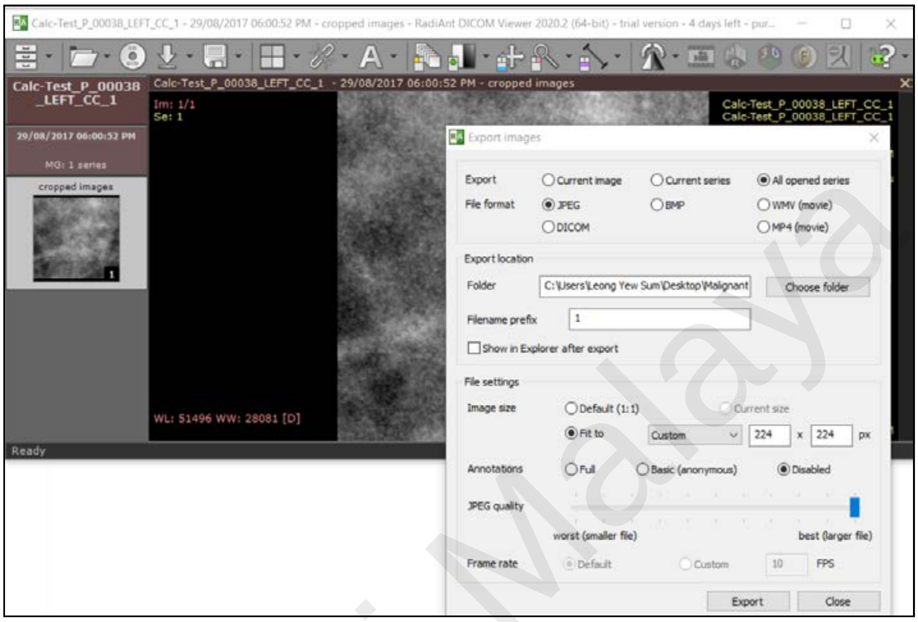


Figure 4.2: DICOM software

The downloaded images are in *.dcm format and of different sizes. In order to match the format of the image to the image processing algorithms in Google Colaboratory platform, all the images are converted into *.jpeg format. In order to achieve this, Digital Imaging and Communications in Medicine (DICOM) software was utilized. All of the images extracted were set to be in *.jpeg format. In addition to that, the images were resized into 224 x 224 pixels to facilitate training of the model.

4.3 Removal of Artifacts

Wedges and labels in the raw mammography picture may cause needless disruptions during the mass detection procedure (Boss, Thangavel, & Daniel, 2013). Hence, they need to be removed before any training is performed. Before any preprocessing work was performed, the notebook on Google Collab was set to be under GPU Runtime to allow faster computational work. Based on manual observation, the ROI images of breast calcification downloaded from the TCIA database were found to be free from labelling of name or numbers. However, removal of artifacts were still performed just in case there are hidden or unobvious artifacts. Removal of artifacts on the breast mammogram images were performed on Google Colab Platform using OpenCV programming library. Techniques presented by Xi, Shu & Goubran (2018) were implemented, which involves Otsu Segmentation Method and MorphologyEx Method under Morphological Transformation. Appendix A depicts the step-by-step method on removal of artifacts. In order to ensure that this section of coding works properly, full breast mammogram images with obvious artifacts were tested prior implementation of the coding to the images that will be utilized for this research.

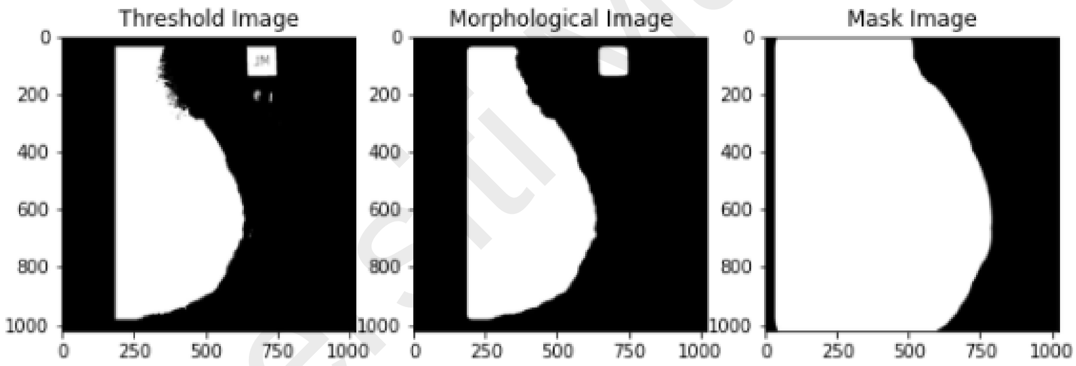


Figure 4.3: Result under coding of artifacts removal

Otsu Thresholding was applied to set the threshold of the artifacts to be 255(white), using cv2.threshold(). Small areas within the white region were enclosed by using cv2.morphologyEx() under MorphologyEx Method. Later, regions with the largest contour (breast) were identified using cv2.findContours(). cv2.morphologyEx() were utilized again to make a mask image out of the largest contour, and the background besides the largest contour were turned black. Lastly, the original image and the mask image were compared bit-by-bit using cv2.bitwise_and(), where the regions which appear black on the mask are applied to the original image.

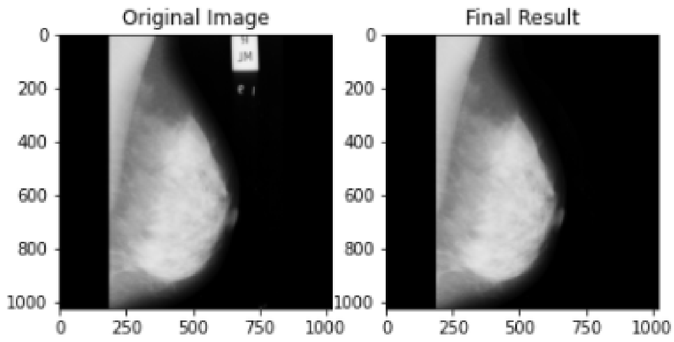


Figure 4.4: Comparison on application of artifacts removal

Figure 4.4 shows the comparison of the breast image with artifacts before and after the application of the techniques based on Xi, Shu & Goubran (2018), in which the artifacts were completely removed.

4.4 Image Enhancement

The major forms of noises that influence mammography pictures include salt and pepper noise, speckle noise, gaussian noise, and poisson noise. These disturbances cause issues with fine analysis and proper interpretation of the breast image produced from a mammographic exam, resulting in incorrect diagnosis (Joseph, John & Dhas, 2017). In this research, three types of filters, namely adaptive median filter, mean filter, and median filter were applied on the same image and the MSE and PSNR value for each filter was compared to select the best filter for image enhancement.

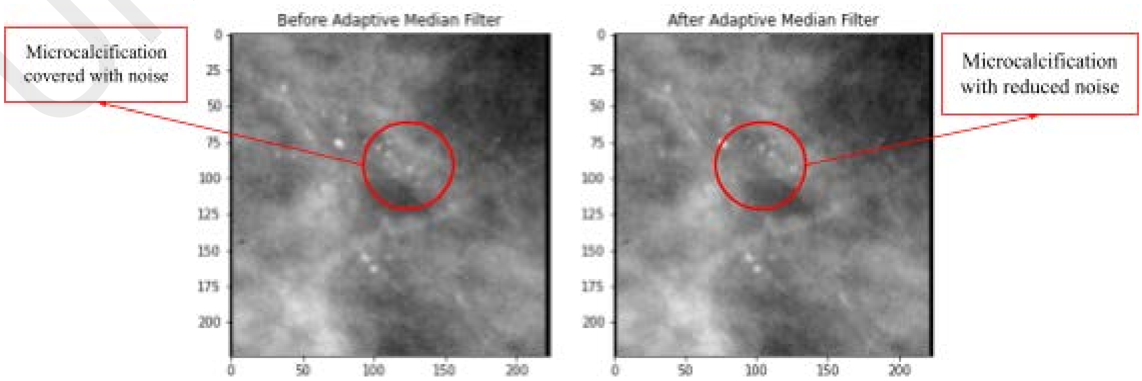


Figure 4.5: Comparison on application of adaptive median filter

Adaptive median filter was applied on the ROI images in order to eliminate noise. Appendix B depicts the step-by-step coding for application of adaptive median filter. Figure 4.5 shows the before and after images of applying adaptive median filter. As we can see, the image appears much smoother after applying adaptive median filter.

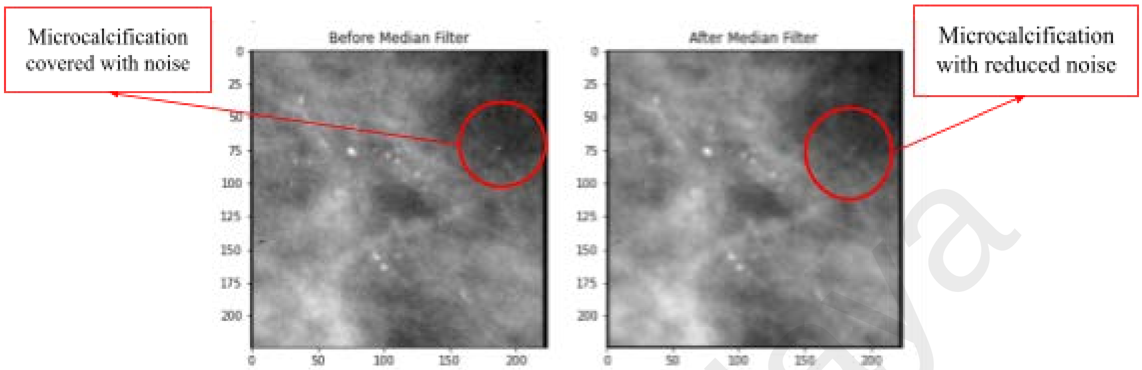


Figure 4.6: Comparison on application of median filter

Appendix C depicts the step-by-step coding for application of median filter. For every 3x3 area of the image, the value of the pixel is replaced by the median value. The result for application of the median filter can be seen in Figure 4.6, where the noise of the image was significantly lower.

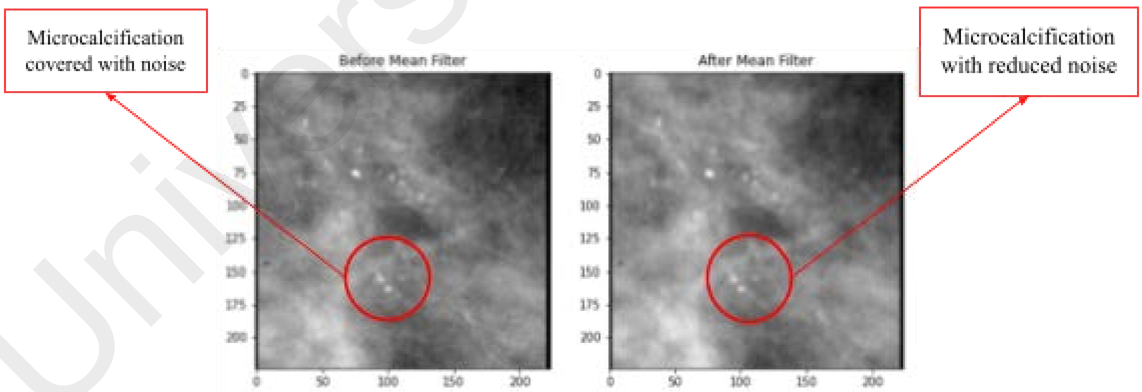


Figure 4.7: Comparison on application of mean filter

Appendix D depicts the step-by-step coding for application of mean filter. The application of the mean filter utilizes cv2.filter2D() function in OpenCV to perform the linear filtering operation. Figure 4.7 shows the comparison on the before and after images of mean filter application.

In order to compare the effectiveness of the filters, the values for Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) were calculated. Appendix E shows the step-by-step coding for obtaining PSNR and MSE value from a filtered image. Table 4.1 below shows the PSNR and MSE value for all three filters.

Table 4.1: PSNR and MSE value for filter

Parameter	Adaptive Median Filter	Median Filter	Mean Filter
PSNR	42.3863	37.5911	36.9511
MSE	3.7536	11.3233	13.1213

Based on the result obtained from Table 4.1, the value for MSE is lowest for adaptive median filter. This indicates that the error difference between the original image's values and the degraded image's values for adaptive median filter is the least among all three types of filters. Hence, it is utilized in this study in order to enhance the images by removing image noises before feeding them into the CNN models for training.

The comparison for adaptive median, mean and median filter for breast mammogram images were also reported in the study of Chashmi & Chehelamirani (2019). Table 4.2 depicts the PSNR and MSE result obtained for three types of filter based on several MIAS databases and the authors concluded that adaptive median filter is the best filter for noise reduction.

Table 4.2: MSE and PSNR values for samples of image (Cashmi & Chehelamirani, 2019)

Image No	Median Filter		Mean Filter		Adaptive Median Filter	
	MSE	PSNR	MSE	PSNR	MSE	PSNR
Mdb001	65.8468	30.5837	30.8829	33.2336	6.7584	39.8323
Mdb002	63.1807	30.125	41.04	31.9987	16.4629	35.9657
Mdb003	52.2811	30.9474	38.339	32.2944	15.9076	36.1147

Based on the result from Table 4.2, the PSNR value for median filter is quite close to mean filter, and the value for adaptive median filter is definitely significantly higher, which is similar to the result obtained from this research (refer Table 4.1). Based on the result obtained in Table 4.1, the PSNR value for the image that we have selected is

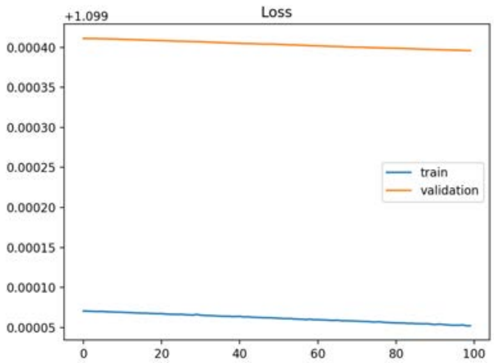
42.3863, which is much higher compared to all PSNR values for three images from Table 4.2. This is probably because of the dataset of images that were put into use. For instance, the dataset that is used in this research is from CIBS-DDSM, which possesses a much higher resolution as compared to the image from MIAS dataset (refer Table 2.3).

In the study of Ramani, Vanitha, & Valarmathy (2013), comparison between adaptive median filter, median filter, mean filter and Wiener filter has been made and the author has also agreed that adaptive median filter is the most suited approach when compared to other filters, since adaptive median filter image quality is superior compared to other filters.

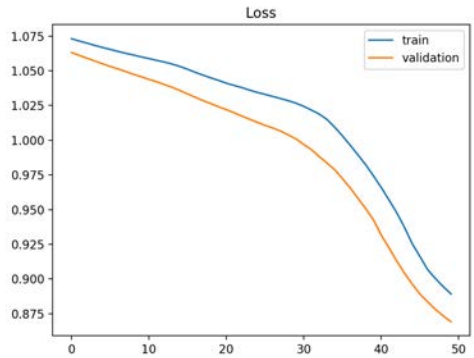
4.5 CNN Model Architecture

In the training period, underfitting and overfitting are two main problems that usually occur. According to Zhang et. al. (2019), underfitting indicates that the training for neural networks is insufficient and learning precision is low, whereas overfitting indicates that learnt mappings function well for training data but not for test data, implying a lack of flexibility and adaptability.

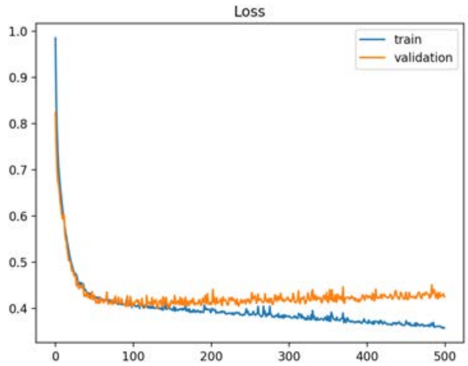
Study of Chen, Liu and Peng (2019) stated that model underfitting and overfitting needs to be fundamentally balanced. Underfitting can be avoided by training with a more parameter-rich model or with weaker regularisation. Overfitting is particularly worrisome since the evaluation overestimates the performance of generalisation on previously unknown data. Bagui et al. (2017) has also agreed that in comparison to the underfitting issue, it has gotten a lot more attention. Feature selection, hyper-parameter tuning, ensemble models, and regularisation are all methods for limiting overfitting issues.



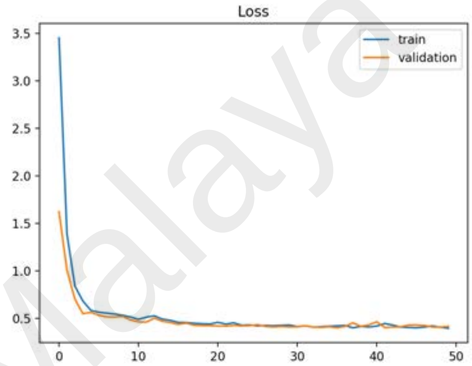
(a)



(b)



(c)



(d)

Figure 4.8: Example of train and validation learning curves (Brownlee, 2019)

In order to identify whether the model is underfitting or overfitting, the training and validation loss curves are observed. During training, a loss curve is one of the most commonly used graphs to debug a neural network. It provides an overview of the training process as well as the network's learning trajectory. During the forward pass, the loss is calculated for each batch.

As stated by Brownlee (2019), a good fit is defined as a training and validation loss that gradually reduces to a point of stability with a small gap between the two final loss values. If the plot of training loss reduces to a point of stability and the plot of validation loss declines to a point of stability and has a small gap with the training loss, the plot of learning curves indicates a good match. Figure 4.9 depicts examples of underfitting curve (a), underfitting curve that requires some additional training (b), overfitting curve (c) and a good fit curve (d). However, if training of a good fit continues, it will almost certainly result in overfit.

Study of Afaq & Rao (2020) reported that during the training phase, when the model is learning both the training set and the validation set. Hence, both the training and validation error will continue to decrease. This indicates that the learning process is proceeding smoothly. However, as soon as the validation error increases, the training process should be terminated. This is because the loss begins to escalate, it is at this moment where overfitting begins to develop.

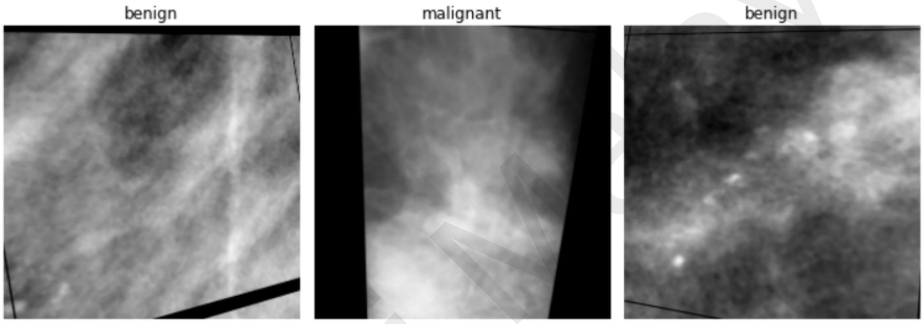


Figure 4.9: Output of data augmentation

Because of its ease of use and capacity to 'inject' prior knowledge into the training process, data augmentation is one of the popular techniques that is particularly effective in data-limited situations to avoid overfitting. In this study, data augmentation technique is applied to the images under get_transforms() to avoid overfitting. Figure 4.8 depicts the result of the images after application of data augmentation technique which includes flipping, zooming, rotating and lighting.

In order to import a pretrained model from fastai library, create_cnn() was utilized and the pretrained network was called using specific parameters. For instance, VGG16 was called using vgg16_bn, Resnet34 and Resnet50 were called using resnet34 and resnet50 respectively and Alexnet was called using alexnet.

valid_pct () splits the dataset into training and testing sets at a particular ratio of 0.80 testing and 0.20 validation. In total, there are 5288 training images and 1322 validation images. A confusion matrix is generated based on the validation images. In actual practice, it is very important to keep the FN value as low as possible. This is to identify as many malignant cases as possible to avoid misdiagnosis of malignant breast microcalcification.

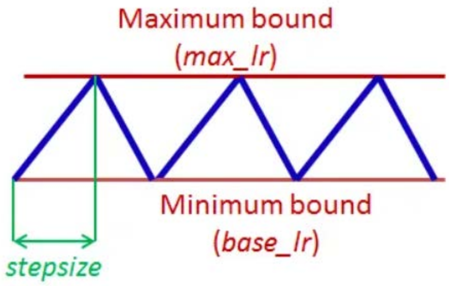


Figure 4.10: Cylindrical learning rate (CLR) method from fastai Library

During training, CLR (learn_fit_one_cycle()) was used to train the model in a specific learning rate border. This technique is preferred over the fit() method since it performs better in terms of speed and accuracy (Mavropalias, 2019). Rather than utilising a constant or decreasing learning rate, the CLR approach enables the learning rate to fluctuate between appropriate minimum and maximum boundaries and is computationally cheap and eliminates the need to identify the ideal learning rate.

Table 4.3: Output of VGG16 model

Test	CNN Model	BS	LR	Epoch	TL(%)	VL(%)	Error Rate (%)	Acc (%)
1	VGG16	32	8e-6,1e-4	15	42.7083	43.9302	23.4493	76.5507
2	VGG16	64	8e-6,1e-4	15	76.4934	50.4612	22.4917	77.5083
3	VGG16	64	8e-6,1e-4	30	26.2982	45.7910	18.0787	81.9213
4	VGG16	32	2e-6,1e-3	15	30.7861	32.0147	16.4522	83.5478
5	VGG16	64	2e-6,1e-3	15	25.4205	25.4679	10.9682	89.0318
6	VGG16	64	2e-6,1e-3	30	7.5000	8.4696	3.0257	96.9743

Table 4.4: Output of Resnet34 model

Test	CNN Model	BS	LR	Epoch	TL(%)	VL(%)	Error Rate (%)	Acc (%)
7	Resnet34	32	8e-6,1e-4	15	42.2252	43.1934	21.4070	78.5930
8	Resnet34	64	8e-6,1e-4	15	41.1351	42.8464	21.5582	78.4418
9	Resnet34	64	8e-6,1e-4	30	12.6166	36.0723	16.3888	83.6112
10	Resnet34	32	2e-6,1e-3	15	35.0093	30.7748	14.2965	85.7035
11	Resnet34	64	2e-6,1e-3	15	26.2728	26.9305	10.8926	89.1074
12	Resnet34	64	2e-6,1e-3	30	7.6075	9.5925	2.6475	97.3525

Table 4.5: Output of Alexnet model

Test	CNN Model	BS	LR	Epoch	TL(%)	VL(%)	Error Rate (%)	Acc (%)
13	Alexnet	32	8e-6,1e-4	15	52.147	48.5449	26.0968	73.9032
14	Alexnet	64	8e-6,1e-4	15	49.9790	46.5579	25.416	74.5840
15	Alexnet	64	8e-6,1e-4	30	42.8953	44.6564	24.2814	75.7186
16	Alexnet	32	2e-6,1e-3	15	46.3035	42.8736	22.1044	77.8956
17	Alexnet	64	2e-6,1e-3	15	44.2203	42.6651	22.0121	77.9879
18	Alexnet	64	2e-6,1e-3	30	39.0666	35.3782	16.9440	83.0560

Table 4.6: Output of Resnet50 model

Test	CNN Model	BS	LR	Epoch	TL(%)	VL(%)	Error Rate (%)	Acc (%)
19	Resnet50	32	8e-6,1e-4	15	39.0362	41.5517	20.5749	79.4251
20	Resnet50	64	8e-6,1e-4	15	35.1833	40.6826	19.5159	80.4841
21	Resnet50	64	8e-6,1e-4	30	21.1929	36.9642	14.2965	85.7035
22	Resnet50	32	2e-6,1e-3	15	20.5363	37.8652	15.5068	84.4932
23	Resnet50	64	2e-6,1e-3	15	29.6796	24.4782	10.6657	89.3343
24	Resnet50	64	2e-6,1e-3	30	10.8362	5.8117	2.4206	97.5794

Identifying ideal batch size for CNNs is important as it helps the network to reach maximum accuracy in the quickest possible time, particularly for complicated datasets, such as a medical picture dataset (Kandel & Castelli, 2020).

Based on the results obtained in this study, when the learning rate and epochs remains, the accuracy of the model increases when the number of batch sizes increases from 32 to 64. Which shows that the accuracy of the model increases when the number of batch sizes increases. For instance, by referring to Test 10 and 11, when other parameters are constant, increase in batch size from 32 to 64 has resulted in increase in accuracy with an additional value of 4.67%.

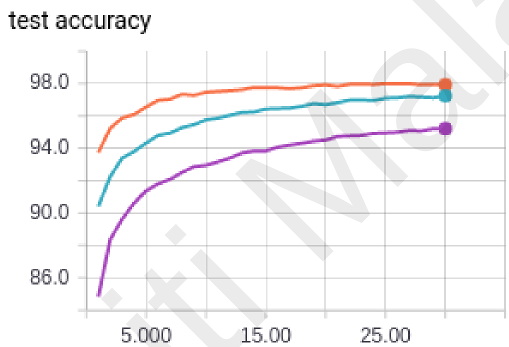


Figure 4.11: Test accuracy based on different batch size (batch size 64 for orange, 254 for blue and 1024 for purple, MNIST dataset) (Shen, 2018)

According to Shen (2018), when the batch size increases, the accuracy tends to increase as well. This is found to be similar to our result. Radium (2017) has further supported this statement. According to the author's findings, the larger the batch size, the greater the network accuracy, implying that batch size has a significant influence on CNN performance.

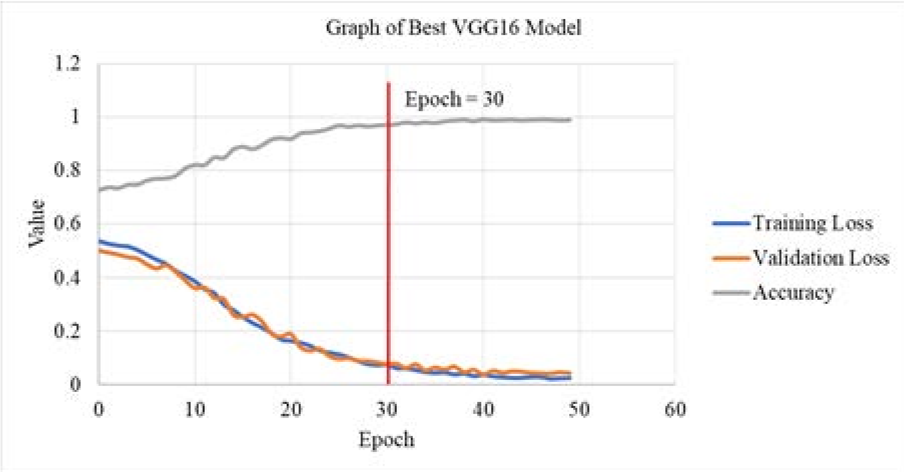


Figure 4.12: Graph of best VGG16 model

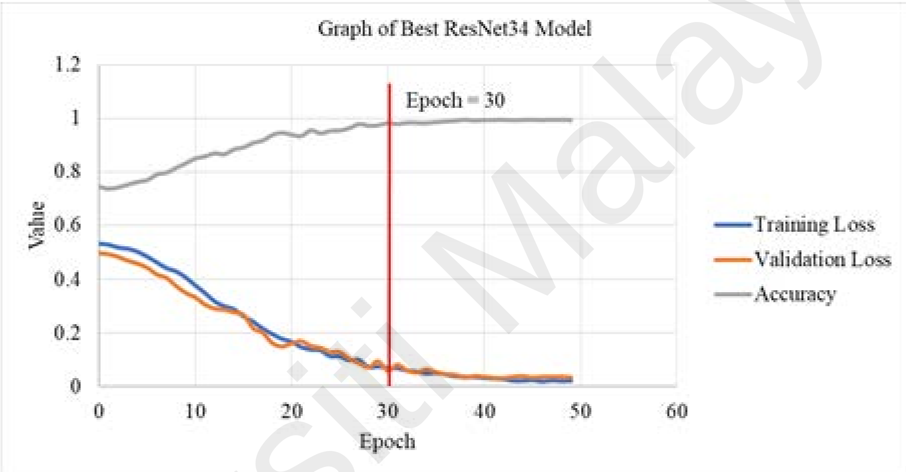


Figure 4.13: Graph of best Resnet34 model

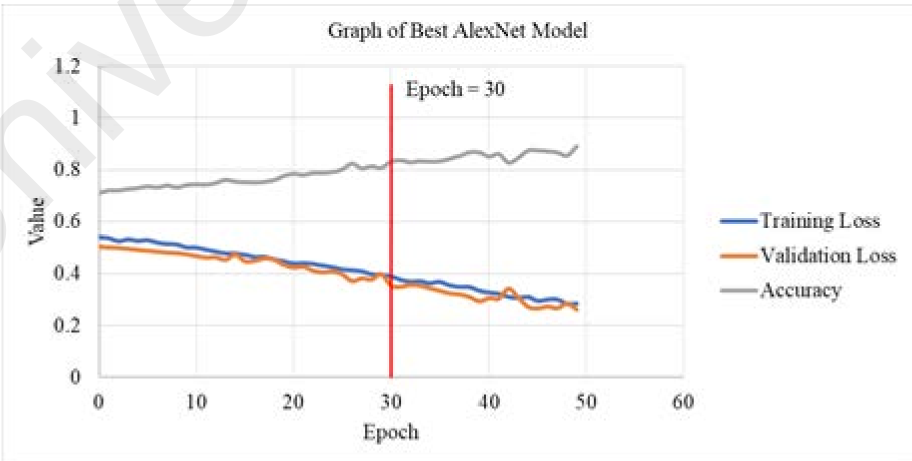


Figure 4.14: Graph of best Alexnet model

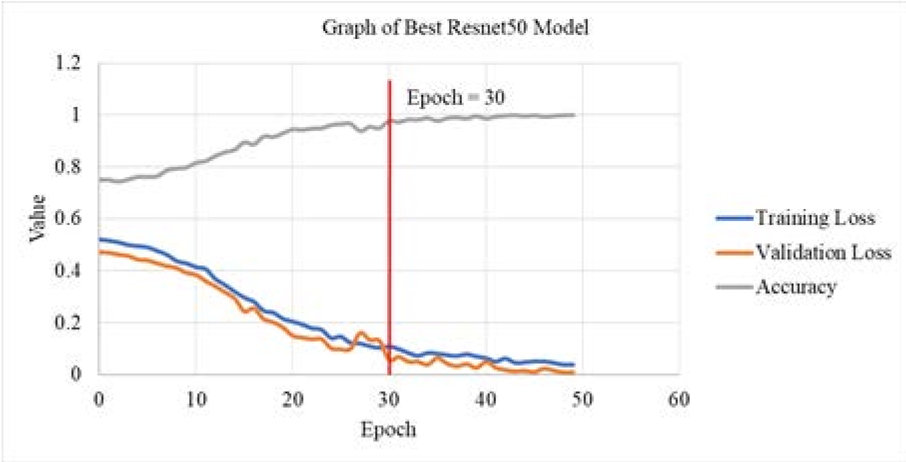


Figure 4.15: Graph of best Resnet50 model

Based on the results obtained in Table 4.3, Table 4.4, Table 4.5 and Table 4.6, it can be observed that better accuracy was achieved with smaller learning rates of 2e-6,1e-3 as compared to 8e-6,1e-4. In addition to that, as the number of epochs increases, the accuracy tends to increase as well. For instance, in test 6, the accuracy of VGG16 has managed to reach 96.9743% as compared to test 5 89.0318%, with an increase of 15 epochs. On the other hand, the increase in the number of epochs from 15 to 30 in Test 17 and Test 18 has resulted in an increase of accuracy from 77.99% to 83.06%, which is about a difference of 6.50%.

Graphical illustrations that depict the training loss, validation loss and accuracy for different models were plotted in this study. By referring to Figure 4.12 and Figure 4.15, upon reaching 30 epochs, the losses and accuracy starts to flatten out, suggesting overfitting. Therefore, the number of epochs for all the models is fixed at 30. This action is supported by Swathi (2018), the author has discovered that the early stopping method is the best method which can avoid overfitting and underfitting. When the network begins to overfit the data, the error on the validation set will begin to rise on a regular basis. The training should be stopped when the validation error increases for a predefined number of epochs.

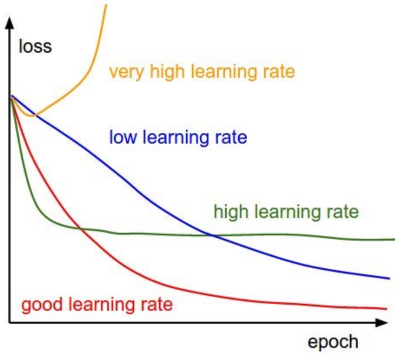


Figure 4.16: Loss curve (Karpathy, 2021)

Based on the study of Karpathy (2021), the accuracy will appear to be quite linear if learning rates are low. If the learning rates are higher, the accuracy will start to seem exponential. Learning rates that are too high will cause an adverse effect on the accuracy whereby the loss starts to increase exponentially over the batches. On the other hand, learning rates that are fairly higher than the optimum value will cause the loss to decline at a very fast rate, and appear to be constant before much loss can be reduced (green line in Figure 4.16). This is because the optimization has too much "energy" and the parameters are bouncing about in a chaotic manner.

A good learning rate shall be able to reduce the loss slowly and achieve minimal loss towards the end of the epochs (red line in Figure 4.16). By comparing the loss curve for the models in this study, the training and validation loss result is very much satistisfies. For instance, by referring to the loss curve of the Resnet50 model (epoch 30) as in Figure 4.15, the training and validation loss is considered acceptable because the loss is not increasing nor achieving linearity before minimal loss is achieved, which means that the result obtained at epoch 30 is not overfitting.

4.6 Comparison of Model with Existing Work

During the early stages of deep learning development, the architecture was simple and rudimentary. However, as deep learning became more popular, more researchers created new architectures with deeper CNN in radiomics of mammographic imaging to improve breast cancer diagnosis (Pang et al., 2020).

VGG net requires much more parameters in order to thoroughly evaluate its performance (Tsochatzidis et al., 2019). The use of VGG16 (pretrained on ImageNet) was modified in the study of Khamparia et al. (2021) and Li et al. (2021) in order to classify microcalcification images into benign or malignant cases from the DDSM database and obtained accuracy of 94.3% and 87.0% respectively. The VGG16 model trained in our study is able to surpass both models by reaching an accuracy of 96.97% with batch size 64 and the epoch 30.

Cai et al. (2019) utilized Alexnet (pretrained on Imagenet) in classifying breast microcalcification images and managed to achieve an accuracy of 79.1% upon utilizing 10-fold cross validation technique (300 epochs, learning rate of 0.01 and database from SYUCC and NAHSMU). In this research, the application for cross validation in the dataset was not performed, but the accuracy achieved is much higher, which is 83.1% with only an epoch of 30. The difference in accuracy may be due to the difference of learning rate as well as the dataset used for machine learning.

27	0.121974	0.308126	0.121029	0.878971	02:00
28	0.108851	0.307963	0.115734	0.884266	02:00
29	0.110513	0.297422	0.109682	0.890318	02:00

Figure 4.17: Learning rate of 8e-6,1e-4 for Resnet50

27	0.119694	0.161868	0.062784	0.937216	01:51
28	0.109841	0.134682	0.046899	0.953101	01:51
29	0.103276	0.131036	0.052194	0.947806	01:50

Figure 4.18: Learning rate of 2e-6,1e-3 for Resnet50

Study of Wilson & Martinez (2001) explained that for big, complicated tasks, a smaller learning rate can frequently increase generalisation accuracy substantially. The enhanced generalisation accuracy outweighs the extra time required by slower learning rates. Using a learning rate that is too low, on the other hand, is a waste of effort, and does not guarantee the best accuracy has been obtained.

As proposed by Brownlee (2019), a slower learning rate may allow the model to learn a set of weights that is more optimum or even globally optimal. This might explain why smaller learning rates may also be able to produce models with higher accuracy. Based on Figure 4.17 and Figure 4.18, a smaller learning rate of 2e-6,1e-3 for Resnet50 has managed to obtain significantly higher accuracy compared to 8e-6,1e-4 with a shorter amount of time.

The use of the Alexnet model can be observed in the study of Cai et al. (2019). The authors had utilized Alexnet pretrained on ImageNet to classify 900 images from SYUCC and NAHSMU database. Later, a 10-fold cross validation was applied on the dataset and their model had achieved an accuracy of 79.10%. However, the accuracy that was achieved in this study is much higher, which is 83.06%. The difference in the result might be due to the different database of images that was used. For instance, this research utilizes ROI calcification images of CIBS-DDSM database which provides higher resolution.

Study of Hekal, Elnakib & Moustafa (2021) have classified 1852 calcification images of CIDB DDSM database into CNN pretrained models of modified AlexNet and modified Resnet50, of which the FC8 layer in AlexNet or FC1000 layer in ResNet50 is replaced with a shallow classifier (SVM). With 20 epochs, the accuracy for breast microcalcification for Resnet50 has managed to reach 91% while AlexNet has reached 90%. Although the accuracy for the Alexnet model in this study was lower (83.1%), the accuracy for Resnet50 managed surpassed with a value of 97.6%.

Heenaye-Mamode et al. (2021) have utilized the CBIS-DDSM and UPMC database for the multiclass categorization of different anomalies which includes calcification. The improved Resnet50 model managed to obtain an 88% accuracy. The Resnet50 model from the study of Tsochatzidis, Costaridou, & Pratikakis (2019) utilized CBIS-DDSM database images only and obtained an accuracy value of 74.9%. Last but not least, the study of Hakim, Prajitno, & Soejoko (2021) utilizes 354 images from INbreast dataset and their model has managed to obtain an accuracy of 90.3%. The Resnet50 model in this study is able to surpass existing work with an accuracy value of 97.6%. The main difference between the models is the image that is fed to the machine for training. For instance, this research directly utilizes ROI calcification images of CIBS-DDSM database, which enables the machine to learn accurately the features of malignant and benign calcified cases.

The use of Resnet34 in breast microcalcification can be observed in the study of Xiao et al. (2021), where the authors had utilized 2D Resnet34 together with anisotropic 3D Resnet to classify 495 Digital Breast Tomosynthesis (DBT) microcalcification images, and reached an accuracy value of 76.0%. The model of Resnet34 in this study is able to provide a significantly higher accuracy value, which is 97.4%, probably due to the large amount of images (6611 images) utilized for machine learning, of which is 13 times larger than the study of Xiao et al. (2021).

According to Vedantham et al. (2015), DBT is able to improve breast imaging by delivering a succession of pictures (thin slices) across the breast, which decreases the effect of tissue superimposition. However, Wells (2021) had reported that tomosynthesis is still a novel process, and not all imaging technicians or clinicians are familiar with it. The algorithms used to create the 3-D pictures may also differ according to different cases, and thus affecting test results.

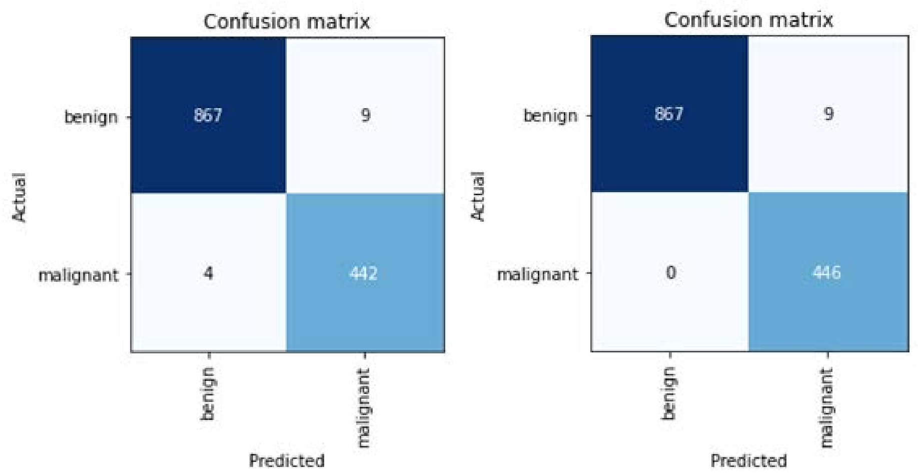


Figure 4.19: Confusion matrix of best VGG16 and Resnet34 model

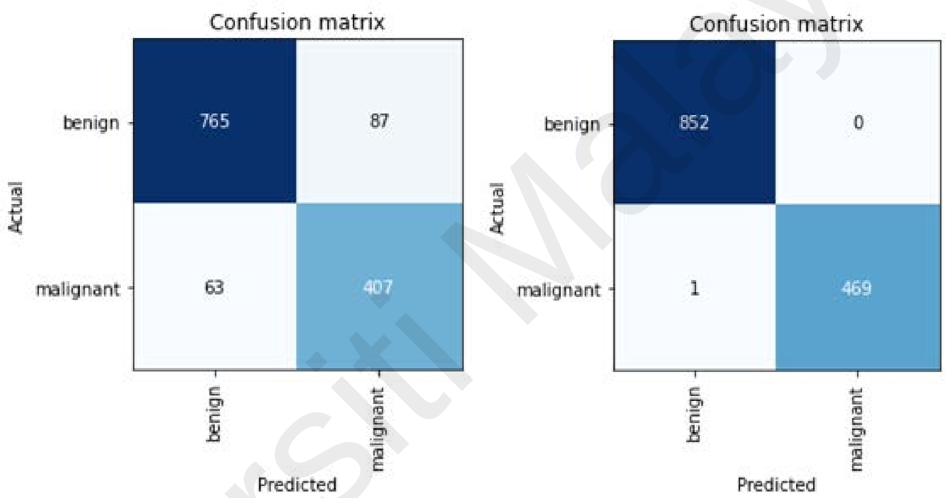


Figure 4.20: Confusion matrix of best Alexnet and Resnet50 model

Table 4.7: Tabulated confusion matrix for best Resnet34, Resnet50, VGG16 and Alexnet model

Architecture	TP	FP	TN	FN	FP/TP (%)	FN/TN (%)
Resnet34	867	9	446	0	1.04	0.00
Resnet50	852	0	469	1	0.00	0.21
VGG16	867	9	442	4	1.04	0.90
Alexnet	765	87	407	63	11.37	15.48

The tabulated value of the confusion matrix can be found in Table 4.7. Overall, the Alexnet model has both the highest percentage of falsely classified benign and falsely classified malignant cases, which is 11.37% and 15.48% respectively. The performance of the Resnet50 is considered as the best model because it only has 1 misclassified image over 1322 images, while Resnet34 has a total of 9 misclassified images. For the case of VGG, it has a total of 13 misclassified images. Based on the values obtained in the confusion matrix, calculation for additional performance measurement was performed and tabulated in Table 4.8.

Table 4.8: Additional performance measurement for best Resnet34, Resnet50, VGG16 and Alexnet model

Architecture	Recall	Precision	Specificity	Accuracy	F-1 Score	MCC
Resnet34	1.0000	0.9897	0.9802	0.9735	0.1818	0.8950
Resnet50	**0.9988**	**1.0000**	**1.0000**	**0.9758**	**0.6664**	**0.9983**
VGG16	0.9954	0.9897	0.9800	0.9697	0.1328	0.9781
Alexnet	0.9239	0.8979	0.8239	0.8306	0.0122	0.7558

Precision in other words, is the total number of positive class values predicted divided by the number of positive predictions. which is also known as Positive Predictive Value (PPV). A high number of False Positives might be indicated by a low accuracy. Recall equals the number of positive forecasts divided by the number of positive class values in the test data. It's also known as the True Positive Rate or Sensitivity. A poor recall rate suggests a high number of False Negatives (Brownlee, 2014).

According to Wood (2021), the F-score, also known as the F1-score, is a measurement that indicates how accurate a model is on a given dataset. The weighted average of Precision and Recall is the F1 score. As a result, this score considers both false positives and false negatives. Although it is not as intuitive as accuracy, F1 is typically more useful than accuracy, especially if the class distribution is unequal (Joshi, 2016). Based on the result obtained in Table 4.8, the F1-score for ResNet50 is the highest (66.64%), indicating that this model has the best quality for multi class classification.

The MCC is the most credible statistical metric since it is only high if all four confusion matrix categories are correctly predicted. As a result, MCC is the most useful and accurate metric for assessing binary classifications (Chicco & Jurman, 2020). MCC has a range of values ranging from -1 to +1. A perfect model has a score of +1, whereas a bad model has a score of -1 (Alkhaleefah et al.,2020). According to Table 4.8, Resnet50 is able to achieve the highest score of MCC with a value of 0.9983.

4.6 Summary

In summary, the best accuracy obtained for VGG16 is 96.97%, for Resnet34 is 97.35%, for Alexnet is 83.06%, while for Resnet50 is 97.58%. By comparing the accuracy value for all the trained values, the model that performs the best is Resnet50. Resnet50 is able to achieve a high accuracy of 97.58% with hyperparameters of batch size 64, cyclic learning rate of 2e-6,1e-3 and epoch of 30, trained with 6611 ROI microcalcification images from CIBS-DDSM database. Based on the confusion matrix generated for the Resnet50 model, the MCC score for Resnet50 achieved a score of 0.9983/1.0000, suggesting high accuracy. The conclusion, objectives met as well as future work for further improvement will be discussed in Chapter 5.

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion

An automated microcalcification detection in mammography for early breast cancer diagnosis using deep learning techniques has been successfully developed in this study. Prior the use of deep learning algorithm for machine learning, the collected mammogram images had undergo preprocessing operations which includes conversion of images from DICOM to *.jpeg format, resizing to 224 x 224 pixels, removal of artifacts, and image enhancement by application of adaptive median filter. Moving on, transfer learning technique for CNN architectures is employed to build a breast cancer image classifier. The pretrained CNN models were downloaded and utilized from fastai library in Google Colaboratory platform via OpenCV language. The studied models include Resnet34, Resnet50, VGG16 and Alexnet. The classification performance of machine learning based models in distinguishing between benign and malignant cases of breast cancer were assessed and the result from this study shows that Resnet50 achieves the highest accuracy with a value of 97.58%, followed by Resnet34 of 97.35%, VGG16 of 96.97% and finally Alexnet of 83.06%.

Breast cancer is a significant threat to women or men all over the world, and improving the existing state of breast cancer detection systems is a critical challenge. The exponential rise of deep learning in radiomics enables not only the extraction of valuable characteristics, but also the complete use of enormous data sets for improved breast cancer diagnosis and precision treatment. In this work, reviews on recent research on the topic of using deep learning in classification of microcalcification images into benign and malignant cases for breast cancer detection has been studied. This paper focuses on the use of mammogram images, specifically ROI images of microcalcification from CIBS-DDSM database, to train several pre-trained CNN models from fastai library in Google Colaboratory platform.

Many research focusing on deep learning for breast cancer have been published as a result of the widespread awareness about breast cancer and the quick development of deep learning algorithms. The majority of the research employs breast microcalcification images in conjunction with breast mass lesions to assess the architecture's effectiveness in categorising breast images into malignant and benign cases, therefore, there are limited papers that focus solely on breast microcalcification classification using CNN architecture. The goal of this study was to take a comprehensive approach to the most recent results on breast microcalcification categorization.

5.2 Study Limitation

There were several study limitations from the findings of this research. First of all, the data that was used in this study were all obtained from the CBIS-DDSM database, which was limited. In order to increase the amount of images for machine learning, data augmentation was utilized to increase the data size that was fed into the machine. However, this might cause the model to possibly remember the repeated patterning of the dataset for classification into benign or malignant cases. On the other hand, the ROI images provided were in random sizes. Resizing all of the ROI images into 224x224 (either upsizing or downsizing) might result in data compression and loss of useful features or information of the image.

5.3 Future Work

This study has developed an automated system for microcalcification categorization based on the analysis of ROI microcalcified images from CIBS-DDSM database. The suggested method achieved considerable results, demonstrating its capacity to be employed in clinical breast cancer analysis. Yet, there are several ways in which this work could be improved. The following are three options for improving or analysing the suggested system's performance.

1. K-fold cross validation technique could be incorporated in the algorithm. Cross-validation can be used to evaluate a model's ability to predict new data that was not utilised in the estimation process, in order to identify issues such as overfitting or selection bias, as well as to provide insight into how the model will generalise to a different dataset. Cross-validation combines metrics of prediction fitness to get a more accurate estimate of model prediction performance.

2. The AUC-ROC curve could be computed in future studies. The Receiver Operator Characteristic (ROC) curve is a binary classification issue assessment measure. It's a probability curve that displays the TPR against the FPR at different threshold levels, effectively separating the signal from the noise. The Area Under the Curve (AUC) is a summary of the ROC curve that measures a classifier's ability to discriminate between classes. The AUC indicates how well the model distinguishes between positive and negative classes. The greater the AUC, the better the model performs.

3. Different sources of breast images could be incorporated in order to identify and compare the effectiveness of the model in classifying different sources of microcalcification images. For instance, DBT, a novel technology that can be understood as '3D mammogram images', could be included to identify how well the model performs across different sources of images.

REFERENCES

Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: advances, challenges and applications. BMC bioinformatics, 20(11), 1-20.

Afaq, S., & Rao, S. (2020). Significance Of Epochs On Training A Neural Network. International Journal of Scientific and Technology Research, 19(6), 485-488.

Al-Haija, Q. A., & Adebanjo, A. (2020). Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (pp. 1-7). IEEE.

Alam, N., & Islam, M. J. (2014). Pectoral muscle elimination on mammogram using K-means clustering approach. International Journal of Computer Vision & Signal Processing, 4(1), 11-21.

Alkhaleefah, M., & Wu, C. C. (2018). A hybrid CNN and RBF-based SVM approach for breast cancer classification in mammograms. In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 894-899). IEEE.

Alkhaleefah, M., Ma, S. C., Chang, Y. L., Huang, B., Chittem, P. K., & Achhannagari, V. P. (2020). Double-shot transfer learning for breast cancer classification from X-ray images. Applied Sciences, 10(11), 3999.

Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018). Breast cancer classification using machine learning. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) (pp. 1-4). IEEE.

Artificial Neural Networks (ANN) and Different Types. (2021). Elprocus. https://www.elprocus.com/artificial-neural-networks-ann-and-their-types/

Baeldung. (2020). How ReLU and Dropout Layers Work in CNNs. Baeldung CS. https://www.baeldung.com/cs/ml-relu-dropout-layers

Barot, P. (2020). Artificial Intelligence vs. Machine Learning vs. Deep Learning. BoTree Technologies. https://www.botreetechnologies.com/blog/artificial-intelligence-vs-machine-learning-vs-deep-learning-difference/

Basile, T. M. A., Fanizzi, A., Losurdo, L., Bellotti, R., Bottigli, U., Dentamaro, R., ... & La Forgia, D. (2019). Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system. Physica Medica, 64, 1-9.

Batchu, S., Liu, F., Amireh, A., Waller, J., & Umair, M. (2021). A Review of Applications of Machine Learning in Mammography and Future Challenges. Oncology, 1-8.

Bagui, S., Fang, X., Kalaimannan, E., Bagui, S. C., & Sheehan, J. (2017). Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features. Journal of Cyber Security Technology, 1(2), 108-126.

Beeravolu, A. R., Azam, S., Jonkman, M., Shanmugam, B., Kannoorpatti, K., & Anwar, A. (2021). Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN. IEEE Access, 9, 33438-33463.

Behera, J. M., Bhandari, K., Kaur, K., & Behera, S. (2017). A review on Segmentation Techniques of Image Processing. Indian Journal of Scientific Research, 135-139.

Belisarius. (2011). Image Segmentation using Mean Shift explained. Stack Overflow. https://stackoverflow.com/questions/4831813/image-segmentation-using-mean-shift-explained

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1798-1828.

Bhandari, A. K., Maurya, S., & Meena, A. K. (2018). Social spider optimization based optimally weighted Otsu thresholding for image enhancement. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. American journal of epidemiology, 188(12), 2222-2239.

Boss, R., Thangavel, K., & Daniel, D. (2013). Automatic mammogram image breast region extraction and removal of pectoral muscle. arXiv preprint arXiv:1307.7474.

Breast Calcifications. (2021). Mayo Clinic. https://www.mayoclinic.org/symptoms/breast-calcifications/basics/definition/sym-20050834

Breast cancer. (2021). World Health Organization. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Brownlee, J. (2019). How to use Learning Curves to Diagnose Machine Learning Model Performance. Machine Learning Mastery. https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/

Brownlee, J. (2020). What is a Confusion Matrix in Machine Learning. Machine Learning Mastery. https://machinelearningmastery.com/confusion-matrix-machine-learning/

Brownlee, J. (2019). Understand the Impact of Learning Rate on Neural Network Performance. Machine Learning Mastery. https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/

Cai, H., Huang, Q., Rong, W., Song, Y., Li, J., Wang, J., ... & Li, L. (2019). Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms. Computational and mathematical methods in medicine, 2019.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., ... & Zdeborová, L. (2019). Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), 045002.

Cashmi, A. J., & Chehelamirani, M. (2019). Using Adaptive Median Filter for Noise Removal from Image to Diagnose Breast Cancer. Merit research Journal of Engineering, Pure and Applied Sciences, 5(1), 014-018.

Chen, P. H. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. Nature materials, 18(5), 410-414.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1), 1-13.

Cheriyedath, S. (2021). Microcalcifications in Breast Cancer. News-Medical.Net. https://www.news-medical.net/health/Microcalcifications-in-Breast-Cancer.aspx

Dey, N., Ashour, A., Shi, F., & Balas, V. E. (2018). Soft Computing Based Medical Image Analysis. Academic Press.

Dietz, M. (2017). Understand Deep Residual Networks — a simple, modular learning framework that has redefined state-of-the-art. Medium. https://medium.com/@waya.ai/deep-residual-learning-9610bb62c355

Es-salhi, R., Daoudi, I., Tallal, S., & Medromi, H. (2016, May). A survey on segmentation techniques of mammogram images. In International Symposium on Ubiquitous Networking (pp. 545-556). Springer, Singapore.

Fanizzi, A., Basile, T. M., Losurdo, L., Bellotti, R., Bottigli, U., Dentamaro, R., ... & La Forgia, D. (2020). A machine learning approach on multiscale texture analysis for breast microcalcification diagnosis. BMC bioinformatics, 21(2), 1-11.

Fathy, W. E., & Ghoneim, A. S. (2019). A deep learning approach for breast cancer mass detection. International Journal of Advanced Computer Science and Applications, 10(1), 175-182.

Ganin, Y., & Lempitsky, V. (2015, June). Unsupervised domain adaptation by backpropagation. In International conference on machine learning (pp. 1180-1189). PMLR.

Gayathri, B. K., & Raajan, P. (2016, January). A survey of breast cancer detection based on image segmentation techniques. In 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp. 1-5). IEEE.

George, M. J., & Dhas, D. A. S. (2017). Preprocessing filters for mammogram images: a review. In 2017 Conference on Emerging Devices and Smart Systems (ICEDSS) (pp. 1-7). IEEE.

Goel, V. (2018, September 29). Building a Simple Machine Learning Model on Breast Cancer Data. Towards Data Science. https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3

Hakim, A. N. R., Prajitno, P., & Soejoko, D. S. (2021, March). Microcalcification detection in mammography image using computer-aided detection based on convolutional neural network. In AIP Conference Proceedings (Vol. 2346, No. 1, p. 040001). AIP Publishing LLC.

Hedjazi, M. A., Kourbane, I., & Genc, Y. (2017). On identifying leaves: A comparison of CNN with classical ML methods. In 2017 25th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Heenaye-Mamode Khan, M., Boodoo-Jahangeer, N., Dullull, W., Nathire, S., Gao, X., Sinha, G. R., & Nagwanshi, K. K. (2021). Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). Plos one, 16(8), e0256500.

Hekal, A. A., Elnakib, A., & Moustafa, H. E. D. (2021). Automated early breast cancer detection and classification system. Signal, Image and Video Processing, 1-9.

Hsieh, Y. C., Chin, C. L., Wei, C. S., Chen, I. M., Yeh, P. Y., & Tseng, R. J. (2020, November). Combining VGG16, Mask R-CNN and Inception V3 to identify the benign and malignant of breast microcalcification clusters. In 2020 International Conference on Fuzzy Theory and Its Applications (iFUZZY) (pp. 1-4). IEEE.

Huang, Q., Luo, Y., & Zhang, Q. (2017). Breast ultrasound image segmentation: a survey. International journal of computer assisted radiology and surgery, 12(3), 493-507.

Jafari, S. H., Saadatpour, Z., Salmaninejad, A., Momeni, F., Mokhtari, M., Nahand, J. S., ... & Kianmehr, M. (2018). Breast cancer diagnosis: Imaging techniques and biochemical markers. Journal of cellular physiology, 233(7), 5200-5213.

Jiang, X., Xie, F., Liu, L., Peng, Y., Cai, H., & Li, L. (2018). Discrimination of malignant and benign breast masses using automatic segmentation and features extracted from dynamic contrast‑enhanced and diffusion‑weighted MRI. Oncology letters, 16(2), 1521-1528.

Joseph, A. M., John, M. G., & Dhas, A. S. (2017, March). Mammogram image denoising filters: A comparative study. In 2017 Conference on emerging devices and smart systems (ICEDSS) (pp. 184-189). IEEE.

Joshi, R. (2016). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. Exsilio Solutions. https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/

Justanial, E. & Althothali, A. (2021) Mammogram Segmentation Techniques: A Review. International Journal of Advanced Computer Science and Applications, 5(12), 520-529.

Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. ICT express, 6(4), 312-315.

Khairnar, S., Thepade, S. D., & Gite, S. (2021). Effect of image binarization thresholds on breast cancer identification in mammography images using OTSU, Niblack, Burnsen, Thepade's SBTC. Intelligent Systems with Applications, 10, 200046.

Khamparia, A., Bharati, S., Podder, P., Gupta, D., Khanna, A., Phung, T. K., & Thanh, D. N. (2021). Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. Multidimensional systems and signal processing, 32(2), 747-765.

Kashif, M., Malik, K. R., Jabbar, S., & Chaudhry, J. (2020). Application of machine learning and image processing for detection of breast cancer. In *Innovation in Health Informatics* (pp. 145-162). Academic Press.

Karpathy, A. (2021). CS231n: Convolutional Neural Networks for Visual Recognition. cs231n - Stanford University. https://cs231n.github.io/neural-networks-3/

Kaur, H., Virmani, J., & Thakur, S. (2019). A genetic algorithm-based metaheuristic approach to customize a computer-aided classification system for enhanced screen film mammograms. In U-Healthcare Monitoring Systems (pp. 217-259). Academic Press.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159-190.

Koo, M. M., von Wagner, C., Abel, G. A., McPhail, S., Rubin, G. P., & Lyratzopoulos, G. (2017). Typical and atypical presenting symptoms of breast cancer and their associations with diagnostic intervals: Evidence from a national audit of cancer diagnosis. Cancer epidemiology, 48, 140-146.

Kooi, T., Gubern-Merida, A., Mordang, J. J., Mann, R., Pijnappel, R., Schuur, K., ... & Karssemeijer, N. (2016). A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography. In International Workshop on Breast Imaging (pp. 51-56). Springer, Cham.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging: general overview. Korean journal of radiology, 18(4), 570-584.

Lehman, C. D., Wellman, R. D., Buist, D. S., Kerlikowske, K., Tosteson, A. N., Miglioretti, D. L., & Breast Cancer Surveillance Consortium. (2015). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA internal medicine, 175(11), 1828-1837.

Li, Y., Chen, H., Cao, L., & Ma, J. (2016). A survey of computer-aided detection of breast cancer with mammography. J Health Med Inf, 4(7), 1-6.

Li, J., Pei, Y., Yasin, A., Ali, S., & Mahmood, T. (2021). Computer Vision-Based Microcalcification Detection in Digital Mammograms Using Fully Connected Depthwise Separable Convolutional Neural Network. Sensors, 21(14), 4854.

Lopes, R. d. l. F. (2018). Wild Data Part 3: Transfer Learning. Stratio. https://blog.stratio.com/wild-data-part-three-transfer-learning/

Lytras, M. D., & Visvizi, A. (2018). Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research. Sustainability, 10(6), 1998.

Mahesh, B. (2020). Machine Learning Algorithms-A Review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

Maladkar, K. (2018). Why ResNets Are A Major Breakthrough In Image Processing. Analytics India Magazine. https://analyticsindiamag.com/why-resnets-are-a-major-breakthrough-in-image-processing/

Markham, K. (2021). Simple guide to confusion matrix terminology. Data School. https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

Massodi, P., Safdarian, N., & Kalantar, B. (2015, May). Detection and classification of breast cancer using feature extraction from mammography images by image processing technique and neural network. In International Conference on Non-Linear System & Optimization in Computer & Electrical Engineering.

Mavropalias, K. (2019). Understanding Fastai's fit_one_cycle method. Kostas Mavropalias. https://iconof.com/1cycle-learning-rate-policy/

Mehta, R., & Aggarwal, N. K. (2014). Comparative Analysis of Median Filter and Adaptive Filter for Impulse Noise A Review. International Journal of Computer Applications, 975, 8887.

Mezic, A. (2021). Why Unsupervised Machine Learning is the Future of Cybersecurity. Technative. https://technative.io/why-unsupervised-machine-learning-is-the-future-of-cybersecurity/

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.

Muhammad, M., Zeebaree, D., Brifcani, A. M. A., Saeed, J., & Zebari, D. A. (2020). Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images: A Review. Journal of Applied Science and Technology Trends, 1(3), 78-91.

Muthukrishnan, R., & Radha, M. (2011). Edge detection techniques for image segmentation. International Journal of Computer Science & Information Technology, 3(6), 259.

Nabi, J. (2019). Hyper-parameter Tuning Techniques in Deep Learning. Towards Data Science. https://towardsdatascience.com/hyper-parameter-tuning-techniques-in-deep-learning-4dad592c63c8

Nayak, T., Bhat, N., Bhat, V., Shetty, S., Javed, M., & Nagabhushan, P. (2019). Automatic segmentation and breast density estimation for cancer detection using an efficient watershed algorithm. In Data Analytics and Learning (pp. 347-358). Springer, Singapore.

Nelson, H. D., O'Meara, E. S., Kerlikowske, K., Balch, S., & Miglioretti, D. (2016). Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data. Annals of internal medicine, 164(4), 226-235.

Newton, E. (2019). What is the prevalence of false-positive and false-negative mammography results in breast cancer screening? Mediscape. https://www.medscape.com/answers/1945498-167946/what-is-the-prevalence-of-false-positive-and-false-negative-mammography-results-in-breast-cancer-screening

Nixon, M. S., & Aguado, A. S. (2012). Basic image processing operations. Feature Extraction & Image Processing for Computer Vision. Elsevier, 83-136.

Odat, A., Otair, M., & Shehadeh, F. (2015). Image denoising by comprehensive median filter. International Journal of Applied Engineering Research, 10(15), 36016-36022.

Pang, T., Wong, J. H. D., Ng, W. L., & Chan, C. S. (2020). Deep learning radiomics in breast cancer with different modalities: Overview and future. Expert Systems with Applications, 158, 113501.

Papandrianos, N., Papageorgiou, E., Anagnostis, A., & Feleki, A. (2020). A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans. Applied Sciences, 10(3), 997.

Peak Signal-to-Noise Ratio as an Image Quality Metric. (2016). National Instrument. https://www.ni.com/en-my/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html

Prasad, S. (2021). What is Image Segmentation? AnalytixLabs. https://www.analytixlabs.co.in/blog/what-is-image-segmentation/

PSNR. (2021). MathWorks. https://www.mathworks.com/help/vision/ref/psnr.html

Purva, H. (2020). Top 4 Pre-Trained Models for Image Classification with Python Code. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/08/top-4-pre-trained-models-for-image-classification-with-python-code/

Radiuk, P. M. (2017). Impact of training set batch size on the performance of convolutional neural networks for diverse datasets. Information Technology and Management Science, 20(1), 20-24.

Ramani, R., Vanitha, N. S., & Valarmathy, S. (2013). The pre-processing techniques for breast cancer detection in mammography images. International Journal of Image, Graphics and Signal Processing, 5(5), 47.

Redaniel, M. T., Martin, R. M., Ridd, M. J., Wade, J., & Jeffreys, M. (2015). Diagnostic intervals and its association with breast, prostate, lung and colorectal cancer survival in England: historical cohort study using the Clinical Practice Research Datalink. PloS one, 10(5), e0126608.

Riese, F. M., & Keller, S. (2020). Supervised, Semi-Supervised, and Unsupervised Learning for Hyperspectral Regression. In Hyperspectral Image Analysis (pp. 187-232). Springer, Cham.

Safdarian, N., & Hedyezadeh, M. (2019). Detection and classification of breast cancer in mammography images using pattern recognition methods. Multidisciplinary Cancer Investigation, 3(4), 13-24.

Saha, S. (2018). A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Towards Data Science. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Shareef, SR., (2014). Breast cancer detection based on watershed transformation. IJCSI, 11(1), 237-45.

Shen, K. (2018). Effect of batch size on training dynamics. Mini Distill. https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e

Shrivastava, N., & Bharti, J. (2019). A comparative analysis of medical image segmentation. In International Conference on Advanced Computing Networking and Informatics (pp. 459-467). Springer, Singapore.

Shorten, C. (2019). Introduction to ResNets. Towards Data Science. https://towardsdatascience.com/introduction-to-resnets-c0a830a288a4

Smith, K. (2016). CBIS-DDSM. The Cancer Imaging Archive (TCIA). https://wiki.cancerimagingarchive.net/plugins/servlet/mobile?contentId=22516629#content/view/22516629

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. CA: a cancer journal for clinicians, 65(1), 5-29.

Spruit, M., & Lytras, M. (2018). Applied data science in patient-centric healthcare: Adaptive analytic systems for empowering physicians and patients.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Savage, J. (2015). Mammographic image analysis society (MIAS) database v1.21 [Dataset]. Available from https://www.repository.cam.ac.uk/handle/ 1810/250394.

Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., ... & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. International journal of biological sciences, 13(11), 1387.

Suradi, S. H., Abdullah, K. A., & Isa, N. A. M. (2021, April). Breast Lesions Detection Using FADHECAL and Multilevel Otsu Thresholding Segmentation in Digital Mammograms. In International Conference on Medical and Biological Engineering (pp. 751-759). Springer, Cham.

Swathi, P. (2018). Analysis on Solutions for Over-fitting and Under-fitting in Machine Learning Algorithms. International Journal of Innovative Research in Science, Engineering and Technology, 7(12), 12401-12404.

Thanh, D. N., Erkan, U., Prasath, V. S., Kumar, V., & Hien, N. N. (2019, April). A skin lesion segmentation method for dermoscopic images based on adaptive thresholding with normalization of color models. In 2019 6th International Conference on Electrical and Electronics Engineering (ICEEE) (pp. 116-120). IEEE.

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

Use rolling-ball algorithm for estimating background intensity. (2021). Scikit-Image. https://scikit-image.org/docs/stable/auto_examples/segmentation/plot_rolling_ball.html

van Bommel, R. M., Weber, R., Voogd, A. C., Nederend, J., Louwman, M. W., Venderink, D., ... & Duijm, L. E. (2017). Interval breast cancer characteristics before, during and after the transition from screen-film to full-field digital screening mammography. BMC cancer, 17(1), 1-7.

Vagssa, P., Doudou, N. M., Jolivo, T., Videme, O., & Kolyang, D. T. (2020). Pectoral muscle deletion on a mammogram to aid in the early diagnosis of breast cancer. International Journal of Engineering, Science and Technology, 12(3), 57-65.

Vedantham, S., Karellas, A., Vijayaraghavan, G. R., & Kopans, D. B. (2015). Digital breast tomosynthesis: state of the art. Radiology, 277(3), 663-684.

Waks, A. G., & Winer, E. P. (2019). Breast cancer treatment: a review. Jama, 321(3), 288-300.

Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). Discrimination of breast cancer with microcalcifications on mammography by deep learning. Scientific reports, 6(1), 1-9.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), 1-40.

Wells, D. (2021). What You Need to Know About Tomosynthesis for Breast Cancer. Health Line. https://www.healthline.com/health/breast-cancer/tomosynthesis

Wellings, E., Vassiliades, L., & Abdalla, R. (2016). Breast cancer screening for high-risk patients of different ages and risk-which modality is most effective?. Cureus, 8(12).

Wilkinson, L., Thomas, V., & Sharma, N. (2017). Microcalcification on mammography: approaches to interpretation and biopsy. The British journal of radiology, 90(1069), 20160594.

Wilson, D. R., & Martinez, T. R. (2001, July). The need for small learning rates on large problems. In IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222) (Vol. 1, pp. 115-119). IEEE.

Wu, L., Perin, G., & Picek, S. (2020). I Choose You: Automated Hyperparameter Tuning for Deep Learning-based Side-channel Analysis. IACR Cryptol. ePrint Arch., 2020, 1293.

Xian, M., Zhang, Y., & Cheng, H. D. (2015). Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. Pattern Recognition, 48(2), 485–497. doi:10.1016/j.patcog.2014.07.026

Xi, P., Shu, C., & Goubran, R. (2018, June). Abnormality detection in mammography using deep convolutional neural networks. In 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-6). IEEE.

Xiao, B., Sun, H., Meng, Y., Peng, Y., Yang, X., Chen, S., ... & Zheng, J. (2021). Classification of microcalcification clusters in digital breast tomosynthesis using ensemble convolutional neural network. BioMedical Engineering OnLine, 20(1), 1-20.

Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., & Carson, P. L. (2019). Medical breast ultrasound image segmentation by machine learning. Ultrasonics, 91, 1-9.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), 611-629.

Zhang, H., Zhang, L., & Jiang, Y. (2019, October). Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP) (pp. 1-6). IEEE.

Zhao, H., Zou, L., Geng, X., & Zheng, S. (2015). Limitations of mammography in the diagnosis of breast diseases compared with ultrasonography: a single-center retrospective analysis of 274 cases. European journal of medical research, 20(1), 1-7.

Zucca-Matthes, G., Urban, C., & Vallejo, A. (2016). Anatomy of the nipple and breast ducts. Gland surgery, 5(1), 32.