BREAST CANCER DETECTION USING ML

ZIYAD ELEYAN

A REPORT SUBMITTED AS PART OF THE REQUIREMENTS FOR THE DEGREE
OF MSC IN APPLIED STATISTICS AND DATA SCIENCE
AT THE BIRZEIT UNIVERSITY

August 2023

Supervisor Dr. Hassan Abu Hassan

Abstract

This research aims to build a computer-aided diagnosis platform to help medical experts analyze mammogram images, diagnose breast cancers, and integrate the medical information system in Palestine with the medical imaging prediction model. Furthermore, use Artificial Intelligent to analyze, predict and classify medical cases. The research problem is that breast cancer is the most commonly diagnosed cancer and the leading cause of cancer death among women in Palestine. Early breast cancer screening and detection are crucial for decreasing death rates and successful treatment. The primary objectives are to use Artificial Intelligence (AI) to aid early and rapid breast cancer diagnosis and build an experimental framework using state-of-the-art Deep Learning methods. The study's methodology is to collect and reprocess the data to increase the quality of the data, and the processing data will use to learn the Machine Learning model to predict the patient's cases. The Palestinian Ministry of Health (MOH)'s dataset consists of 783 cases divided into 233 without a tumor, 226 with a benign tumor, 163 with a benign tumor, and 160 with a malignant tumor. Images are 45.6% normal, 15.7% malignant, and 38.3% benign. The MOH's dataset of mammograms is rich in quantity but needs more stability in classification and storage. The Machine Learning (ML) model can predict cases of breast cancer patients with or without an infection with an accuracy of 83% and can be deployed on a web application program to help in the MOH's work. We have conducted Six experiments to reach the best results and build the most accurate model. The best model was VGG-16, with a validation loss of 0.4271 and a validation accuracy of 83.26%. The F1 score decreased when the model performance was conducted on the MIAS public dataset to 63.87%. Data quality, classification, tabulation, and suitability are critical factors in raising the model's efficiency.

Acknowledgements

I want to express my special thanks and gratitude to my teacher, family, and friends who supported me in doing this project.

Declaration

I confirm that the work contained in this MSc project report has been composed solely by myself and has not been accepted in any previous application for a degree. All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.

Signed .... Date ....
Ziyad Eleyan

Contents

Abstract ii

Acknowledgements iii

Declaration iv

1 Introduction 1

- 1.1 Background 1
- 1.2 Research Problem 1
- 1.3 Research Objectives 2
- 1.4 Importance of the study 3
- 1.5 Thesis Structure 3
- 1.6 Conclusion 3

2 Literature Review 1

- 2.1 Introduction 1
- 2.2 Artificial Intelligent 2
  - 2.2.1 Machine Learning 2
  - 2.2.2 Deep Learning 2
  - 2.2.3 Convolution Neural Network 3
  - 2.2.4 Transfer Learning 3
- 2.3 Medical Images 3
  - 2.3.1 Medical Image Compression 4
  - 2.3.2 Medical Image Pre-processing 4
  - 2.3.3 Augmentation 4
  - 2.3.4 Medical Images Public Datasets 4
  - 2.3.5 Medical Image Classification 5
- 2.4 Challenges 7
- 2.5 Conclusion 7

3 Methods 8
3.1 DataSet 8
3.2 Data Prepossessing 10
3.3 Learning Process 12
3.4 Model Evaluation 15
3.5 Conclusions 16

4 Implementation 17
4.1 Dataset 17
4.2 Data Prepossessing 21
4.3 Learning Process 23
4.4 Model Evaluation 24
4.5 Conclusions 24

5 Experiments & Results 26
5.1 Conclusions 31

6 Conclusion 32
6.1 Conclusions 32
6.2 Future Work 32

A BCD Functional & Non-Functional Requirements 39
A.1 Functional Requirements 39
A.2 Non-Functional Requirements 43

B Project Management 47

C Breast Cancer Detection App 49

D Presentation Slides 52

E Outputs of Experiment One (Model One) 55

F Outputs of Experiment Two (Model Two) 59

G Outputs of Experiment Three (Model Three) 63

H Outputs of Experiment Four (VGG 16 Model) 66

I Outputs of Experiment Five (Inception V3 Model) 69

J Outputs of Experiment Six (Resnet Model) 72

List of Tables

4.1 MOH DataSet Sample 18
4.2 Mammogram Results Description 18
4.3 Parameters Description 18
4.4 Age & Date Description 19
4.5 Curated Breast Imaging Subset of DDSM (CBIS-DDSM) Data Sample . 19
4.6 CBIS-DDSM Stat Description 20
4.7 CBIS-DDSM Parameters 20

5.1 Models Training Result 28
5.2 Confusion Matrix for Visual Geometry Group 16 (Vgg16) Model 28
5.3 Confusion Matrix for ResNet Model 29
5.4 head of the information for MIAS 30
5.5 Confusion Matrix for MIAS Dataset 31

A.1 BCD Functional Requirements 42
A.2 BCD Non-Functional Requirements 46

B.1 Project Tasks 48

List of Figures

2.1 Transfer Learning 6

3.1 CBIS-DDSM: Breast Cancer Image Dataset Samples 9
3.2 MOH Breast Cancer Image Dataset samples 10
3.3 CBIS-DDSM Mammogram Images Preprocessing 11
3.4 MOH Mammogram Images Preprocessing 11
3.5 CBIS-DDSM Proposed Data Structuer 12
3.6 MOH Proposed Data Structuer 12
3.7 Project Workflow 13
3.8 Model One Learning Workflow 13
3.9 U-Net Architecture (Wang et al., 2020) 14
3.10 Model Two Learning Workflow 15

4.1 CBIS-DDSM Mask 17
4.2 MOH Sample label 17
4.3 Box Plot for Age 19
4.4 Raw Data for MOH Dataset 21
4.5 Image Preprocessing 21
4.6 Image Preprocessing 22
4.7 DataSet One 22
4.8 DatSet Two 22

5.1 Unsplit Dataset 26
5.2 Dataset One 27
5.3 Dataset Three 27
5.4 Vgg Accuracy Graph 29
5.5 Vgg Loss Graph 29
5.6 Mias DataSet Sample 30

6.1 Future Design 33

C.1 BCD APP 1 49
C.2 BCD APP 2 50
C.3 BCD APP 3 51

List of Abbreviations

ML	Machine Learning	ii
DL	Deep Learning	1
MOH	Palestinian Ministry of Health	ii
AI	Artificial Intelligence	ii
Vgg16	Visual Geometry Group 16	vii
CBIS-DDSM	Curated Breast Imaging Subset of DDSM	vii
BC	Breast Cancer	1
X-ray	Energetic High-Frequency Electromagnetic Radiation	1
CNNs	Convolutions Neural Networks	1
MRI	Magnetic Resonance Imaging	1
ANN	Artificial neural networks	3
CV	Computer Vision	1
CT	Computed tomography	2
NLP	Natural language processing	3
TL	Transfer Learning	3
BCD	Breast Cancer Detection	3
NIFTI	Neuroimaging informatics technology	4
MIAS	Mammographic Image Analysis Society	4
BCS-DBT	Digital Breast Tomosynthesis	4
DICOM	Digital Imaging and Communication in Medicine	4
DDSM	Digital Database for Screening Mammography	4
CXR	Chest X-rays	5
CNN-based	Convolutions Neural Network based	5
GANs	Generative Adversarial Networks	13
OCR	Optical character recognition	11
IoU	Intersection over union	15
BI-RADS	Breast-Imaging Reporting and Data System	18
SGD	Stochastic Gradient Descent	23
HTML	Hyper Text Markup Language	24
CSS	Cascading Style Sheets	24
HIS	Health Information System	34

Glossary

Artificial Intelligent Artificial intelligence focuses on developing computers capable of human-like thought processes, extending human intelligence through computers.. ii

Artificial Neural Network ANNs are parallel computational models mimicking human brain function, consisting of simple processors and weighted connections, generating simple non-linear outputs.. 2

BI-RADS The Breast Imaging Reporting and Data System (BI-RADS) was initiated by the American College of Radiology (ACR) in the 1980s to standardize mammography practice reporting. Its data-driven dictionary of imaging features offers opportunities for quality assurance, communication, research, and patient care. The BI-RADS lexicon's history provides a valuable guide for future lexicon development. The system ranks test findings according to seven categories, ranging from normal to highly suspicious or malignant. It also includes four additional categories for mammograms, describing the level of breast density.. 17

Computer Vision Computer vision is an AI field that uses digital images and videos to derive meaningful information and make recommendations, surpassing human capabilities through advanced cameras, data, and algorithms.. 2

Deep Learning Deep learning imitates human brain learning and uses complex images, text, and audio patterns to enable precise insights and predictions.. ii

Generative adversarial networks GANs are AI algorithms solving generative modeling problems using training examples and probability distribution, generating high-resolution images and using deep learning for prediction improvement.. 7

Inception-v3 Inception v3 is a convolutional neural network for image analysis and object detection, with 78.1% accuracy on the ImageNet dataset.. 6

Machine Learning Machine learning enhances computer performance without human intervention, enabling adaptability and problem-solving in data science problems through continuous learning and experience.. ii

Natural Language Processing Natural Language Processing involves computational techniques for analyzing and representing natural texts for human-like language processing in various tasks.. 2

ResNet ResNet, introduced in 2015, excels in deep residual learning, replacing VGG-16 layers and efficiently training networks with 100-1000 layers.. 3, 6

Transfer Learning Transfer learning enhances target learners' performance by transferring knowledge from related domains, reducing data dependence.. 3

VGG-16 VGG16, a 16-layer convolutional neural network, achieves 92.7% accuracy in ImageNet dataset, replacing large filters on AlexNet with smaller 3x3 filters.. ii

Chapter 1

Introduction

1.1 Background

ML methods (Mitchell, 1997) and Deep Learning (DL) (Goodfellow et al., 2016a) have been successfully applied to a wide range of applications across various domains, including medical imaging (Spanhol et al., 2015), engineering (Elyan et al., 2020), constructions, sports, and others.

Convolutions Neural Networks (CNNs) (Krizhevsky et al., 2012), in particular, played a crucial role in significantly advancing the field of image processing and understanding. These advances have greatly helped improve various medical applications, such as Energetic High-Frequency Electromagnetic Radiation (X-ray) classification, heart segmentation, robotic surgery, etc.

The main objective of this project is to provide a platform for computer-aided diagnostics that will assist medical professionals in analyzing mammography pictures and identifying breast malignancies. The solution will be developed for testing purposes at some medical points of care in Palestine.

1.2 Research Problem

Breast Cancer (BC) is a common disease affecting women worldwide (Key et al., 2001). In Palestine, the situation is even worse. According to the Palestinian Ministry of Health MOH figures (PHIC, 2020), BC is the most commonly diagnosed cancer and the leading cause of cancer death among women in Palestine. In 2020, more than 526 new BC cases were reported in Palestine. This is almost 16.5% of the total cancer cases in the country and is equivalent to 19.1 people per 100,000 population. The figures of the MOH indicate an increase in female diagnoses of this disease, with an incidence rate reaching 38.4 per 100,000 female population. Early BC screening and detection are crucial for decreasing death rates and successful treatment. In this project, we aim to build an intelligent data-driven framework to process, analyze and predict signs of BC from relevant medical images (e.g., mammograms, Ultrasounds, X-rays, etc.). The recent progress in DL and Deep CNNs resulted in significant improvement in the Computer Vision domain. Today, intelligent algorithms showed superior performance over humans and medical images in analyzing and understanding medical images. Furthermore, various complex tasks were solved using the latest AI and DL domain development, including image classification, segmentation, and object detection and recognition (e.g., localizing abnormality in a medical image). Hence, using such advanced technologies and algorithms can significantly advance health and well-being in Palestine, especially in an area such as cancer early detection and recognition. This can help the medical staff with an accurate decision support system to help save lives. The project team will include the student Ziyad Eleyan and the supervision team Dr. Hassan Abu Hassan. All results and findings will be shared with the MOH, and any possible dissemination will be subject to MOH approval.

1.3 Research Objectives

The main objective of the research is to integrate the medical information system in Palestine with the medical imaging system and use artificial intelligence AI to analyze, predict and classify medical cases. The overall objectives of this project can be outlined as follows:

- Use AI to aid early and rapid BC diagnosis to decrease BC death.
- Build and evaluate an experimental framework using state-of-the-art DL methods to detect and classify the tumor as malignant or benign.
- Make a web application portal to upload and classify medical images of BC and host the web application in the MOH. Report and discuss results and findings to evaluate the accuracy in real cases and improve the performance of this application.
- Propose a solution to build a technical system linking the Palestinian health information system with health medical images, including mammograms, Computed tomography (CT) scans, and ultrasound with biopsy results and the adoption of AI technology to determine the susceptibility to BC in the future.
- Outline future directions for developing AI-driven solutions for medical-related

problems in Palestine.

- Critically evaluate and review existing literature on medical image classification using ML and DL approaches and Breast Cancer Detection (BCD) using DL, emphasizing BC.

1.4 Importance of the study

The importance of the study lies in developing early detection tools for BC with high accuracy, low cost, and in a short time. These tools can reduce cancer patients' death incidence through early disease detection. In addition, these tools can be generalized to most other cancers in Palestine.

1.5 Thesis Structure

The remaining parts of this thesis are as follows:

Chapter 2 Literature Review. In this chapter, a detailed critical evaluation of existing related literature will be thoroughly reviewed, evaluated, and summarised.

Chapter 3 Methods. In this chapter, methods and datasets will be presented, justified, and discussed.

Chapter 4 Implementation. This chapter deals with the technical implementation of the experimental framework for BC detection.

Chapter 5 Experiments and Results. In this chapter, experiments and results will be presented, discussed, and interpreted.

Chapter 6 Conclusion. Here, final findings, conclusions, and future directions will be outlined.

1.6 Conclusion

AI is among the most advanced scientific fields and has overlapped with all scientific fields, including medicine. Cancers, in general, and BC are among the real threats to humanity, and early cancer detection is one of the most critical factors that lead to treatment. There is a high chance that AI can be used for early cancer detection to protect patients with low cost and high accuracy.

Chapter 2

Literature Review

In this chapter, a review of relevant literature will be examined and critically evaluated. First, computer vision core tasks will be introduced in the context of medical images. This will be followed by a focused review of work related to breast cancer classification using deep learning-based methods. Finally, challenges will be discussed, and a conclusion will be drawn.

2.1 Introduction

Diagnosis of cancer screening has been investigated for more than 40 years. BC can be found and diagnosed using medical imaging techniques such as diagnostic mammography X-ray, magnetic resonance imaging, ultrasound (sonography), and thermography. However, a biopsy is an accurate method to determine if cancer is present. The surgical (open), vacuum-assisted, core needle and fine needle aspiration techniques are the most widely used for biopsies.

AI is effectively used in medicine to diagnose and lower errors. AI, for instance, helps with early disease detection. AI is also used in diagnostic procedures and screenings. Significant advancements have been made in automatic medical imaging analysis in recent years. For instance, DL-based techniques have been successfully used to identify COVID-19, and they have also been used to manage Magnetic Resonance Imaging (MRI) images.

Computer Vision (CV) aims to enable the computer to process and evaluate visual content, such as 2D, movies, and 3D images. CV is extensively used in various industries, including oil and gas, fishing and agriculture, medical image analysis, robotic surgery, etc. Three major CV tasks include image classification, object identification and recognition from images, and image segmentation. Image classification challenges are the most frequent CV issues. The complete image must be categorized to classify a picture into one or more separate classes in classification issues. However, there are several challenges in using computer vision in practical situations, such as poorly defined problems and the quantity and quality of data.

2.2 Artificial Intelligent

AI is a general term that implies using a computer to mimic intelligent human behavior with minimal human intervention (Hamet and Tremblay, 2017). Also, AI is the set of activities that make the machine intelligent (Nilsson, 2009). AI applies to various medical items, such as robotics, medical diagnosis, statistics, and human biology. AI includes ML, Neural Networks, CV, Robotics, Speech Processing, Natural Language Processing, and DL (Stone et al., 2016). AI is broad and includes reinforcement, unsupervised, and supervised machine learning. However, this thesis will concentrate on supervised learning.

2.2.1 Machine Learning

ML allows computers to find hidden patterns (Bishop and Nasrabadi, 2006). ML is a data analysis method that automates the construction of an analytic model (Hackeling, 2017). ML allows the computer to learn without being explicitly programmed. This is achieved by using a large amount of labeled training data. Therefore, ML is classified as supervised, unsupervised, and reinforcement learning. Supervised learning is about learning from labeled data. For example, if we have a data set of 100 images labeled as benign and malignant cases, we can build a model to predict the diagnoses for new patients. Supervised learning algorithms are trained using the addressed set, such as the input where the selected output is known. The network receives a set of corresponding inputs and outputs, and the algorithm learns by comparing its actual output with the real output to find errors. Supervised learning is commonly used in the health sector, where historical data predict possible future events (Caruana and Niculescu-Mizil, 2006).

2.2.2 Deep Learning

Artificial Neural Network simulates the human biological brain with one input, hidden, and output layer. DL refers to neural networks with more than one hidden layer. DL enables the computer to build a complex concept from a simple idea. DL is associated with transforming and extracting features that attempt to establish a relationship between brain stimuli and associated neural responses. DL allows machines to solve complex problems even when using a data set that is very diverse, unstructured, and inter-connected (Goodfellow et al., 2016b).

2.2.3 Convolution Neural Network

In CV, Artificial neural networks (ANN)s will increase the parameters and lose the two-dimension information when flattening the images. So CNNs are created to convolve the images by multiple filters to extract the image features. The CNNs model is trained to determine the best filter weight values. And these will help to reduce the number of parameters as ANN by focusing on local connectivity. Not all neurons will be fully connected, only connected to a subset of local neurons in the next layer. CNNs are fed into another convolution layer, allowing the network to discover patterns in the patterns, usually with more complexity for later convolutions layers. Lenet-5, AlexNet, GoogLeNet, and ResNet are the famous CNNs architecture.

2.2.4 Transfer Learning

The lack of a large and publicly available annotated database is one of the biggest challenges facing DL, so Transfer Learning is one of the solutions that help to overcome this problem (Ribeiro et al., 2016). Transfer Learning (TL) is an ML pattern that forces additional source data from other interconnected fields for learning (Zhu et al., 2011). TL has solved various practical issues, including speech recognition, image processing, and Natural language processing (NLP). The TL can be categorized into transductive, inductive, unsupervised, and destructive (Niu et al., 2020). TL trains the model with big data set to transfer the knowledge of this model to another small data set with minor tuning for the original model. Furthermore, TL can leverage an existing neural network architecture proven to work on problems similar to our problem, which has already learned patterns on similar data to our data that will produce an accurate module with fewer data.

2.3 Medical Images

Medical images provide spatially resolved information about the human body. Various medical applications use different modalities (Oakden-Rayner, 2020). The most common medical modality used is radiography, magnetic resonance imaging, and ultrasound imaging (Ehrhardt, 2015). Medical image screening in all medical fields includes emergency medicine, oncology, orthopedics, and surgery. For example, a mammogram is a projection radiology image of the breast used to diagnose early signs of breast cancer. Medical image screening tasks can broadly be classified, recognized, and segmented.

2.3.1 Medical Image Compression

Digital Imaging and Communication in Medicine (DICOM) is the standard for medical imaging and related information. DICOM defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use(Bidgood et al., 1997). Most hospital digital image data are stored in DICOM format. However, DICOM is mainly used in a clinical setting and often requires a conversation with other formats for ML. Neuroimaging informatics technology (NIFTI) is an open format for storing medical image data that efficiently hold medical image data together with necessary metadata. However, NIFTI is not a clinical standard, mainly used in research settings (Larobina and Murino, 2014).

2.3.2 Medical Image Pre-processing

Preprocessing is essential for breast cancer detection, and various techniques are used to improve image quality, remove noise, preserve edges, and enhance and smoothen the image (Ramani et al., 2013). Preprocessing is the data preparation for DL prediction (Jeyavathana et al., 2016). DL method may need a particular input format to improve data access during training. Medical image preprocessing includes augmentation, scaling, normalization, noising removal, and complex artifact correction. The appropriate preprocessing for medical imaging in the proper order can enhance the undertaking of the DL model (Masoudi et al., 2021).

2.3.3 Augmentation

One of the most important methods that help increase medical images is augmentation. DL needs augmentation when insufficient annotated medical images are required to train the model (Frid-Adar et al., 2018). Data augmentation has become a widespread method for expanding the size of a training dataset, specifically in domains where big datasets aren't available. Data augmentation has become a general method for increasing the size of a training dataset, specifically in domains where big datasets aren't available (Chlap et al., 2021).

2.3.4 Medical Images Public Datasets

Various datasets are available in the public domain. This data has helped advance research and development in medical image analysis and improvement. Typical datasets include Mammographic Image Analysis Society (MIAS) (Suckling et al., 2015), mini-MIAS, CBIS-Digital Database for Screening Mammography (DDSM), DDSM, and Digital Breast Tomosynthesis (BCS-DBT) Dataset. A specific example of the breast cancer public set includes the histopathology dataset The BreakHis generated from breast tissue biopsy slides for 82 patients, generating 7909 images with four magnification sizes. The data set has two primary classes, with four types for each category. The proposed model accuracy for the BreakHis dataset is 80-85%, depending on the image magnification, feature extraction, and classifier algorithm (Spanhol et al., 2015).

2.3.5 Medical Image Classification

Deep transfer learning is used in most of the existing research in one way or another to address medical image classification problems. For example, Umar Ibrahim et al. presented a framework using a pre-trained AlexNet Model to classify Chest X-rays (CXR) images into different categories (e.g. COVID19, normal, bacterial, viral pneumonia, and other types of abnormalities) (Ibrahim et al., 2021). The authors evaluated their methods using four open datasets, including CXR dataset[1], Kaggle's Data[2], Kaggle Covid Dataset[3], and a dataset of CT scan images and CXR (Kermany, 2018) and reported relatively high sensitivity and accuracy with a marginally lower F1 score. Interestingly, no preprocessing or data augmentation has been done, demonstrating the potential of using trained models to complete such classification tasks.

Similarly, Abbas et al. classified COVID-19 from CXR scan images based on TL (Abbas et al., 2020). The authors employed widely used CNNs techniques to extract image features, including Vgg16, ResNet, etc. After reducing the dimensionality of the data using a dimension reduction technique on the features space, they used a clustering algorithm (K-means) to find within-class similarity before supplying the features to the final classifier. A 98% accuracy rate was reported. However, because only two datasets with a limited number of instances were used for evaluation, it is challenging to determine how generalizable the suggested methods are to other unobserved data. X and y presented CNNs-based methods to classify X-ray images as infected.

In recent years, Convolutions Neural Network based (CNN-based) methods have gained significant popularity in detecting pneumonia and COVID-19 in CXR scan images, especially after the outbreak of COVID-19 in December 2020. However, training CNN-based methods, especially deep models, require many training instances (images), and such data is only sometimes readily available. One common alternative to overcome this requirement is to use a popular method called TL (Pan and Yang, 2009). In a nutshell, TL means reusing a model created to perform a certain task to solve a similar one/s. Technically speaking, this means using some of the most common architectures of CNNs that were trained using many images to perform a specific classification task in solving other equivalent tasks (e.g., pneumonia detection). Various CNN-based models

---

[1] https://github.com/ieee8023/covid-chestxray-dataset

[2] https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[3] https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset

exist in the literature and are often used for TL. These include AlexNet (Krizhevsky et al., 2012), GoogleNet (Szegedy et al., 2015), Inception-v3 (Szegedy et al., 2016), Vgg16 and 19 (Simonyan and Zisserman, 2015), ResNet (He et al., 2016) and others. It should be noted that most of these common CNNs architectures have been trained using the ImageNet Dataset, which contains more than 1.2 million images. Reusing these and other models is relatively straightforward and only requires freezing the lower layers of the pre-trained models. The motivation is that the CNNs models' lower layers learn generic features common across various images and domains, such as edges, corner points, etc. The only requirement to apply transfer learning includes changing the input layers to align with the input size of the data and the output layer (final outcome/s). However, this approach needs to be fine-tuned in some situations, where all lower layers are frozen while the deep layers (the last few layers of the model) are retrained using the new dataset. Figure 2.1 provides a schematic diagram to illustrate this technique which has significantly accelerated research, development, and application of DL-based methods across different domains., and made it possible to train deep models using relatively small-sized datasets and with moderate computing power.

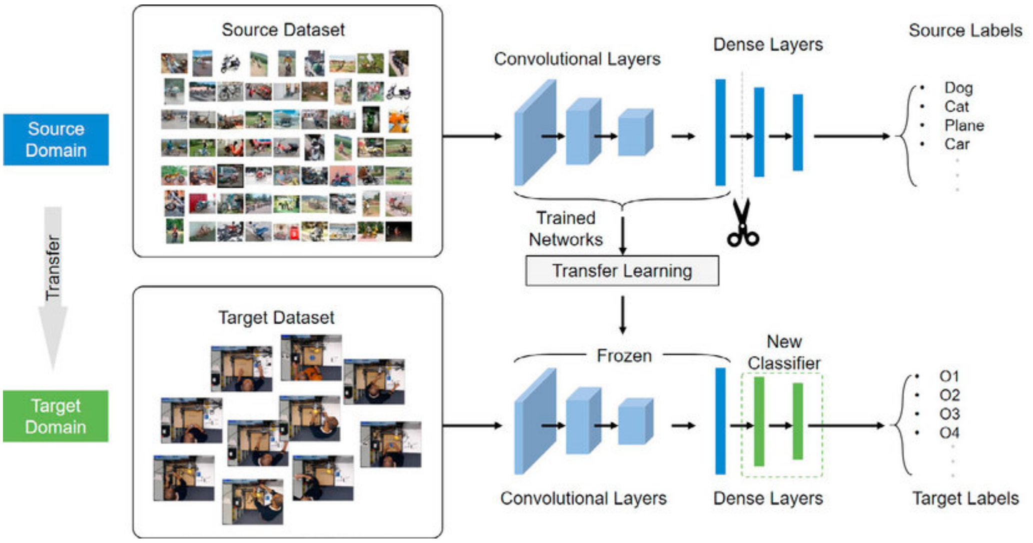


Figure 2.1: Transfer Learning

One of the most cited papers in this area (Minaee et al., 2020) also used deep TL to detect COVID-19 from CXR images. They used common CNNs architectures (e.g., ResNet, DenseNet, and their variants) pre-trained on the public dataset image. They also fine-tuned the final layer in these models using 2000 images collected from public resources. Using simple data augmentation techniques such as Data augmentation, including image flipping, slight rotation, and adding minor distortions to the training data, they reported overall sensitivity and specificity rates greater than 90% on a test set of 3000 images. The authors released the data and the methods which make the results reproducible. However, further experiments are required to assess the model's performance across other data distributions. Furthermore, it would be interesting to see if fine-tuning additional deeper layers would result in better results.

Rahaman et al. Used a customized dataset of 860 CXR scan images from public resources. The dataset samples were chosen to maintain balanced distribution between the different categories (pneumonia, COVID-19, and normal cases), and basic data augmentation techniques were applied to account for over-fitting. The experiment used various pre-trained standard models, including Vgg16 and 19, ResNet, DenseNet, and Inception. The result showed superior performance of VGG19 with precision and recall reaching 92% and 90%, respectively (Rahaman et al., 2020). It should be noted, however, that balanced distribution is not often guaranteed in medical data. Therefore these results might not hold in the presence of a biased imbalanced distribution of samples.

2.4 Challenges

Most challenges in AI Medical Images are not considered technical. Among these challenges are the amount of data available, the quality of this data, and the accuracy of the correct classification. The core challenge is the lack of appropriately annotated large-scale data sets (Altaf et al., 2019). The imbalance of samples in datasets is one issue that affects conventional CV tasks. In order to train a model for identifying breast cancer in mammograms, for example, a dataset might only have a small number of malignant samples but a very high number of normal samples.

Sharing medical data is far more complex than sharing other types of information. Effectively, data privacy is both a societal and a technical issue. To restrict its usage or exposure, researchers believe anonymizing patient data will remove substantial problems. In addition, linking the data to a particular individual becomes harder when such details (ID, medical history, name, and age) are discarded (Razzak et al., 2017).

2.5 Conclusion

The literature on machine learning and its relationship to medical imaging was reviewed. Review the technical and logistical challenges of medical imaging and the mechanisms to overcome these obstacles. Lack of data and imbalance are among the most common problems facing researchers. Deep Transfer learning is considered one of the most effective solutions to these obstacles. Generative adversarial networks are the most effective solution for the imbalanced dataset.

Chapter 3

Methods

This chapter describes the project's method. First, build and evaluate an experimental framework using state-of-the-art DL methods. Full mammogram scans will serve as 2D inputs into an image localization; the CBIS-DDSM public breast Cancer Image Dataset will be used to train the model. The model will output a predicted mask for each mammogram. The mammogram images from the backend MOH database will extract and separate to detect the tumor and classify it as malignant or benign. Diagnosis, this dataset will be reprocessed to increase the size of this data, the size of the image, and the label of the images, by data augmentation technique and openCV to resize the images.

3.1 DataSet

The Learning algorithm for detecting breast cancer will be developed using two secondary datasets. The first dataset is an updated and standardized version of the DDSM known as CBIS-DDSM. There are 2,620 digitized film mammography studies in the DDSM. It includes instances with certified pathology data for benign and malignancy conditions. The DDSM is a vital tool for creating and evaluating decision support systems due to the size of the database and ground truth inspection. A trained mammographer chose and curated a subset of the DDSM for the CBIS-DDSM collection. After being decompressed, the images were changed to DICOM format. Additionally included are updated ROI segmentation, bounding boxes, and pathologic diagnosis for training data. Figure 3.1 shows sample mammogram images from CBIS-DDSM dataset. Note that the images in this dataset are stored in the DICOM format, including the patient information in the header of the mammogram images.

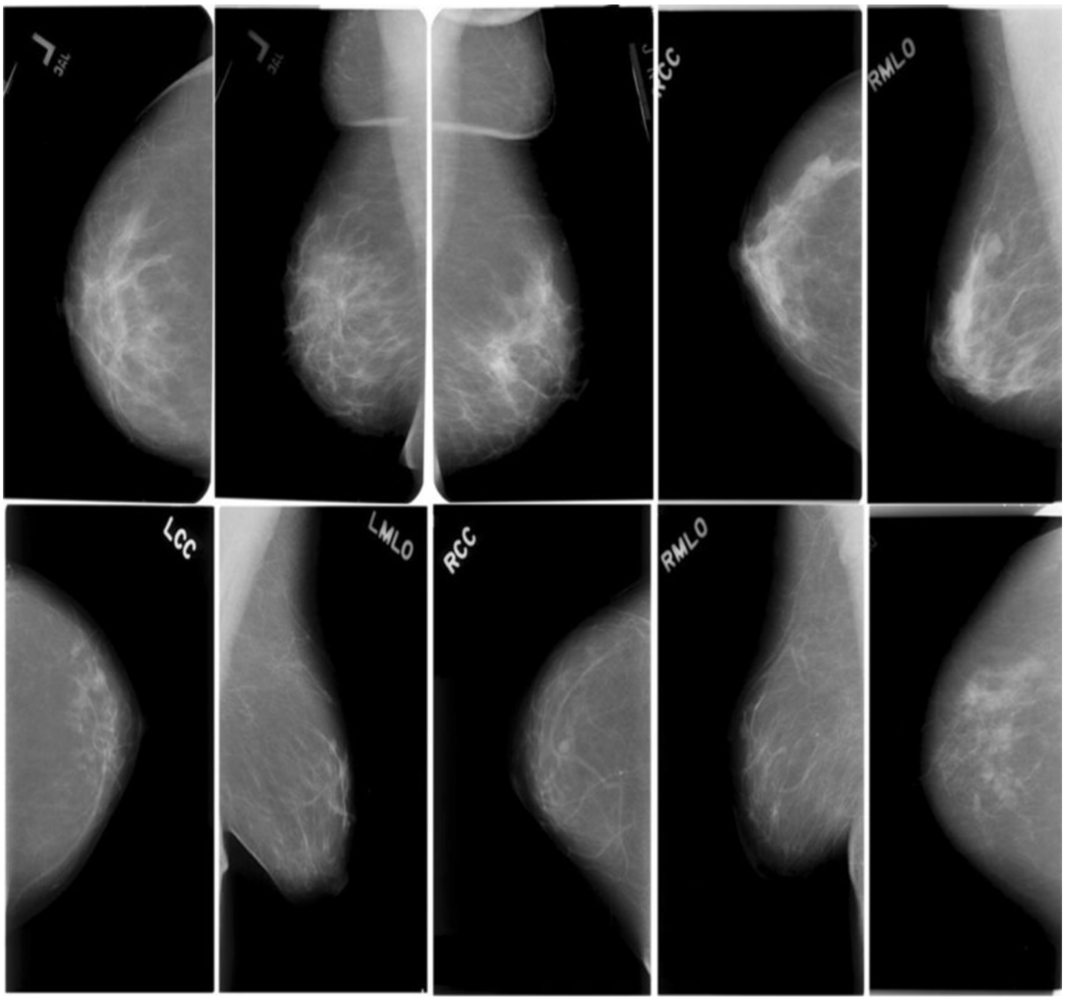


Figure 3.1: CBIS-DDSM: Breast Cancer Image Dataset Samples

The second dataset is MOH dataset, consisting of 783 cases distributed over 2033 images. MOH collected and classified those images from 2017 until February 2022. The diagnoses are based on the international standard framework for breast cancer diagnoses, including mammogram diagnoses and ultrasound checks. The diagnosis results are approved after biopsy tests occur. The cases are divided into 233 patients diagnosed without a tumor, 226 patients with a benign tumor raid two, 163 patients with a benign tumor raid three, and 160 patients with a malignant tumor. The percentage of normal images is 45.6%. The images that contain a malignant tumor are 15.7%. In addition, Up to 38.3% of the images have benign tumors of the first and second types. The dimension of each image is 4510*5844 pixels, and the total size of the dataset is 16 GB. Each image contains a scan of the breast, information about the patient, and the diagnosis result.

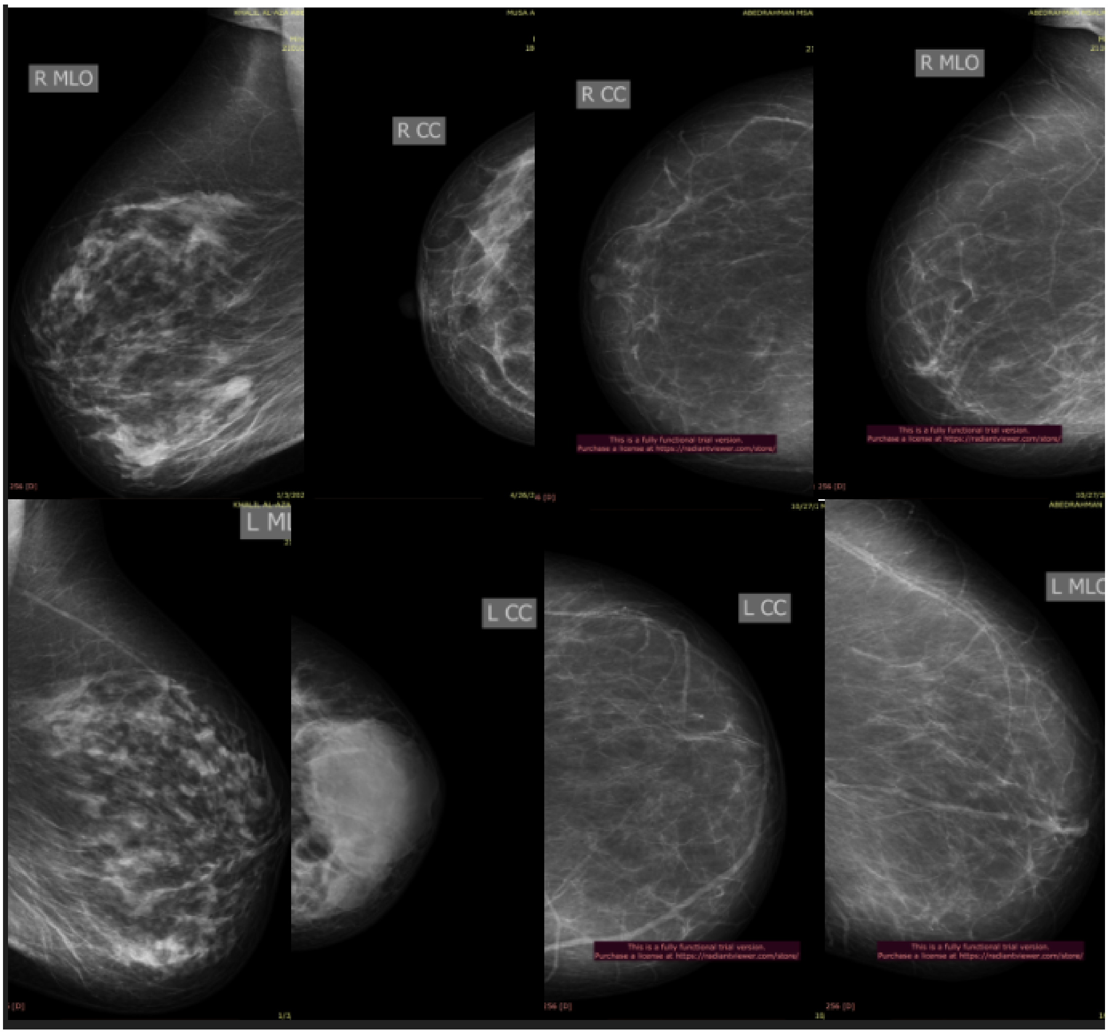


Figure 3.2: MOH Breast Cancer Image Dataset samples

3.2 Data Prepossessing

To prepare the data for the model, we must first enhance its quality, make it simpler for the model to learn characteristics and patterns from the best available data, and concentrate on the area of interest. For instance, we need to prepare two data sets so that we may enter them into the algorithm to train the model. We will remove the border, normalize the photos, remove the artifact, flip the images horizontally, improve the clash, and pad the square for the set of CBIS-DDSM images. In addition, we will remove the border, flip the mask horizontally, and pad it to square for the CBIS-DDSM mask set. Figure 3.3 depicts the area where the border needs to be removed, the artifact, and the location where the image quality enhances the result when we preprocess the images.

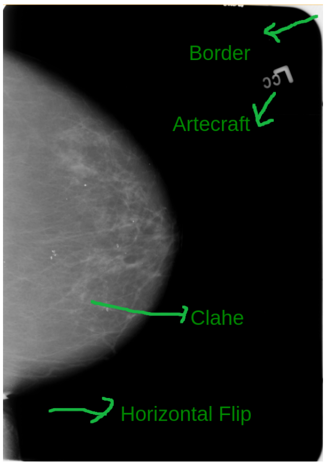


Figure 3.3: CBIS-DDSM Mammogram Images Preprocessing

The MOH data set contains images in png format, so the personal information is displayed in the image, and we need to use the Optical character recognition (OCR) algorithm to detect the sentence and numbers in the image to remove this extra information. In addition, we will remove the border, artifact, horizontal flip, clash enhancement, and the pad in the square. Finally, resize the images to 224*224, normalize, and standardize all images to fast data access during the training and evaluation phase. Figure 3.4 shows the sample from the MOH data set, the information we should remove, and the weakness we should improve, such as the artifact and the clahe.

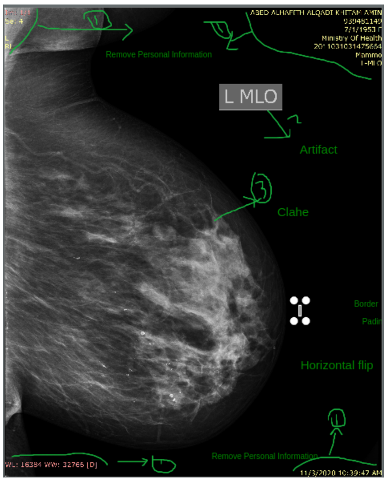


Figure 3.4: MOH Mammogram Images Preprocessing

The calcification and mass categories are two important breast cancer categories in CBIS-DDSM. The complete images, the cropped images, and the mask where the cancer is identified are all included in each case. As a result, the folder structure must be constructed based on the cancer category; Figure 3.5 illustrates the suggested layout for CBIS-DDSM. Folders containing the diagnosis information, patient information, and the png image are located in the MOH data collection. The information needs to be separated into two groups: normal and abnormal. The photos were divided according to the diagnosis in each division. The suggested data format for the MOH data collection is depicted in Figure 3.6.

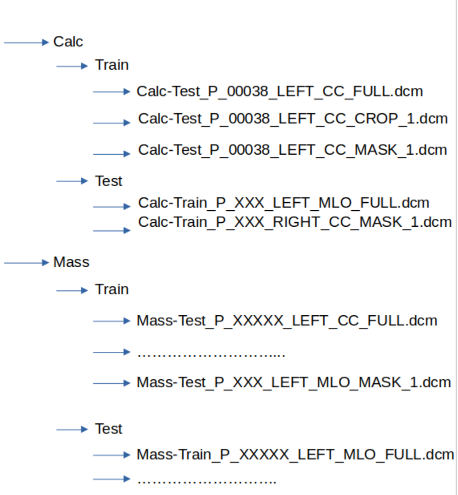

Calc
Train
Calc-Test_P_00038_LEFT_CC_FULL.dcm
Calc-Test_P_00038_LEFT_CC_CROP_1.dcm
Calc-Test_P_00038_LEFT_CC_MASK_1.dcm
Test
Calc-Train_P_XXX_LEFT_MLO_FULL.dcm
Calc-Train_P_XXX_RIGHT_CC_MASK_1.dcm
Mass
Train
Mass-Test_P_XXXXX_LEFT_CC_FULL.dcm
....
Mass-Test_P_XXX_LEFT_MLO_MASK_1.dcm
Test
Mass-Train_P_XXXXX_LEFT_MLO_FULL.dcm
....

Figure 3.5: CBIS-DDSM Proposed Data Structuer

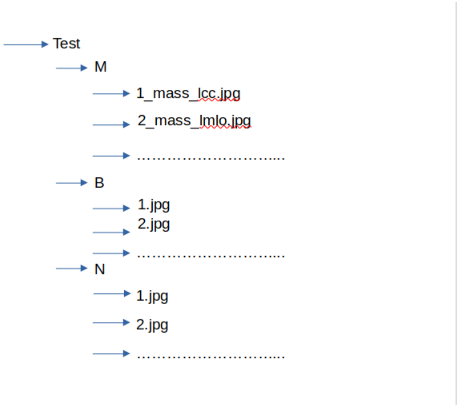


Figure 3.6: MOH Proposed Data Structuer

3.3 Learning Process

The primary research objective of this article is to build a web application portal to upload and classify medical images of breast cancer and host the web application in the MOH. The proposed project workflow is produced in Figure 3.7 to achieve this goal. We must train two models in two different data sets to achieve what we expected. The primary function of model one is to classify the image as normal or abnormal. The role of model two is to classify and detect the tumor in images. For training of the models, the DL algorithm uses a set of breast cancer input images to identify the image features that, when used, will result in the correct classification of the image, that is, defining benign or malignant tumor as compared with the provided labels for these input images. CNNs are created to determine the best filter weight values. CNNs allows the network to discover patterns. Dropout is deployed with CNNs to prevent over-fitting during the training of the models, and units are randomly dropped along with their connections.

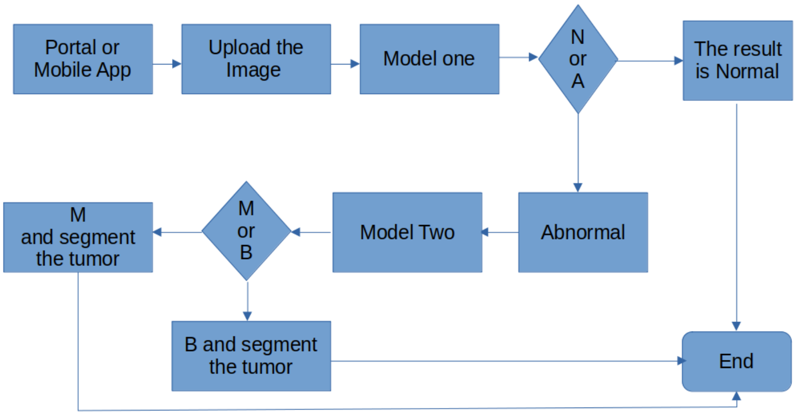


Figure 3.7: Project Workflow

A web application portal will be created using Streamlit. Streamlit is an open-source Python app framework that simplifies creating custom web apps for machine learning and data science. It is compatible with major libraries like Sci-kit-learn, Keras, PyTorch, NumPy, Pandas, and Matplotlib. Streamlit's ease of use and simple API make it ideal for small data or prototyping larger apps. However, it has limited third-party support, making it a newer framework with limited third-party support. A framework will be made according to Figure 3.7. The first and second models on the portal will be deployed.

Model one learning starts with binary classification (Normal / Abnormal) without considering magnification. We Split the dataset into Train (80%) and Test (20%). Then we use pre-trained models or transfer learning without fine-tuning as a baseline. VGG19, ResNet101, and InceptionV3 for deep computer vision classification apply to the training. We are exploring, fine-tuning, data augmentation, and comparing with the baseline. If the result is unacceptable, we will use Generative Adversarial Networks (GANs) to generate more data and then apply and reevaluate. Figure 3.8 show the high-level workflow for the model training. The MOH data set will train model one for normal and abnormal detection.

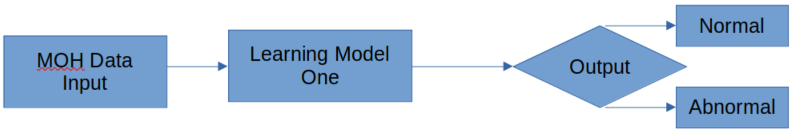


Figure 3.8: Model One Learning Workflow

The CBIS-DDSM data set will use to learn model two to classify and segment the tumor in the abnormal images. The U-Net is a convolution neural network developed for biomedical image segmentation and will be used to detect and segment cancer in mammogram images. U-Net Has two main parts, the encoder block and the decoder block. The encoder block learns features from mammogram images and classifies the image as malignant or benign. The decoder block then presents these low-resolution features discovered by the Encoder block back up to the full resolution of the original input image, and the output is a full-resolution segmentation map. The Encoder block is a sequence of convolutional layers and max-pooling layers. Vgg16 was selected for the encoder block without the final dense layers. The Encoder learns from the first layers of CNNs' low-level features, while the decoder learns from the final layers of high-level features; This allows the U-Net to map components extracted across all levels into the final full-resolution, predicted segmentation mammogram images. Figure 3.9 shows the U-Net algorithm's architecture and the Encoder and decoder's function.

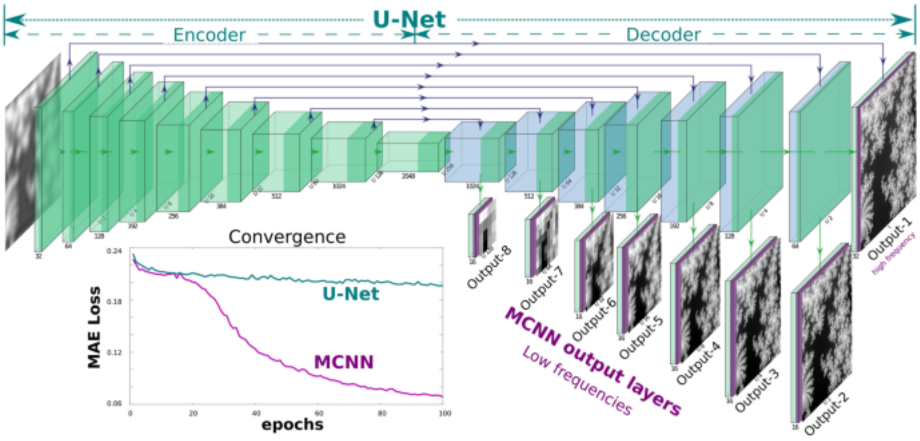


Figure 3.9: U-Net Architecture (Wang et al., 2020)

The CBIS-DDSM data set is used to learn model two and extract information and feature, as shown in Figure 3.10. The learning phase uses U-Net and Vgg16 as CNNs. The output of the classification is malignant or benign and segments the tumor. The learning mode used to learn the two models is detecting the normal and abnormal cases and classifying the abnormal images as malignant and benign using MOH and CBIS-DDSM to segment the tumor location.

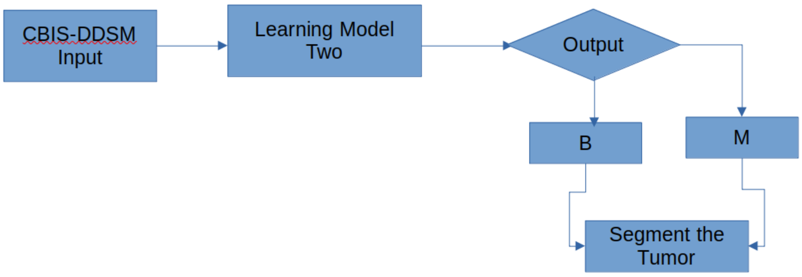


Figure 3.10: Model Two Learning Workflow

3.4 Model Evaluation

The most critical phase is evaluating the model learning to improve the performance of the learning phase. In the project structure Figure 3.7, there are two models. Each model has specific evaluation criteria to evaluate the learning performance and prediction. Model one is a pure classification problem, and multi-literature is related. Accuracy for model one will calculate as the sum of two accurate predictions (TP + TN) divided by the total number of data sets (P + N), which means accuracy = $(TP + TN)/(P + N)$. The disadvantage of accuracy is that it masks the issue of class imbalance, so we will calculate the Recall as the number of accurate positive predictions (TP) divided by the total number of positive (P), which means Recall = $(TP)/(TP + FN)$. Precision will calculate the number of correct positive predictions (TP) divided by the total number of positive predictions (TP + FP), which means Precision for model one = $(TP)/(TP + FP)$. F-score is a measure of the accuracy of the test. It is calculated based on Precision and reminders. The final method to evaluate model one is the ROC curve, a graph that visualizes the trade-off between True Positive Rate and False Positive Rate.

$$F - score = \frac{2 * Precision * Recall}{Precision * Recall} \tag{3.1}$$

The evaluation criteria for model two is more complicated because the literature in biomedical computer vision segmentation is a little, and we need accuracy in classifying abnormality and predicting the tumors. Intersection over union (IoU) is a good metric for measuring overlap between two bounding boxes or masks. If the prediction is correct, then IoU = 1. Equation 3.2 displays IoU is the percentage overlap between the predicted segmentation and the ground truth, divided by the union area between the predicted and ground truth. We will use IoU to evaluate the accuracy of the prediction of tumors in abnormal images.

$$IoU = \frac{Area_of_Predicted_mask \cap Area_of_Truth_mask}{Area_of_Predicted_mask \cup Area_of_Truth_mask} \tag{3.2}$$

3.5 Conclusions

The main conclusions for this chapter are that building a system to predict breast cancer, whether it is normal or abnormal, and to determine the location of cancer in case of infection requires training two models, one to classify patients as infected or not infected, and the MOH data set will be used in this phase. The F-score will be used to evaluate and increase the model one efficiency. Vgg16 algorithm and transfer learning will also be used to build the model. The second model will classify infected cases as benign or malignant cancer and determine the tumor's location. The IoU will also be used to evaluate model two and improve its efficiency, and the U-Nit algorithm and transfer learning will be used to build model two. The CBIS-DDSM data set will be used to build model two.

The CBIS-DDSM data set was used because the MOH dataset does not have a classification of cancer location or mask to detect the tumor location in the mammogram images, and its size is insufficient. Building a segmentation model to classify and detect breast cancer requires specifying the area of cancer on the mammogram images to make this system

Chapter 4

Implementation

This chapter looks at how the project was carried out. The mammogram images are preprocessed in the data sets. U-Net and Vgg16 are used for image classification and segmentation. Models are evaluated to improve performance. The web portal is built to process all the classification and segmentation of mammogram images.

4.1 Dataset

Breast cancer mammograms and chest scan images from 2017 to 2022 are needed with the doctor's diagnosis for each case. The sample size we need to build CNNs to detect and localize the abnormality in the image is 2000 images selected from 4518 patients from the MOH dataset in addition to CBIS-DDSM. The first dataset is stored in the backend system in the MOH. The research output will be implemented in MOH as a plugin in the IBNSENA Health information system. The dataset we will process will not contain identifying patient information. MOH data sets contain labels for each image, which will be used to classify the mammogram into normal that contains BI-RADSs one or abnormal cases that classify one of the class from 2 to 5, as shown in Figure 4.2 and Table 4.1. The CBIS-DDSM dataset contains a mask for localizing the tumor, which will be used to locate the tumor in the abnormal mammogram images, as shown in Figure 4.1.

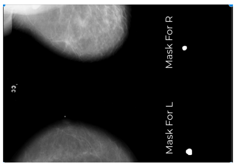


Figure 4.1: CBIS-DDSM Mask

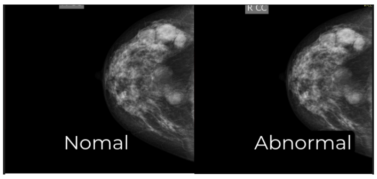


Figure 4.2: MOH Sample label

#	Date	Age	Mammogram Result	BI-RADS	Date of Birth
0	2022	58	Scattered fibro glandular breasts tissue.No de...	1	Mar 01.1964
1	2022	53	Scattered fibro glandular breasts tissue.No de...	1	Feb 21.1969
2	2022	54	Scattered fibro glandular breasts tissue.No de...	1	Dec 13.1968
3	2022	59	Scattered fibro glandular breasts tissue.No de...	1	Sep 14.1963
4	2022	76	Scattered fibro glandular breasts tissue.No de...	1	Apr 5.1946

Table 4.1: MOH DataSet Sample

Table 4.2 shows the description and class for Breast-Imaging Reporting and Data System (BI-RADS) from one to six, the selected cases from the MOH dataset contain BI-RADS one to five.

#	LN	Result	Description
0	0	Incomplete	Additional imaging evaluation to prior mammograms is needed
1	I	Negative	Normal
2	II	Benign	Non-cancerous finding
3	III	Probably benign finding	Follow-up in a short time frame is suggested.
4	IV	Suspicious abnormality	Biopsy should be considered
5	V	Highly suggestive of malignancy	Appropriate action should be taken
6	VI	Known biopsy-proven malignancy	Appropriate action should be taken

Table 4.2: Mammogram Results Description

There are no null parameters in the MOH dataset. As shown in Table 4.3, there are 4518 cases in the MOH dataset.

#	Column	Non-Null	Count	Dtype
0	Number	non-null	4518	int64
1	Date	non-null	4518	int64
2	Name of Patient	non-null	4518	object
3	Age	non-null	4518	int64
4	Mammogram Result	non-null	4518	object
5	BI-RADS	non-null	4518	object
6	Date of Birth	non-null	4518	object

Table 4.3: Parameters Description

Table 4.4 show the description of the statistics for the Age and Date parameters, as shown in the table the minimum age in the dataset is 22 and the maximum is 93, both of them are outlier.

	count	mean	std	min	25%	50%	75%	max
Date	4518.0	2018.636565	1.415249	2017.0	2018.00	2018.0	2020.00	2022.0
Age	4518.0	52.963701	9.295913	22.0	46.00	52.0	59.00	93.0

Table 4.4: Age & Date Description

The normal Age mean is 51.42, the benign age mean is 56.85, the benign callback age mean is 51.67 =, and the malignant age mean is 53.06. and the advanced malignant age mean is 57.5. there is no big difference between the means in different classes. Figure 4.3 shows the box plot for age in all classes. There is an outlier above 80 and under 28 years old. All the samples used are from Bethlehem governorate.

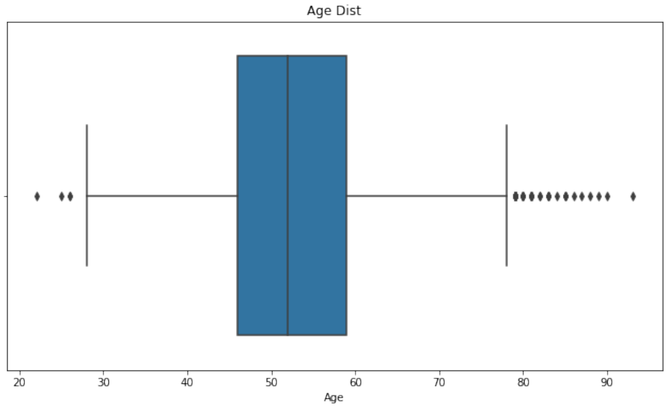


Figure 4.3: Box Plot for Age

	patient_id	breast density	left or right breast	image view	abnormality id	abnormality type	calc type	calc distribution	assessment	pathology	subtlety	image file path	cropped image file path	ROI mask file path
0	P_00005	3	RIGHT	CC	1	calcification	AMORPHOUS	CLUSTERED	3	MALIGNANT	3	Calc-Training	Calc-Training	Calc-Training
1	P_00005	3	RIGHT	MLO	1	calcification	AMORPHOUS	CLUSTERED	3	MALIGNANT	3	Calc-Training	Calc-Training	Calc-Training
2	P_00007	4	LEFT	CC	1	calcification	PLEOMORPHIC	LINEAR	4	BENIGN	4	Calc-Training	Calc-Training	Calc-Training
3	P_00007	4	LEFT	MLO	1	calcification	PLEOMORPHIC	LINEAR	4	BENIGN	4	Calc-Training	Calc-Training	Calc-Training
4	P_00008	1	LEFT	CC	1	calcification	NaN	REGIONAL	2	BENIGN	3	Calc-Training	Calc-Training.	Calc-Training

Table 4.5: CBIS-DDSM Data Sample

Table 4.5 describe the data function, the breast type, and the tumor class. This data has two major categories: one for mass tumors and the second for calcification tumors.

	breast density	abnormality id	assessment	subtlety
count	1546.000000	1546.000000	1546.000000	1546.000000
mean	2.663648	1.415265	3.258732	3.411384
std	0.937219	0.903571	1.229231	1.179754
min	1.000000	1.000000	0.000000	1.000000
25%	2.000000	1.000000	2.000000	3.000000
50%	3.000000	1.000000	4.000000	3.000000
75%	3.000000	1.000000	4.000000	4.000000
max	4.000000	7.000000	5.000000	5.000000

Table 4.6: CBIS-DDSM Stat Description

The statistical description shown in Table 4.6, breast density is the essential function because it affects the image quality. The general information about the parameter null count and the type of data related to these parameters are shown in Table 4.7. there are three main types of data, int64, category, and object data type. Abnormality_type function to classify if the tumor is calcification or mass. Pathology to detect the type of cancer, if it is malignant or benign.

	Column	Non-Null Count	Dtype
0	patient_id	1546 non-null	object
1	breast_density	1546 non-null	int64
2	left_or_right_breast	1546 non-null	category
3	image_view	1546 non-null	category
4	abnormality id	1546 non-null	int64
5	abnormality_type	1546 non-null	category
6	calc_type	1546 non-null	category
7	calc_distribution	1546 non-null	category
8	assessment	1546 non-null	int64
9	pathology	1546 non-null	category
10	subtlety	1546 non-null	int64
11	image file path	1546 non-null	object
12	cropped image file path	1546 non-null	object
13	ROI mask file path	1546 non-null	object

Table 4.7: CBIS-DDSM Parameters

4.2 Data Prepossessing

MOH dataset contains images in jpg format that will be less flexible, and it takes more time to improve the images. From Figur 4.3, we note that the images include personal information; there is Clahe in the images, and the artifact must be removed. In addition, there is a text above the image that must be removed.

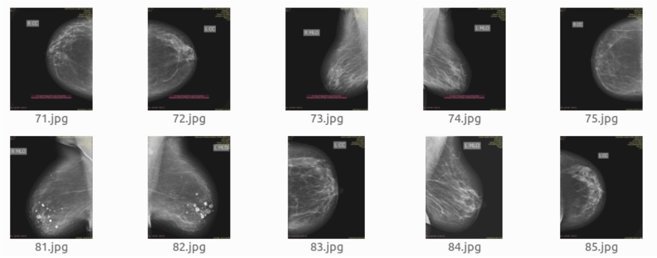


Figure 4.4: Raw Data for MOH Dataset

First of all, to preprocess the images, we crop the border from the two sides 1% and from up and bottom of the images 4%. Normalize the images to modify the data of each channel/tensor so that the mean is zero and the standard deviation is one. MinMax normalization using equation 4.1. Removing artifacts and enhancing Clahe is one of the most critical phases in image enhancement. The Sortconturbyarea function detects the most significant area and removes the other. The Sortconturbyarea function computes the bounding rectangle for each contour and outputs the sorted contours. xLargestBlobs function finds the largest contours. The applyMask function applies the mask that detects from xLargestBlobs to the image. The White area is kept, and the black will remove. Finally, the checkLRFlip function will check the direction of the image horizontally, if it's right or left, and make it right.

$$X_(norm)^i = \frac{X^i - X_(min)}{X_(max) - X_(min)} \tag{4.1}$$

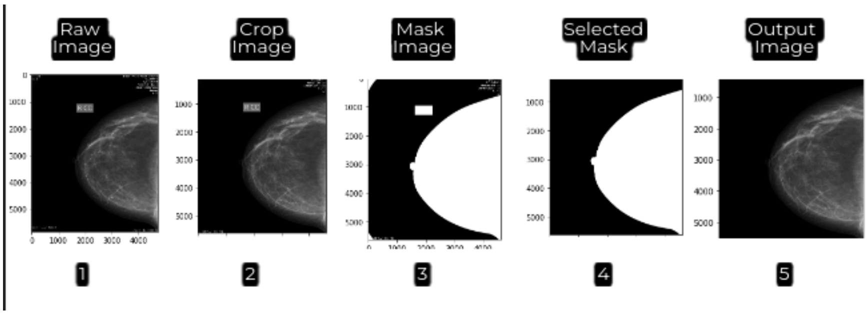


Figure 4.5: Image Preprocessing

Figure 4.4 shows the workflow procedure of image preprocessing; we note five major editing phases. The selected mask is most vital because it selects the area of interest and removes unwanted things from the image.

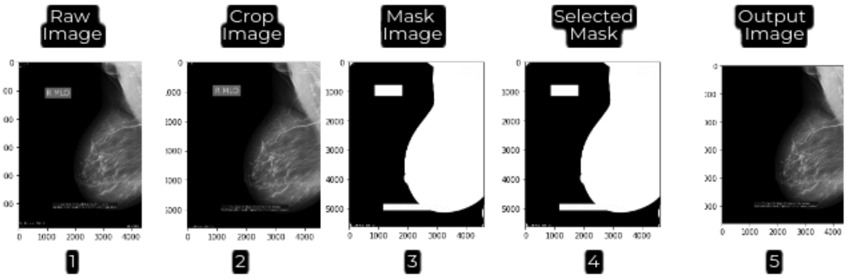


Figure 4.6: Image Preprocessing

We notice from Figure 4.5 that the selection of the region or applied mask did not include the region that we want to study, which means the selected mask contains the region of interest in addition to the bounding box that includes text, and this is due to the overlap of this region. so text remover program build using keras_ocr to remove the text from the images. After removing the text, some images need manual improvement, so the data set is split into two data sets. The first data set is ready to learn the model and includes 898 images for abnormal and 766 for normal Figure 4.6; the other needs manual manipulation to be ready and includes 201 images for abnormal and 181 for normal cases Figure 4.7.

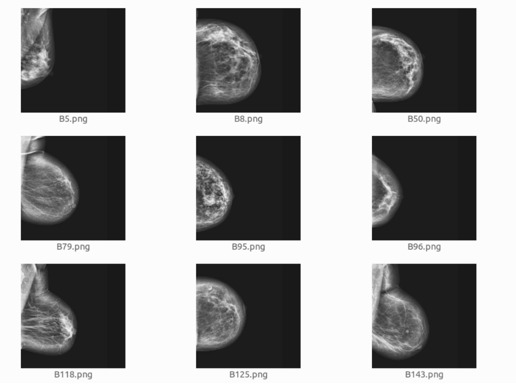


Figure 4.7: DataSet One

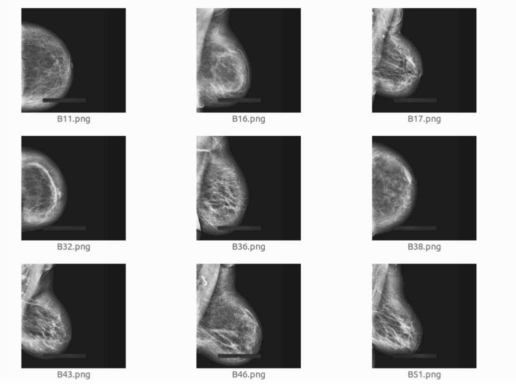


Figure 4.8: DatSet Two

After manual editing with the Photoshop application for data set two, we build Data Set Three, which includes 898 images for abnormal and 1421 for normal cases. The split folder method divides dataset one into three folders: test folders, which contain 75% of images, and the val folder, which includes 5% of cases. Moreover, the test folder contains 15% of cases, and the same task was done for dataset three.

4.3 Learning Process

The learning process is based on the quality of the data. More than one data set and multiple learning models have been created. The first data set contains all the images, but some must be processed. The first model was built, which consisted of Two Convolution layers, one max pol layer, and two fully connected layers. We use Cross Entropy Loss as our loss function. We use Stochastic Gradient Descent (SGD) with a learning rate of 0.001 for our gradient descent algorithm or Optimizer. We set the momentum to 0.9. Equations 4.2 and 4.3 show how the gradient is calculated. Model one learn from the unspilt data set, dataset one and three. Each learning phase is a separate experiment.

$$ch_x(t) = StepSize * f'(x(t-1)) + mom * ch_x(t-1) \tag{4.2}$$

$$x(t) = x(t-1) - ch_x(t) \tag{4.3}$$

The model one network structure is as follows:

- (conv1): Conv2d(3, 32, kernel_size=(3, 3), stride=(1, 1))
- (conv1): Conv2d(3, 32, kernel_size=(3, 3), stride=(1, 1))
- (conv2): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1))
- (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
- (fc1): Linear(in_features=1000000, out_features=128, bias=True)
- (fc2): Linear(in_features=128, out_features=4, bias=True)

Transfer learning used in building models used ResNet-18, ResNet-50, Vgg16, and inception_v3 pre-trained model. ResNet-18 is a convolutional neural network that includes 18 layers deep, the pre-trained version of the network trained on more than a million images from the ImageNet database. The pre-trained network can classify images into 1000 object categories, such as keyboard, pencil, and many animals. ResNet-50 is a convolutional neural network that is 50 layers deep. The pre-trained version of the neural network was trained on more than a million images from the ImageNet database. Vgg16 is a CNNs that is 16 layers deep. Inception-v3 is a pre-trained CNNs model that is 48 layers deep. All pre-trained models were downloaded, and the last fully connected layers were edited to classify two outputs. In the learning phase, we update the weight in the last fully connected layer. In another way, we update the weight for all by turning off the requires_grad that will not compute backward.

The portal was built using Streamlit. The first and second models were deployed on the portal. We were creating visually appealing apps without front-end development. Streamlit releases us from focusing on any front-end framework or coding in Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript.

4.4 Model Evaluation

Model learning evaluation is one of the most critical stages in building AI models to build a model capable of predicting with high accuracy and without overfitting. Cross entropy loss and SGD were used to evaluate the ML model through the learning phase by applying the two techniques in the test data baches. Several tools, including cross-entropy loss, measure ML performance of a classification model. In cross-entropy, the loss is expressed as a number ranging from 0 to c, where c is the number of classes. The class is 0 or 1 in our model, with 0 being a perfect model. The goal is generally to get the model as close to 0 as possible. Cross entropy loss measures the difference between the discovered probability distribution of a machine learning classification model and the predicted distribution. The mathematics formula that calculated the loss or error displayed in Equation 4.4 where y is normal class, $1 - y$ is abnormal class, and p is the probability of normal class.

$$l = -(y * \log(p) + (1 - y) * \log(1 - p)) \tag{4.4}$$

Gradient Descent is an optimization method that seeks the optimal value for an objective function. Finding the model parameters that produce the highest levels of accuracy on training and test datasets is the main objective of gradient descent. A variation of the Gradient Descent approach used to enhance machine learning models is called SGD. It tackles the classic Gradient Descent methods' computational inefficiencies while working with massive datasets in machine learning applications. In SGD, the gradient is computed, and the model parameters are updated using only a random training example (or a small batch) rather than the entire dataset for each iteration.

4.5 Conclusions

The most important conclusion in this finding is the success of extracting a dataset of images for breast cancer mammograms that can build an intelligent model to predict cases normal or abnormal for patients. The extracted dataset can be relied upon as a classifier but not to build a model that can detect and segment the tumor's location in the breast. This data needs more work to identify the tumor in each image, and then it can build a machine-learning model to predict and segment the tumor's location. CBIS-DDSM is a public dataset in which a learning model can be built to detect the tumor's location. However, due to time constraints and because of the consumption of most of the time in collecting and improving the Ministry of Health dataset and many studies on this public dataset, it was sufficient to study the dataset of the Ministry of Health. Also, to build a good learning model, more than one part of the dataset in the Ministry of Health was adopted, and more than one learning method was adopted in building the model.

Chapter 5

Experiments & Results

This chapter assesses the project as a whole and presents the test findings. Six experiments were conducted to achieve the best results and build the most accurate model. This chapter will review each experiment, and its effects will be discussed.

The model one was built from the ground up in the first experiment. The model consists of two Conv layers, two Max layers, and two fully connected layers. The unseparated dataset has been processed. Figure 5.1 shows the data set sample used in the experiment, as we notice that some of the images contain borders that still need to be processed. Table 5.1 shows that the validation loss in this experiment is 1.33, and the validation accuracy is 33.77 %. This result is unacceptable because of the model structure or the dataset used in the learning. To examine the reason for this result, we separated the dataset used in this experiment into two datasets, the first containing well-processed images and the second containing the dataset that needed manual processing.

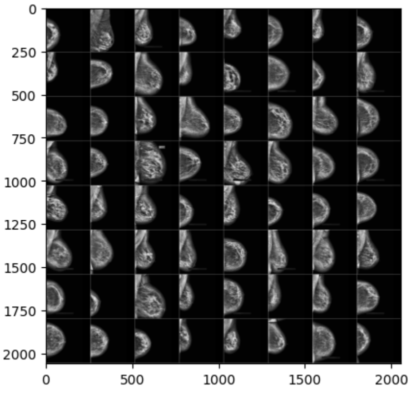


Figure 5.1: Unsplit Dataset

Because the results from the first experiment were not accepted, the second experiment was conducted with the same model structure. The first database has been adopted, which contains well-processed images. Figure 5.2 shows the sample images from dataset one. The model accuracy in the training phase is 53.9%, and the accuracy in the test phase is 53.6 %. The training loss is 0.68, and the test loss is 0.7. This result shows that it is an improvement but needs revision. The reason is that when we split the data, we removed the images containing borders but decreased the number of images in the dataset. So we will manually reprocess the unclear image and added to the dataset.

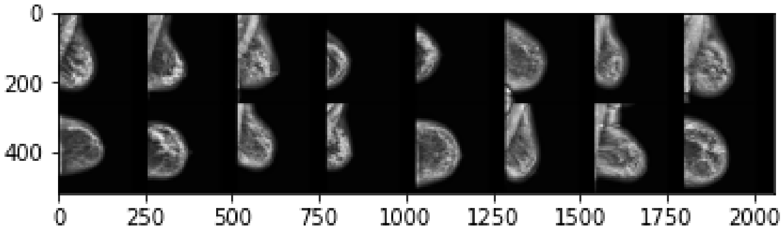


Figure 5.2: Dataset One

The third experiment was conducted to enhance the model's performance and prepare the dataset in which the images were processed automatically and manually. This group contains 1065 normal and 673 abnormal mammogram images in the training folder. Figure 5.3 shows the sample of dataset three, and we note that some of the images are cropped. The validation accuracy in this experiment is 74.46 %. The validation loss is 0.474, which is an acceptable result, and we note the excellent improvement in model performance after enhancing the dataset.

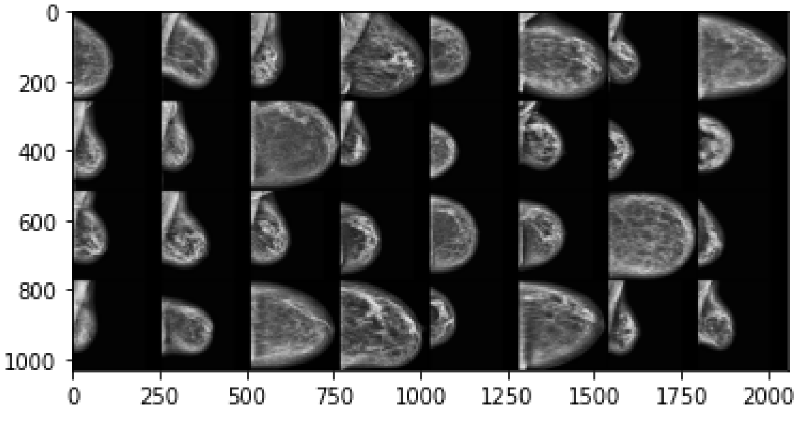


Figure 5.3: Dataset Three

The following three experiments are to improve the model's structure and performance.

Transfer learning will be used in the model creation process. Four models have been used: Resnet18, Vgg16, and Inception_v3, in addition to Resnet50. The last layer in each model will be modified, and the output will change from 1000 classes to two classes, normal and abnormal. The weights update was done in two ways, the first method was to update the weights of the last layer, and the second was to update the weights on all layers. The third database is used to learn the model. As we can see from the results in Table 5.1, the best model is Vgg16 with a validation loss of 0.4271 and validation accuracy of 83.26%.

Models	Train Loss	Test Loss	Train Acc	Test Acc	Dataset
Model_one	1.36	1.33	32.02	33.77	unsplit dataset
Model_one	0.68	0.7	53.9	53.6	Dataset 1
Model_one	0.4854	0.474	74.51	74.46	Dataset 3
Resnet18	0.2605	0.3937	89.41	80.69	Dataset 3
Resnet18 no grad	0.5027	0.4658	71.69	76.61	Dataset 3
Vgg16	0.1724	0.4271	93.96	83.26	Dataset 3
Inception_v3	0.2682	0.4085	0.8746	0.8197	Dataset 3
Inception_v3 no grad	0.5130	0.4778	7261	7425	Dataset 3
Resnet50	0.1024	0.5552	96.84	75.97	Dataset 3
Resnet50 no grad	0.4816	0.4879	74.11	72.53	Dataset 3

Table 5.1: Models Training Result

Table 5.2 shows the evaluation result for the test and validation data set for Vgg16 model. We also note that the accuracy of the validation data is less than the accuracy of the test data.

Func	Val Data	Test Data	AV
True Negative	27	150	
False Positive	17	31	
False Negative	8	54	
True Positive	63	231	
Dataset Accuracy(mean)	78.26 %	81.75 %	80.005 %
Sensitivity	88.73 %	81.05 %	84.89 %
Specificity	61.36 %	82.87 %	72.115 %
Precision	78.75 %	88.16 %	83.455 %
NPV	77.14 %	73.5 %	75.32 %
F1	83.44 %	84.46 %	83.95 %

Table 5.2: Confusion Matrix for Vgg16 Model

Figures 5.4 and 5.5 show the accuracy and loss graph during the training phase. The chart shows that the model overfits after eight epochs, so the best parameters weight will be in epoch seven.

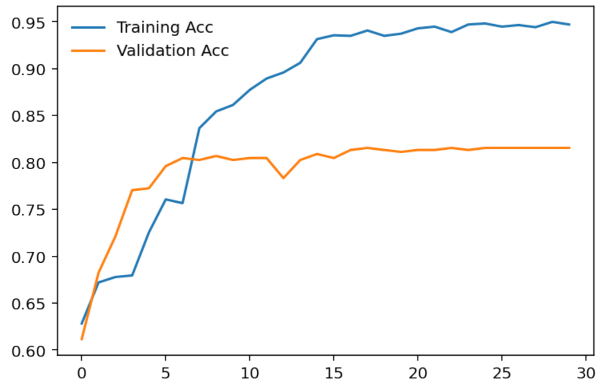


Figure 5.4: Vgg Accuracy Graph

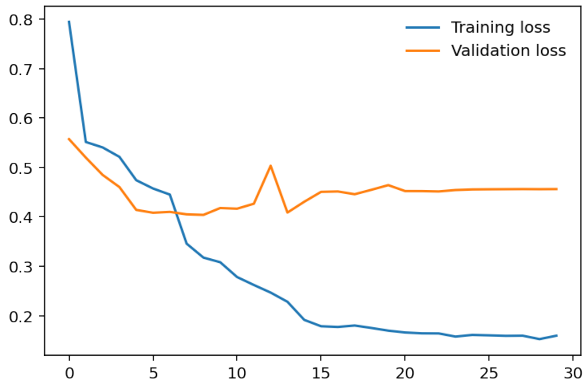


Figure 5.5: Vgg Loss Graph

Table 5.3 shows the evaluation result for the test and validation data set for the Resnet18 model. We note that the NPV is not high according to other values because the dataset is not balanced, and the normal class is more than 60% of the data set.

Func	Val Data	Test Data	AV
TN	29	148	
FP	15	33	
FN	16	56	
TP	55	229	
Testset Accuracy	73.04	80.9	76.97
Sensitivity	77.46	80.35	78.905
Specificity	65.9	81.76	73.83
Precision	78.57	87.4	82.985
NPV	64.44	72.54	68.49
F1	78.01	83.7	80.855

Table 5.3: Confusion Matrix for ResNet Model

For more deep evaluation, the MIAS public data set was selected. MIAS contains 207 normal mammogram images and 155 abnormal mammogram images. Table 5.4 show the head of information for the selected dataset. The mammogram images were converted from .pgm to jpeg. The overall preprocess phases are conducted on all the datasets. The data set is split into Normal and abnormal folders.

REFNUM	BG	CLASS	SEVERITY	X	Y	RADIUS
mdb001	G	CIRC	B	535	425	197
mdb002	G	CIRC	B	522	280	69
mdb003	D	NORM	NaN	NaN	NaN	NaN
mdb004	D	NORM	NaN	NaN	NaN	NaN
mdb005	F	CIRC	B	477	133	30

Table 5.4: head of the information for MIAS

Figure 5.6 shows the Mias dataset sample and the related class for each mammogram image in the text.info file. The Mias dataset is just used for evaluation, which means the model didn't train for this dataset.

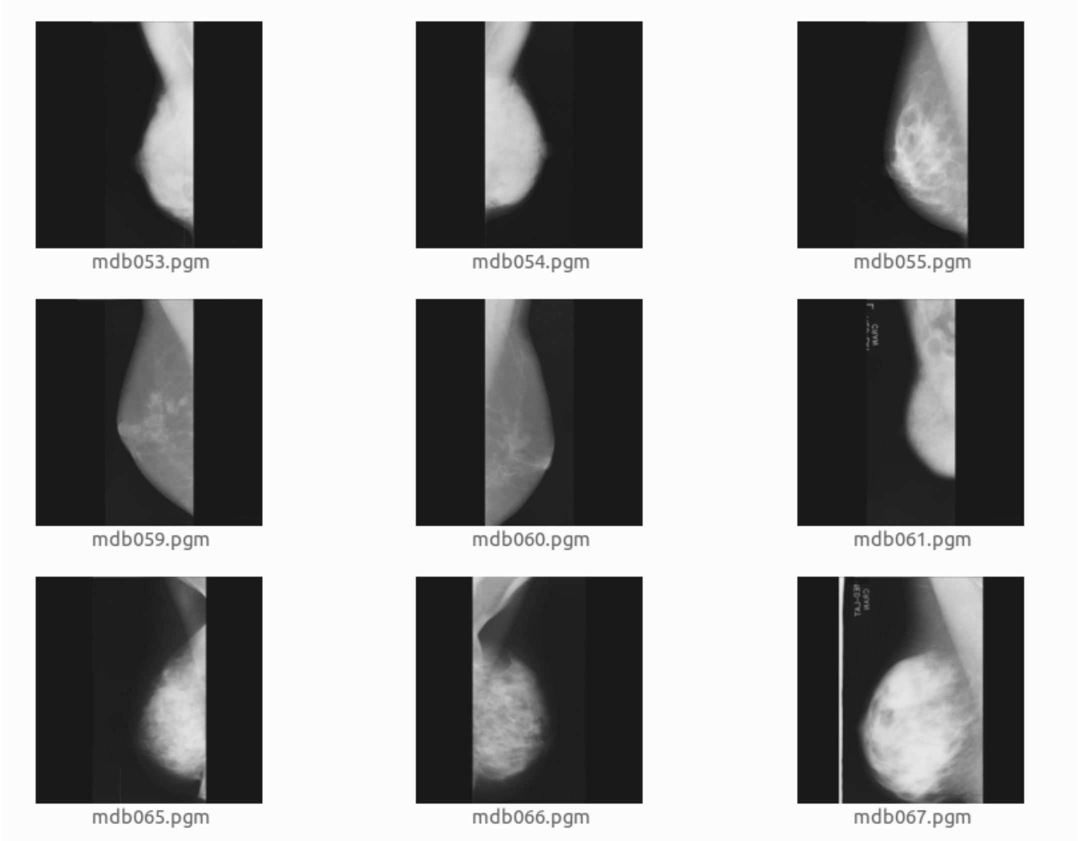


Figure 5.6: Mias DataSet Sample

Table 5.5 shows the confusion matrix result for the VGG16 model on the MIAS dataset. The overall result decreasing. the Testset Accuracy is 57.14%, the Precision is 69.71 %, and the F1 score is 63.87%.

Func	Mias Data
TN	62
FP	53
FN	85
TP	122
Testset Accuracy	57.14
Sensitivity	58.93
Specificity	53.91
Precision	69.71
NPV	42.17
F1	63.87

Table 5.5: Confusion Matrix for MIAS Dataset

5.1 Conclusions

We conclude from this chapter that data quality, classification, tabulation, and suitability in terms of quantity and quality is the most critical factor in raising the model's efficiency. Building the model can start from scratch or using a pre-trained model. This chapter evaluates the overall project and provides the results of tests. Six experiments were conducted to reach the best results and build the most accurate model. Model one was built from the ground up with two Conv layers, two Max layers, and two fully connected layers. The validation loss was 1.33, and the accuracy was 33.77 %. The second experiment was conducted with the same model structure, and the model accuracy was 53.9 % and 53.6 %. The training loss was 0.68, and the test loss was 0.7. The third experiment was conducted to improve the model's performance and prepare a dataset with 1065 normal and 565 abnormal mammogram images. The validation accuracy was 74.46 %, and the validation loss was 0.474. The last three experiments used transfer learning and four models to improve the model's structure and performance. The best model was Vgg16, with a validation loss of 0.4271 and a validation accuracy of 83.26%. In this project, we notice a massive improvement in the model performance when we enhance data equality. The F1 Score decreases to 63.87% when the MIAS dataset is used.

Chapter 6

Conclusion

This chapter summarises the main outcomes and conclusions from this work.

6.1 Conclusions

The main conclusion of this project is that the essential element in building artificial intelligence models is the data collected and the quality of this data in terms of quantity and quality. The database of mammograms in the Ministry of Health is rich in quantity, as it contains data from 2009 until now, but it needs much work to classify, classify, code, and locate tumors. One of the most critical problems the database needs is more stability in classification and storage. The years between 2017 and 2022 were selected, and a database was built from them. The images were processed automatically and manually, and a database containing 2236 mammograms was obtained. It is possible to work on increasing the effectiveness of this data by reviewing it with a specialist in tumors and reclassifying and tabulating it based on the expert's opinion.

The second main output of this project is the ML model that can predict cases of breast cancer patients with or without an infection with an accuracy of 83%, which is a good percentage but needs improvement. In addition to deploying this model on a web application program to help in the work of the MOH, it can also save new mammogram images and classification to help improve the model's accuracy in the future, as the more data, the accuracy will improve for the model.

6.2 Future Work

Many tasks can be done in the future to improve the system, and the most important tasks can be summarized as follows.

- The portal created by Streamlit can be used to collect data, improve its quality, and for initial diagnosis. However, the program must be updated using Django or Flask, and an advanced database should be created to improve system speed and deal with a high demand for diagnosis in the future. In addition to linking the ultrasound, the result of the biopsy, MRI, and the CT scan with the system, the results of these tests are what confirm the presence of the tumor, its size, and its type, and this raises the efficiency of the data, especially the imaging data, and can enhance the results of the diagnosis. Figure 6.1 show the BCD future design.
- It is improving data quality and increasing its number by collecting additional data from private and public centers inside Palestine.
- Using GANs technology to increase the volume of data.
- Using public data to create a model for locating breast cancer.
- Connecting the system with the biopsy result and the ultrasound machine and storing the results in a unified database to improve the system's efficiency and accuracy of the model prediction.

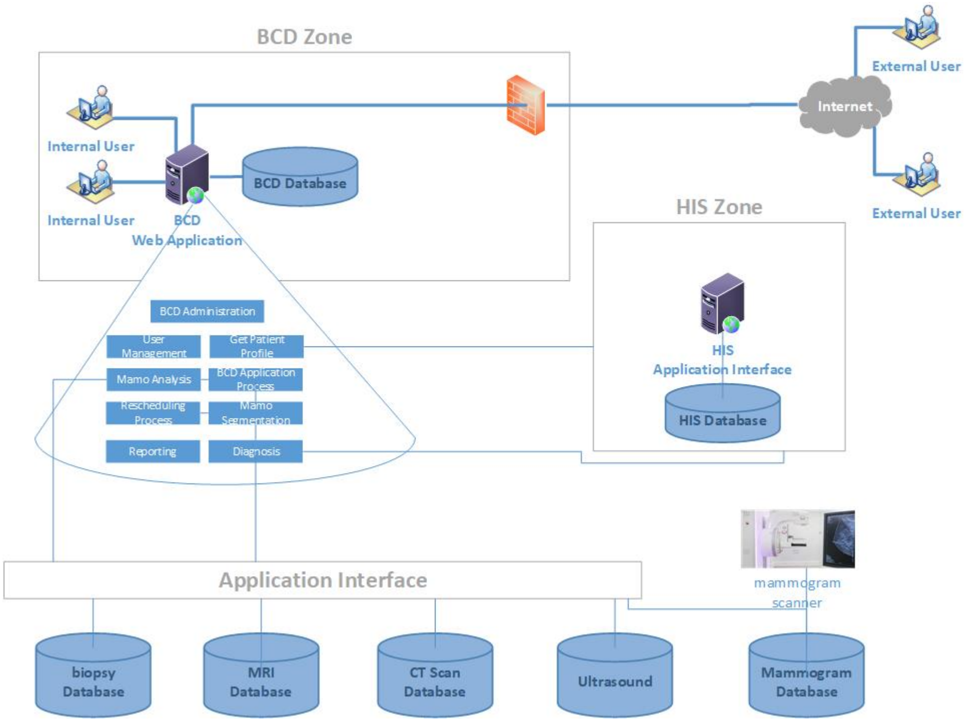


Figure 6.1: Future Design

Figure 6.1 shows the proposed future breast cancer detection BCD system design. We also note from the design that there is application interface was created to be the interface to communicate with databases of medical images from mammograms, CT scans, and MRI, in addition to the ultrasound. The figure shows the BCD and Ibn Sina Medical Information System integration. The proposed system comprises nine parts, including mammogram analysis and segmentation AI model. The BCD starts the process by getting the patient's profile, getting the mammogram images, analyzing the case, then putting the patient in for a biopsy if the result is abnormal, and providing Health Information System (HIS) with the final result. In this design, we will get enhanced and improved breast cancer detection services and high-quality images to improve the BCD AI model accuracy.

Bibliography

Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Classification of COVID-19 in chest x-ray images using DeTraC deep convolutional neural network. *Applied Intelligence*, *51*(2), 854–864. https://doi.org/10.1007/s10489-020-01829-7

Altaf, F., Islam, S. M. S., Akhtar, N., & Janjua, N. K. (2019). Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access*, *7*, 99540–99572. https://doi.org/10.1109/ACCESS.2019.2929365

Bidgood, W. D., Jr, Horii, S. C., Prior, F. W., & Van Syckle, D. E. (1997). Understanding and using DICOM, the data interchange standard for biomedical imaging. *J. Am. Med. Inform. Assoc.*, *4*(3), 199–212.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*, 161–168.

Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, *65*(5), 545–563.

Ehrhardt, M. (2015). *Joint reconstruction for multi-modality imaging with common structure* (Doctoral dissertation). UCL (University College London).

Elyan, E., Jamieson, L., & Ali-Gombe, A. (2020). Deep learning for symbols detection and classification in engineering drawings. *Neural Networks*, *129*, 91–102. https://doi.org/https://doi.org/10.1016/j.neunet.2020.05.025

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, *321*, 321–331.

Goodfellow, I., Bengio, Y., & Courville, A. (2016a). *Deep learning* [http://www.deeplearningbook.org]. MIT Press.

Goodfellow, I., Bengio, Y., & Courville, A. (2016b). *Deep learning*. MIT press.

Hackeling, G. (2017). *Mastering machine learning with scikit-learn*. Packt Publishing Ltd.

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine [Insights Into the Future of Medicine: Technologies, Concepts, and Integration]. *Metabolism*, *69*, S36–S40. https://doi.org/https://doi.org/10.1016/j.metabol.2017.01.011

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., & Yakoi, P. S. (2021). Pneumonia classification using deep learning from chest x-ray images during COVID-19. *Cognitive Computation*. https://doi.org/10.1007/s12559-020-09787-5

Jeyavathana, R. B., Balasubramanian, R., & Pandian, A. A. (2016). A survey: Analysis on pre-processing and segmentation techniques for medical images. *International Journal of Research and Scientific Innovation (IJRSI)*, *3*(6), 113–120.

Kermany, D. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. https://doi.org/10.17632/RSCBJBR9SJ.2

Key, T. J., Verkasalo, P. K., & Banks, E. (2001). Epidemiology of breast cancer. *The Lancet Oncology*, *2*(3), 133–140. https://doi.org/https://doi.org/10.1016/S1470-2045(00)00254-0

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, 1097–1105.

Larobina, M., & Murino, L. (2014). Medical Image File Formats. *Journal of Digital Imaging*, *27*(2), 200–206. https://doi.org/10.1007/s10278-013-9657-9

Masoudi, S., Harmon, S. A., Mehralivand, S., Walker, S. M., Raviprakash, H., Bagci, U., Choyke, P. L., & Turkbey, B. (2021). Quick guide on radiology image preprocessing for deep learning applications in prostate cancer research. *Journal of Medical Imaging*, *8*(1), 010901–010901.

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Jamalipour Soufi, G. (2020). Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *Medical Image Analysis*.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819346

Niu, S., Liu, Y., Wang, J., & Song, H. (2020). A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, *1*(2), 151–166.

Oakden-Rayner, L. (2020). Exploring large-scale public medical image datasets. *Academic radiology*, *27*(1), 106–112.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345–1359.

PHIC. (2020). Health annual report palestine 2020. *MOH.* www.moh.ps

Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., Qi, S., Kong, F., Zhu, X., & Zhao, X. (2020). Identification of COVID-19 samples from chest x-ray images using deep learning: A comparison of transfer learning approaches. *Journal of X-Ray Science and Technology*, *28*(5), 821–839. https://doi.org/10.3233/xst-200715

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.

Razzak, M., Naz, S., & Zaib, A. (2017). Deep learning for medical image processing: Overview, challenges and future.

Ribeiro, E., Uhl, A., Wimmer, G., & Häfner, M. (2016). Exploring deep learning and transfer learning for colonic polyp classification. *Computational and mathematical methods in medicine*, *2016*.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, *63*(7), 1455–1462.

Stone, P., Brooks, R., Brynjolfsson, E., & al. et, e. (2016). Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence. https://apo.org.au/node/210721

Suckling, Parker, Dance, Astley, Hutt, Boggis, & Ricketts. (2015). Mammographic image analysis society (mias) database v1.21 [dataset]. https://www.repository.cam.ac.uk/handle/1810/250394

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wang, F., Eljarrat, A., Müller, J., Henninen, T. R., Erni, R., & Koch, C. T. (2020). Multi-resolution convolutional neural networks for inverse problems. *Scientific Reports*, *10*(1), 5730. https://doi.org/10.1038/s41598-020-62484-z

Zhu, Y., Chen, Y., Lu, Z., Pan, S., Xue, G.-R., Yu, Y., & Yang, Q. (2011). Heterogeneous transfer learning for image classification. *Proceedings of the AAAI conference on artificial intelligence*, *25*(1), 1304–1309.

Appendix A

BCD Functional & Non-Functional Requirements

A.1 Functional Requirements

This specification was prepared to cover the various services BCD required based on the functions' modality. The following table summarizes and describes the features of BCD:

	Functional Requirement	Desc.
1	BCD Public Portal	Design and implement a launching page from within the MOH public portal that allows citizens/organizations to register as users. Create new cases from the internet or follow the active cases.

2	Citizen Registration Process	When accessing the public portal, the citizen will be presented with three methods of logging into the system: Log-In Screen: where a previous user can use his user name and password to login into the system. New User Screen: this screen will allow a new user to enter the required information to login into the system. Anonymous user where no user-specific information will be recorded. The Anonymous user can upload and analyze the mammogram images.
3	New Schedule request Process	For each type of request, the system will display a checklist of mandatory documents that need to be attached to the request and will allow the user to attach Mammogram images and an electronic copy of the documents.
4	In-Progress Case Process	Once the Staff member has marked a request as complete and ready for processing and based on the preconfigured options, the request will be assigned to the staff member responsible for further processing. This assignment could be manual, or it could be done via configured rule.
5	Search	The solution must have the capability to allow MOH staff to search on keywords such as Citizen ID, Request Number, date/time, case request status, request type, telephone number, etc., and retrieve the request records.

6	Reporting	The solution should provide a dashboard (Graphical and textual) that contains defined Key Performance Indicators that would appear on the home pages of the supervisors and managers in the MOH. The suggested KPIs are as follows: Number of Received Request during a period (default to monthly) Percentage of closed Request by status. Distribution of requests by type
7	Closed Request	The system shall not allow any modifications (to the reports, responses, comments)
8	Form Builder	The system should have a form builder to allow the building of forms in a visual style. The fields in the form designer should include text boxes (single and multi-line), checkboxes, drop downs, date pickers, and file upload fields.
9	User Management	The solution shall allow the administrator to define new users. At a minimum, the definition of new users will have the following fields: User name First and Last Name E-Mail Password field Security Challenge “hint” question and answer

10	Security control	The solution shall allow the administrator to define roles with specific privileges and permissions. In addition, to allowing the administrator to create groups of users and attach a group to a role..
11	Storage Control	The solution shall have a screen that allows the administrator to allow attachments, their storage location, and the maximum size per attachment.
12	Case Management	The solution shall be able to split a request into sub-requests, assign them to different departments, and monitor their progress as separate or related requests.
13	Image classification	The solution shall have the ability to classify the image as normal or abnormal
14	Image segmentation	The solution shall be able to classify the image as benign or malignant and localize the tumor with a highlight color in the image.
15	Workflow Builder	The solution shall have a business process designer that allows business analysts and SMEs to create workflow maps using a web-based process designer that will map the flow of a request based on the nature of the request, authority matrix, and organization structure. The workflow maps created will allow for pre and post-actions.
16	Audit Trail	The solution should have audit trail logs that contain the user's login/logout date/time by user ID, terminal (or IP address),

Table A.1: BCD Functional Requirements

A.2 Non-Functional Requirements

Category	Feature
Ease of Use	
	The BCD will support multiple languages for both user interface and content. The BCD will support Arabic data entry and the Arabic user interface.
	Messages and notifications will be provided in various languages (Arabic, English)
	The BCD will support English data entry and an English user interface.
	The BCD will support the ability to determine the language of the user interface for each user.
	The BCD will support entering lookup descriptions in Arabic and English.
	The BCD will present various graphical user interface components, including text items, drop-down lists, hierarchical trees, and buttons.
	The BCD Will include help with the classification scheme, including, at a minimum, easy access to the descriptive metadata for classes, files, sub-files, and volumes.
	The BCD Will include a thesaurus to assist users in selecting terms for keywords, descriptions, etc.
	All error messages produced by the BCD will be meaningful so that users can decide how to correct the error or cancel the process. Ideally, each error message will be accompanied by an explanatory text and an indication of the action(s) that the user can take in response to the error

The BCD user interface will be suitable for users with the widest range of needs and abilities; that is, designed according to suitable accessibility standards and guidelines and compatible with common specialized accessibility software

The BCD will be easy to use and intuitive throughout.

The BCD will be accessible through a Web Based graphical user interface.

The BCD Will not require installing any software or configuration on the client side.

The BCD will be able to display simultaneously multiple records and aggregations

The BCD will allow users to customize aspects of the graphical user interface. Customization will include but need not be limited to, the following changes:

menu and toolbar contents;

screen layout;

use of function keys;

on-screen colors, fonts, and font sizes;

Audible alerts.

The BCD will allow persistent defaults for data entry where desirable. These defaults should include:

user-definable values;

a fixed default value;

values same as the previous item;

values derived from context, e.g., today's date, file reference, user identifier, as appropriate.

The BCD Will be tightly integrated with the organization's e-mail system to allow users to send records and aggregations electronically without leaving the BCD.

The BCD Will indicate whether an e-mail message has an attachment.

The BCD Will allow a user with access to a request/application to route it to another specified user with the required roles and privileges.

Performance and Scalability

The BCD will be able to return the results of a simple search (the hit list) within <3 seconds>and of a complex search (combining four terms) within <10 seconds>regardless of the storage capacity or a number of files and records on the system.

The BCD will be able to retrieve and display within <5 seconds>the first page of cases which has been accessed within the previous <2>months, regardless of storage capacity or a number of files/records on the system.

The BCD will be able to retrieve and display within <20 seconds>the malignant cases which have not been updated within the previous <12>months, regardless of storage capacity or a number of records on the system.

It will be possible to expand the BCD, in a controlled manner, to meet MOH growth up to at least <one hundred>users while providing continuity of service.

	The BCD will be scalable and must be able to be used in small or large sizes, with varying numbers of differently-sized units and across different geographical locations
System Availability	
	The BCD can be configured to be available to users: from 24 by 7
	Planned downtime for the BCD will not exceed <6 hours>per <rolling 0ne year period>.
	Unplanned downtime for the BCD will not exceed <6 hours>per <rolling one year period>.
System Flexibility	
	The BCD will be highly customized and should be easily configurable
	The BCD will include some ‘Report Generator’ and Query Builder so that trained users or system administrators can generate ad-hoc reports and a standard list of reports generated using a standard report writer such as Crystal Reports.
	The BCD will include a facility for system administrators to design new forms or templates.
	The BCD will include a facility to export reports into Microsoft Office (Word, Excel, etc.), HTML, XML, and PDF.
	The BCD provides a facility for Arabic/English reporting capabilities

Table A.2: BCD Non-Functional Requirements

Appendix B

Project Management

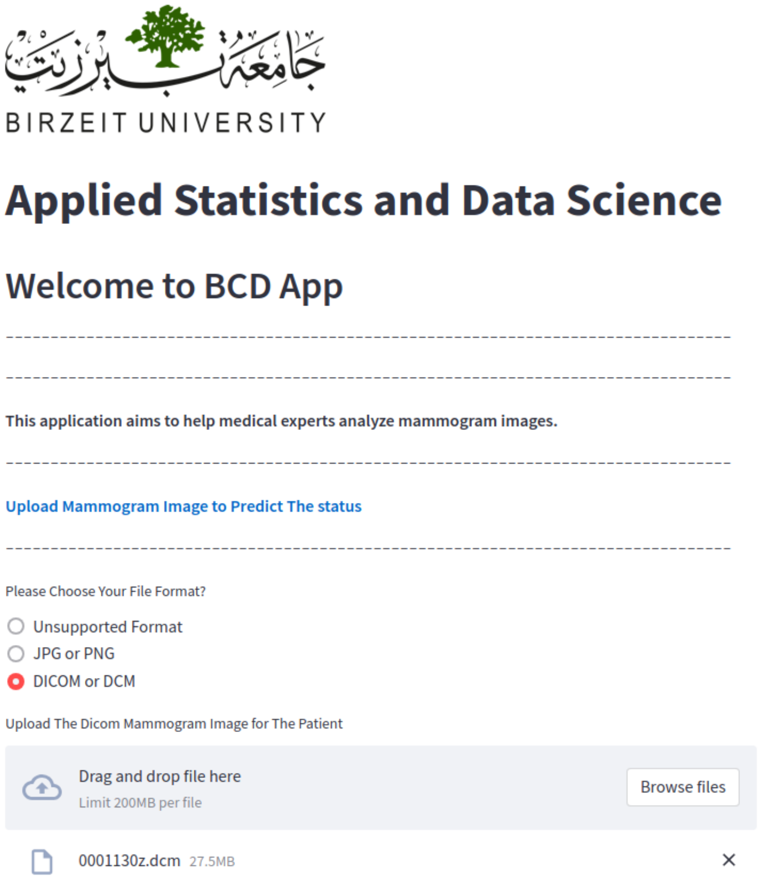
Task Name	Duration	Start	Finish
Summary #1	34 days	Thu 9/8/22	Tue 10/25/22
Meeting with Dr. Hassan to confirm the thesis topics and approve the outline to start the project.	2 days	Thu 9/8/22	Fri 9/9/22
Literature review.	30 days	Mon 9/12/22	Fri 10/21/22
Summary #1 Complete the CH 1 and CH 2	2 days	Mon 10/24/22	Tue 10/25/22
Summary #2	38 days	Wed 10/26/22	Fri 12/16/22
Meet with the Director General of Information Technology at the Ministry of Health to challenge the requirements needed to provide patient data.	3 days	Wed 10/26/22	Fri 10/28/22
Data collection from the Ministry of Health, Bethlehem branch.	14 days	Mon 10/31/22	Thu 11/17/22
Data analysis.	14 days	Fri 11/18/22	Wed 12/7/22
Summary #2 Complete CH 3 Methods presented, justified,	7 days	Thu 12/8/22	Fri 12/16/22
Summary #3	147 days	Mon 12/12/22	Tue 7/4/23
Mammogram images preprocessing	30 days	Tue 12/13/22	Mon 1/23/23

Technical implementation of the experimental framework for BC detection.	30 days	Tue 1/24/23	Mon 3/6/23
Experiments and results will be presented, discussed, and interpreted.	20 days	Thu 3/9/23	Wed 4/5/23
Final findings, conclusions, and future direction	14 days	Thu 4/6/23	Tue 4/25/23
Tuning, enhancement, and finalizing	50 days	Wed 4/26/23	Tue 7/4/23
Summary #3 Thesis Complete	1 day	Tue 7/4/23	Tue 7/4/23

Table B.1: Project Tasks

Appendix C

Breast Cancer Detection App



BIRZEIT UNIVERSITY

Applied Statistics and Data Science

Welcome to BCD App

This application aims to help medical experts analyze mammogram images.

Upload Mammogram Image to Predict The status

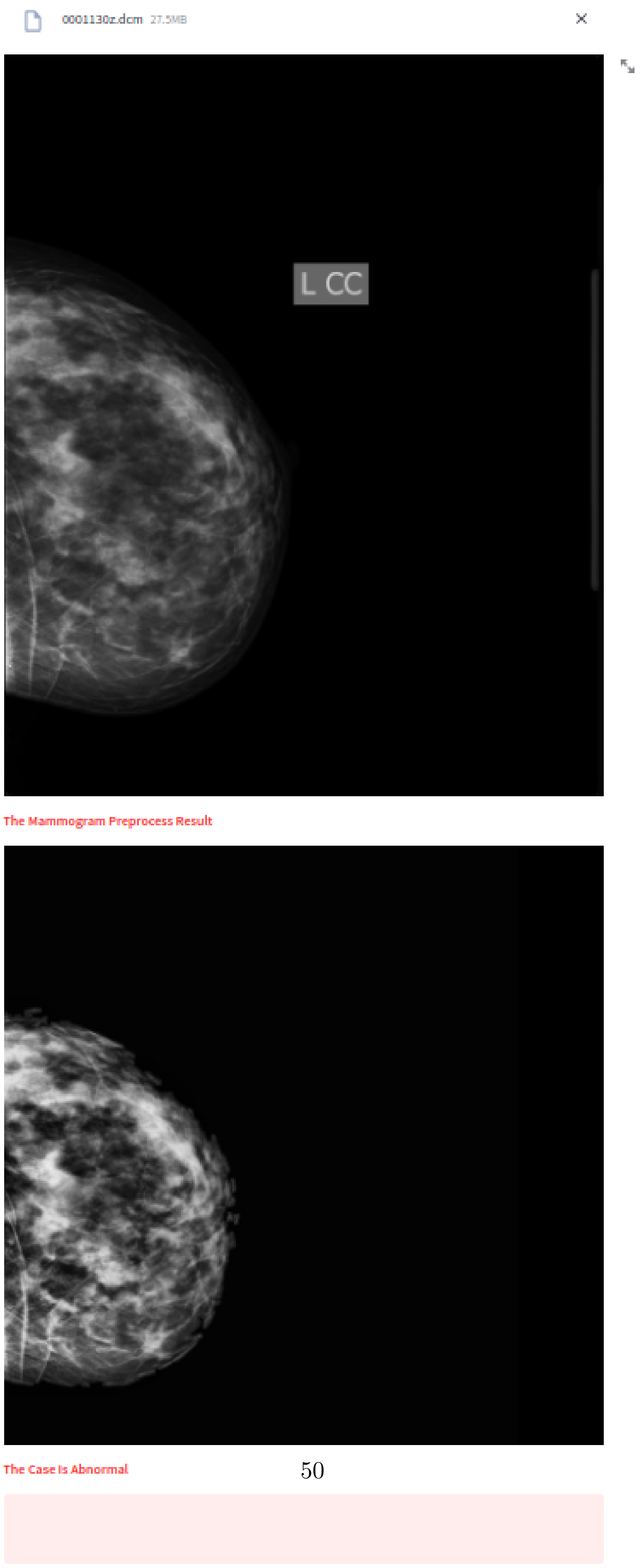
Please Choose Your File Format?

○ Unsupported Format
○ JPG or PNG
● DICOM or DCM

Upload The Dicom Mammogram Image for The Patient

0001130z.dcm 27.5MB

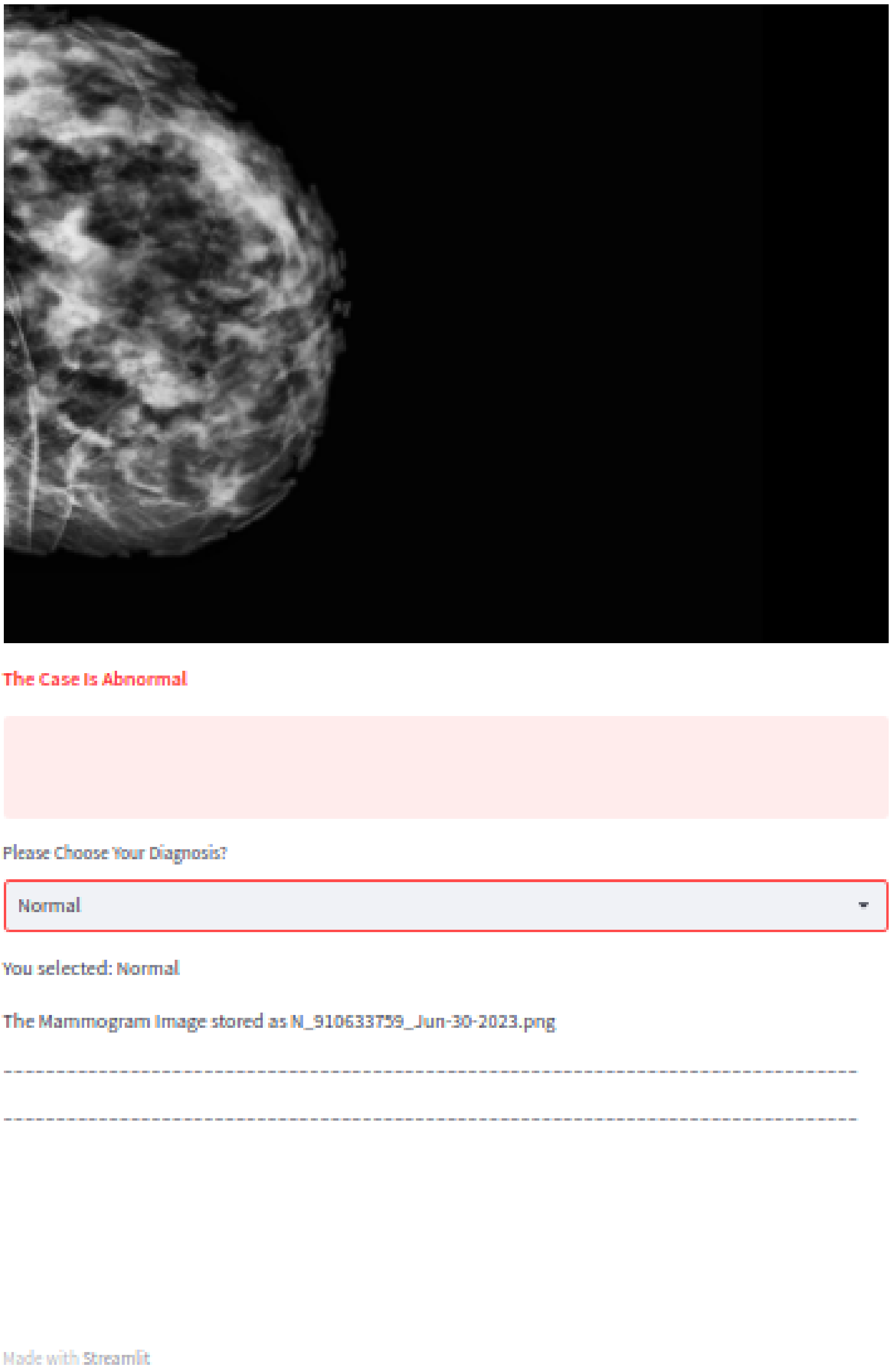
Figure C.1: BCD APP 1



The Mammogram Preprocess Result

The Case Is Abnormal

Figure C.2: BCD APP 2



The Case Is Abnormal

Please Choose Your Diagnosis?

Normal

You selected: Normal

The Mammogram Image stored as N_910633759_Jun-30-2023.png

Figure C.3: BCD APP 3

Appendix D

Presentation Slides

Breast Cancer Detection Using ML

Name Ziyad Eleyan
Regd No: 1203854

Supervisor
Dr Hassan Hassan

Master's Program in Applied Statistics and Data Science
Birzeit University

June 23, 2023

Contents

1. Introduction
2. Background and Literature Survey
3. Methodology
4. Implementation
5. Experiments & Results
6. Conclusion and future work
7. References
8. Acknowledgement

Appendix E

Outputs of Experiment One (Model One)

```
Epoch 1 of 10

Training

20it [12:12, 36.61s/it]

Validating

10it [01:31, 9.16s/it]

Train Loss: 1.3840, Train Acc: 28.47

Val Loss: 1.3818, Val Acc: 29.22

Epoch 2 of 10

Training

20it [12:15, 36.78s/it]

Validating

10it [01:31, 9.18s/it]

Train Loss: 1.3762, Train Acc: 30.56

Val Loss: 1.3593, Val Acc: 29.87

Epoch 3 of 10

Training
```

```
20it [12:15, 36.76s/it]
Validating
10it [01:31, 9.15s/it]
Train Loss: 1.3698, Train Acc: 32.26
Val Loss: 1.3720, Val Acc: 29.22
Epoch 4 of 10
Training
20it [12:14, 36.70s/it]
Validating
10it [01:31, 9.16s/it]
Train Loss: 1.3676, Train Acc: 31.85
Val Loss: 1.3518, Val Acc: 38.96
Epoch 5 of 10
Training
20it [12:13, 36.68s/it]
Validating
10it [01:31, 9.18s/it]
Train Loss: 1.3709, Train Acc: 29.11
Val Loss: 1.3582, Val Acc: 38.96
Epoch 6 of 10
Training
20it [12:17, 36.88s/it]
Validating
10it [01:33, 9.31s/it]
Train Loss: 1.3715, Train Acc: 29.27
Val Loss: 1.3527, Val Acc: 29.22
```

```
Epoch 7 of 10
Training
20it [12:15, 36.79s/it]
Validating
10it [01:32, 9.24s/it]
Train Loss: 1.3639, Train Acc: 32.66
Val Loss: 1.3443, Val Acc: 40.91
Epoch 8 of 10
Training
20it [12:18, 36.93s/it]
Validating
10it [01:32, 9.20s/it]
Train Loss: 1.3624, Train Acc: 32.98
Val Loss: 1.3456, Val Acc: 32.47
Epoch 9 of 10
Training
20it [12:15, 36.78s/it]
Validating
10it [01:31, 9.16s/it]
Train Loss: 1.3627, Train Acc: 32.50
Val Loss: 1.3471, Val Acc: 35.71
Epoch 10 of 10
Training
20it [12:17, 36.86s/it]
Validating
10it [01:32, 9.23s/it]
```

```
Train Loss: 1.3604, Train Acc: 32.02

Val Loss: 1.3382, Val Acc: 33.77

Training time: 137.926 minutes

Saving loss and accuracy plots...
```

Appendix F

Outputs of Experiment Two (Model Two)

```
Epoch 1 of 10

Training

78it [15:05, 11.61s/it]

Validating

27

Train Loss: 0.6925, Train Acc: 53.49

Val Loss: 0.6941, Val Acc: 53.66

Epoch 2 of 10

Training

78it [14:59, 11.53s/it]

Validating

27

Train Loss: 0.6905, Train Acc: 53.97

Val Loss: 0.6957, Val Acc: 53.66

Epoch 3 of 10

Training
```

```
78it [14:39, 11.27s/it]

Validating

27

Train Loss: 0.6901, Train Acc: 53.97

Val Loss: 0.6967, Val Acc: 53.66

Epoch 4 of 10

Training

78it [14:48, 11.39s/it]

Validating

27

Train Loss: 0.6902, Train Acc: 53.97

Val Loss: 0.6958, Val Acc: 53.66

Epoch 5 of 10

Training

78it [14:55, 11.49s/it]

Validating

27

Train Loss: 0.6900, Train Acc: 53.97

Val Loss: 0.6981, Val Acc: 53.66

Epoch 6 of 10

Training

78it [15:42, 12.08s/it]

Validating

27

Train Loss: 0.6899, Train Acc: 53.97

Val Loss: 0.6963, Val Acc: 53.66
```

```
Epoch 7 of 10
Training
78it [16:00, 12.31s/it]
Validating
27
Train Loss: 0.6894, Train Acc: 53.97
Val Loss: 0.7036, Val Acc: 53.66
Epoch 8 of 10
Training
78it [16:17, 12.53s/it]
Validating
27
Train Loss: 0.6905, Train Acc: 53.97
Val Loss: 0.6967, Val Acc: 53.66
Epoch 9 of 10
Training
78it [15:52, 12.21s/it]
Validating
27
Train Loss: 0.6896, Train Acc: 53.97
Val Loss: 0.6958, Val Acc: 53.66
Epoch 10 of 10
Training
78it [15:34, 11.97s/it]
Validating
27
```

```
Train Loss: 0.6887, Train Acc: 53.97
Val Loss: 0.7023, Val Acc: 53.66
Training time: 163.130 minutes
Saving loss and accuracy plots...
```

Appendix G

Outputs of Experiment Three (Model Three)

```
Starting Epoch: 1...
Epoch: 1, Mini-Batches Completed: 10, Loss: 0.121, Test Accuracy = 61.739
Epoch: 1, Mini-Batches Completed: 20, Loss: 0.110, Test Accuracy = 72.174
Epoch: 1, Mini-Batches Completed: 30, Loss: 0.108, Test Accuracy = 71.304
Epoch: 1, Mini-Batches Completed: 40, Loss: 0.104, Test Accuracy = 71.304
Epoch: 1, Mini-Batches Completed: 50, Loss: 0.106, Test Accuracy = 68.696
Starting Epoch: 2...
Epoch: 2, Mini-Batches Completed: 10, Loss: 0.105, Test Accuracy = 69.565
Epoch: 2, Mini-Batches Completed: 20, Loss: 0.104, Test Accuracy = 69.565
Epoch: 2, Mini-Batches Completed: 30, Loss: 0.103, Test Accuracy = 69.565
Epoch: 2, Mini-Batches Completed: 40, Loss: 0.100, Test Accuracy = 64.348
Epoch: 2, Mini-Batches Completed: 50, Loss: 0.094, Test Accuracy = 72.174
Starting Epoch: 3...
Epoch: 3, Mini-Batches Completed: 10, Loss: 0.104, Test Accuracy = 73.043
Epoch: 3, Mini-Batches Completed: 20, Loss: 0.108, Test Accuracy = 62.609
Epoch: 3, Mini-Batches Completed: 30, Loss: 0.098, Test Accuracy = 73.043
```

```
Epoch: 3, Mini-Batches Completed: 40, Loss: 0.107, Test Accuracy = 61.739
Epoch: 3, Mini-Batches Completed: 50, Loss: 0.097, Test Accuracy = 73.913
                         Starting Epoch: 4...
Epoch: 4, Mini-Batches Completed: 10, Loss: 0.099, Test Accuracy = 65.217
Epoch: 4, Mini-Batches Completed: 20, Loss: 0.101, Test Accuracy = 71.304
Epoch: 4, Mini-Batches Completed: 30, Loss: 0.098, Test Accuracy = 73.043
Epoch: 4, Mini-Batches Completed: 40, Loss: 0.100, Test Accuracy = 66.087
Epoch: 4, Mini-Batches Completed: 50, Loss: 0.097, Test Accuracy = 73.043
                         Starting Epoch: 5...
Epoch: 5, Mini-Batches Completed: 10, Loss: 0.099, Test Accuracy = 69.565
Epoch: 5, Mini-Batches Completed: 20, Loss: 0.101, Test Accuracy = 74.783
Epoch: 5, Mini-Batches Completed: 30, Loss: 0.093, Test Accuracy = 71.304
Epoch: 5, Mini-Batches Completed: 40, Loss: 0.099, Test Accuracy = 73.043
Epoch: 5, Mini-Batches Completed: 50, Loss: 0.097, Test Accuracy = 74.783
                         Starting Epoch: 6...
Epoch: 6, Mini-Batches Completed: 10, Loss: 0.096, Test Accuracy = 73.043
Epoch: 6, Mini-Batches Completed: 20, Loss: 0.097, Test Accuracy = 75.652
Epoch: 6, Mini-Batches Completed: 30, Loss: 0.094, Test Accuracy = 74.783
Epoch: 6, Mini-Batches Completed: 40, Loss: 0.099, Test Accuracy = 71.304
Epoch: 6, Mini-Batches Completed: 50, Loss: 0.095, Test Accuracy = 77.391
                         Starting Epoch: 7...
Epoch: 7, Mini-Batches Completed: 10, Loss: 0.101, Test Accuracy = 73.043
Epoch: 7, Mini-Batches Completed: 20, Loss: 0.096, Test Accuracy = 67.826
Epoch: 7, Mini-Batches Completed: 30, Loss: 0.097, Test Accuracy = 74.783
Epoch: 7, Mini-Batches Completed: 40, Loss: 0.094, Test Accuracy = 77.391
Epoch: 7, Mini-Batches Completed: 50, Loss: 0.100, Test Accuracy = 74.783
```

```
Starting Epoch: 8...

Epoch: 8, Mini-Batches Completed: 10, Loss: 0.092, Test Accuracy = 73.913

Epoch: 8, Mini-Batches Completed: 20, Loss: 0.097, Test Accuracy = 72.174

Epoch: 8, Mini-Batches Completed: 30, Loss: 0.101, Test Accuracy = 73.913

Epoch: 8, Mini-Batches Completed: 40, Loss: 0.094, Test Accuracy = 73.913

Epoch: 8, Mini-Batches Completed: 50, Loss: 0.097, Test Accuracy = 75.652

Starting Epoch: 9...

Epoch: 9, Mini-Batches Completed: 10, Loss: 0.091, Test Accuracy = 72.174

Epoch: 9, Mini-Batches Completed: 20, Loss: 0.092, Test Accuracy = 74.783

Epoch: 9, Mini-Batches Completed: 30, Loss: 0.097, Test Accuracy = 75.652

Epoch: 9, Mini-Batches Completed: 40, Loss: 0.095, Test Accuracy = 72.174

Epoch: 9, Mini-Batches Completed: 50, Loss: 0.093, Test Accuracy = 73.913

Starting Epoch: 10...

Epoch: 10, Mini-Batches Completed: 10, Loss: 0.096, Test Accuracy = 73.043

Epoch: 10, Mini-Batches Completed: 20, Loss: 0.099, Test Accuracy = 73.043

Epoch: 10, Mini-Batches Completed: 30, Loss: 0.094, Test Accuracy = 77.391

Epoch: 10, Mini-Batches Completed: 40, Loss: 0.093, Test Accuracy = 76.522

Epoch: 10, Mini-Batches Completed: 50, Loss: 0.092, Test Accuracy = 77.391

Finished Training
```

Appendix H

Outputs of Experiment Four (VGG 16 Model)

```
Epoch 0/24

———-

train Loss: 0.8289 Acc: 0.6082 test Loss: 0.6848 Acc: 0.6137

Epoch 1/24

———-

train Loss: 0.5692 Acc: 0.6427 test Loss: 0.4982 Acc: 0.6845

Epoch 2/24

———-

train Loss: 0.5366 Acc: 0.6784 test Loss: 0.5169 Acc: 0.6824

Epoch 3/24

———-

train Loss: 0.5133 Acc: 0.7066 test Loss: 0.4705 Acc: 0.7639

Epoch 4/24

———-

train Loss: 0.4775 Acc: 0.7342 test Loss: 0.5170 Acc: 0.7124

Epoch 5/24
```

```
———-

train Loss: 0.4381 Acc: 0.7606 test Loss: 0.4268 Acc: 0.7897

Epoch 6/24

———-

train Loss: 0.4167 Acc: 0.7825 test Loss: 0.4778 Acc: 0.7446

Epoch 7/24

———-

train Loss: 0.3393 Acc: 0.8262 test Loss: 0.3767 Acc: 0.8112

Epoch 8/24

———-

train Loss: 0.2985 Acc: 0.8619 test Loss: 0.3780 Acc: 0.8112

Epoch 9/24

———-

train Loss: 0.2919 Acc: 0.8682 test Loss: 0.3883 Acc: 0.8112

Epoch 10/24

———-

train Loss: 0.2540 Acc: 0.8947 test Loss: 0.3876 Acc: 0.8219

Epoch 11/24

———-

train Loss: 0.2483 Acc: 0.8976 test Loss: 0.4271 Acc: 0.8090

Epoch 12/24

———-

train Loss: 0.2160 Acc: 0.9108 test Loss: 0.4148 Acc: 0.8219

Epoch 13/24

———-

train Loss: 0.2008 Acc: 0.9171 test Loss: 0.4167 Acc: 0.8219
```

```
Epoch 14/24

———-

train Loss: 0.1724 Acc: 0.9396 test Loss: 0.4090 Acc: 0.8326
```

Appendix I

Outputs of Experiment Five (Inception V3 Model)

```
Epoch 0/24

———-

train Loss: 0.6186 Acc: 0.6594 test Loss: 0.4795 Acc: 0.7017

Epoch 1/24

———-

train Loss: 0.5566 Acc: 0.6801 test Loss: 0.5013 Acc: 0.7232

Epoch 2/24

———-

train Loss: 0.4977 Acc: 0.7238 test Loss: 0.4484 Acc: 0.7468

Epoch 3/24

———-

train Loss: 0.5023 Acc: 0.7330 test Loss: 0.4891 Acc: 0.7189

Epoch 4/24

———-

train Loss: 0.4646 Acc: 0.7491 test Loss: 0.4243 Acc: 0.7532

Epoch 5/24
```

```
———-
train Loss: 0.4440 Acc: 0.7629 test Loss: 0.4364 Acc: 0.7790
Epoch 6/24
———-
train Loss: 0.4348 Acc: 0.7664 test Loss: 0.4527 Acc: 0.7554
Epoch 7/24
———-
train Loss: 0.3818 Acc: 0.8044 test Loss: 0.3821 Acc: 0.8090
Epoch 8/24
———-
train Loss: 0.3553 Acc: 0.8395 test Loss: 0.4075 Acc: 0.7961
Epoch 9/24
———-
train Loss: 0.3452 Acc: 0.8441 test Loss: 0.4114 Acc: 0.7918
Epoch 10/24
———-
train Loss: 0.3186 Acc: 0.8475 test Loss: 0.4087 Acc: 0.8047
Epoch 11/24
———-
train Loss: 0.3091 Acc: 0.8608 test Loss: 0.4007 Acc: 0.8026
Epoch 12/24
———-
train Loss: 0.3096 Acc: 0.8585 test Loss: 0.3987 Acc: 0.7961
Epoch 13/24
———-
train Loss: 0.2789 Acc: 0.8867 test Loss: 0.4405 Acc: 0.7961
```

```
Epoch 14/24

———-

train Loss: 0.2638 Acc: 0.8930 test Loss: 0.4081 Acc: 0.7983
```

Appendix J

Outputs of Experiment Six (Resnet Model)

```
Epoch 0/24

———-

train Loss: 0.6799 Acc: 0.6410 test Loss: 0.6229 Acc: 0.6760

Epoch 1/24

———-

train Loss: 0.6155 Acc: 0.6577 test Loss: 0.6541 Acc: 0.6609

Epoch 2/24

———-

train Loss: 0.6159 Acc: 0.6933 test Loss: 0.4303 Acc: 0.7511

Epoch 3/24

———-

train Loss: 0.5249 Acc: 0.7181 test Loss: 0.5273 Acc: 0.7167

Epoch 4/24

———-

train Loss: 0.5304 Acc: 0.7284 test Loss: 0.4956 Acc: 0.7768

Epoch 5/24
```

```
———-

train Loss: 0.5193 Acc: 0.7399 test Loss: 0.4203 Acc: 0.7747

Epoch 6/24

———-

train Loss: 0.4916 Acc: 0.7543 test Loss: 0.5445 Acc: 0.6888

Epoch 7/24

———-

train Loss: 0.3754 Acc: 0.8205 test Loss: 0.3949 Acc: 0.8026

Epoch 8/24

———-

train Loss: 0.3646 Acc: 0.8280 test Loss: 0.3870 Acc: 0.8069

Epoch 9/24

———-

train Loss: 0.3372 Acc: 0.8389 test Loss: 0.3833 Acc: 0.7918

Epoch 10/24

———-

train Loss: 0.3220 Acc: 0.8470 test Loss: 0.3980 Acc: 0.8069

Epoch 11/24

———-

train Loss: 0.3281 Acc: 0.8441 test Loss: 0.4079 Acc: 0.7940

Epoch 12/24

———-

train Loss: 0.2986 Acc: 0.8665 test Loss: 0.3938 Acc: 0.7876

Epoch 13/24

———-

train Loss: 0.2887 Acc: 0.8734 test Loss: 0.3912 Acc: 0.7876
```

```
Epoch 14/24

———-

train Loss: 0.2605 Acc: 0.8941 test Loss: 0.3937 Acc: 0.8069
```