Breast Cancer Classification Using Deep Learning

By

Girma Abebe

A thesis Submitted to the Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in

Computer Science and Engineering

Office of Graduate Studies

Adama Science and Technology University

February, 2024

Adama, Ethiopia

Breast Cancer Classification Using Deep Learning

By

Girma Abebe

Advisor: Tilahun Melak (PHD)

A Thesis Submitted to the Department of Computer Science and Engineering,
School of Electrical and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in

Computer Science and Engineering

Office of Graduate Studies
Adama Science and Technology University

February, 2024

Adama, Ethiopia

DECLARATION

I hereby declare that this Master Thesis entitled "**Breast Cancer Classification Using Deep Learning**" is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation

_____	_____	_____
Name of the student	Signature	Date

RECOMMENDATION

I/we, the advisor(s) of this thesis, hereby certify that I/we have read the revised version of the thesis entitled “Breast Cancer classification using Deep Learning “prepared under my/our guidance by Girma Abebe submitted in partial fulfillment of the requirements for the degree of Mater’s of Science in Breast Cancer classification using Deep Learning .Therefore, I/we recommend the submission of revised version of the thesis to the department following the applicable procedures.

Tilahun Melak (PhD)	_____	_____
Major Advisor	Signature	Date
_____	_____	_____
Co-advisor	Signature	Date

APPROVAL

I/we, the advisors of the thesis entitled "**Breast Cancer classification using Deep Learning**" and developed by Girma Abebe, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

Major Advisor/Supervisor	Signature	Date
Co-advisor/ Co-Supervisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by **Girma Abebe** have read and evaluated the thesis entitled "**Breast Cancer classification using Deep Learning**" and examined the candidate during open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in **Computer Science and Engineering**.

Chairperson	Signature	Date
Internal Examiner	Signature	Date
External Examiner	Signature	Date

Final approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

Department Head	Signature	Date
School Dean	Signature	Date
Office of Postgraduate Studies, Dean	Signature	Date

ACKNOWLEDGEMENT

I am deeply grateful to the Almighty God, whose unwavering support and guidance have empowered me to navigate the challenges and accomplishments of completing this thesis.

I extend my sincerest appreciation to my advisor, Dr. Tilahun Melak, whose unwavering encouragement, comprehensive advice, and profound knowledge have been invaluable throughout my M.Sc. thesis journey. Your patience, motivation, and enthusiasm have been a constant source of inspiration, and I am truly grateful for your guidance.

I would also like to express my heartfelt gratitude to my external examiner Dr. Beakal Gizachew and my internal examiners, Dr. Mesfin Abebe, Dr. Bahiru Sheferaw, Dr. Worku Jifara, Dr. Sintayehu Hirpassa, and Dr. Ketema Adare, for their insightful feedback and valuable contributions to the refinement of my thesis.

Furthermore, I extend my thanks to every individual who has played a part in supporting me during my thesis work. Special mention goes to my friend, Ato Briuk, for his assistance with my thesis proposal, and to Ato Ashenafi and Ato Samson from the Ethiopian AI Institute, as well as the staff of the Korean Hospital, particularly Dr. Solomon CEO and Ms. Rebeca, for their support.

I am also deeply grateful to all my teachers at the Department of Computer Science for their dedication and contributions to the success of the graduate program.

To my beloved wife, Selamawit Endalu, and our precious children, Elnathan, Abigail, and Eva, your unwavering love and support have been my rock throughout this journey. I am eternally grateful for the positivity and inspiration you bring to my life every day. A special mention goes to my brilliant mother, Ethalemahu Melese, and all the family members, whose love and encouragement have been a constant source of strength. And my big appreciation goes to my brother Samson Tamirat too for his special encouragement.

In conclusion, I am profoundly thankful to everyone who has been a part of my journey, and I am filled with gratitude for the love, support, and inspiration that have surrounded me throughout this endeavor.

TABLE OF CONTENTS

CHAPTER **PAGE**

LIST OF TABLES xi

LIST OF FIGURES AND ILLUSTRATIONS xii

LIST OF EQUATIONS xiv

LIST OF ACRONYMS AND ABBREVIATIONS xv

ABSTRACT xvi

CHAPTER ONE 17

1. INTRODUCTION 17

1.1. Background of the study 17

1.2. Motivation of the Study 20

1.3. Statement of the problem 20

1.4. Research questions 21

1.5. Objectives 21

1.5.1. General Objective 21

1.5.2. Specific Objectives 21

1.6. Scope and Limitation of the Study 21

1.7. Significance of the study 22

1.8. Organization of the Thesis 22

CHAPTER TWO 23

2. LITERATURE REVIEW AND RELATED WORK 23

2.1. Literature Review 23

2.1.1. Breast anatomy 23

2.1.2. Development of the Breast 25

2.1.3. Types of breast cancer 26

2.1.4. Breast cancer category 27
2.1.5. Stages of Breast cancer 28
2.1.6. Breast Cancer Diagnosis equipment's 29
2.1.7. Mammogram 31
2.1.8. Types of Breast Cancer Detection Technique 33
2.1.8.1. Manual Breast Cancer Detection Technique 33
2.1.8.2. Digital Breast Cancer Detection Technique 34
2.2. Machine learning 34
2.3. Deep learning 35
2.3.1. EfficientNet 36
2.3.1.1. Single dimension Scaling 37
2.3.1.2. Compound scaling 37
2.3.2. EfficentNetB0 architecture 38
2.3.3. EfficientNet Performance 40
2.3.4. Residual Network, is a convolutional neural network (CNN) 41
2.3.5. Image classification with EfficientNet 42
2.4. Related Works 43
CHAPTER THREE 48
3. MATERIALS AND METHODS 48
3.1. Methodology overview 48
3.2. Research Flow 48
3.3. Preprocessing and Dataset Preparation 49
3.4. Augmentation 50
3.5. The proposed architecture 52
3.6. Working Environment 53

3.6.1. Hardware Utilities 53
3.6.2. Software Utilities 54
3.7. Evaluation methods 54
3.7.1. Confusion matrix 54
3.7.2. Receiver Operating Characteristic (ROC) Curve: 57
CHAPTER FOUR 59
4. PROPOSED SYSTEM 59
4.1. Experimental Setup 59
4.2. Dataset organization 59
4.3. Library Used 62
4.4. Importing dataset 62
4.5. Preprocessing 63
4.6. Modeling 64
4.7. Compiling the model 67
4.8. Training the model 68
4.9. Model Evaluation 69
4.9.1. Generate Confusion Matrix 71
4.9.2. Predicting the results 73
4.9.3. Receiver Operating Characteristic (ROC) 75
4.9.4. The loss and accuracy curves 76
4.9.5. Classification report 78
5. RESULTS AND DISCUSSIONS 80
5.1. RESULTS 80
5.2. Hyper parameters 80
5.1. Discussion 81

6. CONCLUSIONS AND RECOMMENDATIONS 83

6.1. Conclusions 83

6.2. Recommendations 83

6.3. Future Work 84

REFERENCE 85

LIST OF TABLES

TABLE PAGE

Table 1: Various Medical Imaging Modalities 30
Table 2:Features of Benign vs Malignant Tumors 36
Table 3: Summary of related works 46
Table 4: The confusion matrix 55
Table 5: Local Dataset organization 60
Table 6: Public Dataset organization 60
Table 6 Parameters used in the study 80

LIST OF FIGURES AND ILLUSTRATIONS

Figure 1: GLOBOCAN 2020: New Global Cancer Data 18
Figure 2: Anatomy of the Breast 25
Figure 3: Benign vs Malignant Tumors 27
Figure 4: The Stages of Breast Cancer 29
Figure 5: An examination a deep learning in breast cancer with its significant impact. 31
Figure 6: Mammography image Scanner 32
Figure 7: Manual Breast Cancer Detection Technique 33
Figure 8: Automatic Breast Cancer Detection Technique 34
Figure 9: EfficentNetB0 Single Dimension Scaling 37
Figure 10: Scaling Network Width for Different Baseline Networks. 38
Figure 11: EfficientNet-B0 baseline network 39
Figure 12: EfficientNet-B0 architecture 39
Figure 13: Scaling up EfficientNet-B0 with different methods. 40
Figure 14: Comparison of EfficientNet family state of the art models 41
Figure 15: Research Flow 49
Figure 16: Original and Augmented images 52
Figure 17: The proposed Architecture for Breast Cancer Classification 53
Figure 18: Train, test and validate split ratio on the dataset 59
Figure 19: Loading and displaying Sample Data 61
Figure 20: Sample Breast Images from Myung Sung Comprehensive Specialized Hospital .. 61
Figure 21: Libraries Used 62
Figure 22: Importing Dataset 63
Figure 23: Data preprocessing and augmentation 64
Figure 24: Image classification model using EfficientNetB0 65
Figure 25: Summary of EfficientNetB0 model 66
Figure 26: Code for compile 67
Figure 27: Train the Model 68
Figure 28: Train the model (train 70%, test 15%, and validate 15%) with EfficentNetB0 69
Figure 29: Evaluating the performance of a model 70
Figure 30: Confusion matrices for the experimentation of the EfficentNetB0 architecture 72

Figure 31: Predicting results as Benign or Malignant 73
Figure 32: Results with prediction 74
Figure 33: Receiver Operating Characteristic (ROC) 75
Figure 34: Receiver Operating Characteristic (ROC) Curve 76
Figure 35: The loss and accuracy curves 77
Figure 36: The loss and accuracy curves 78
Figure 37: Generate evaluation metrics 78
Figure 38: Classification report 79

LIST OF EQUATIONS

Equation 1: Compound scaling 38
Equation 2: Defining the dimensions of the network 38
Equation 3: Accuracy Formula 56
Equation 4: Precision Formula 56
Equation 5: Recall Formula 56
Equation 6: F1-Score Formula 57

LIST OF ACRONYMS AND ABBREVIATIONS

BC	Breast cancer
CAD	Computer-Aided Diagnosis
CEDM	Contrast-Enhanced Digital Mammography
CNN	Convolutional Neural Network
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma in situ
DDSM	Digital Database for Screening Mammography
DL	Deep Learning
DNN	Deep Neural Network
EIT	Electrical impedance imaging
FN	False Negative
FP	False Positive
FPR	False Positive Rate
HP	Histopathology Images
Mg	Mammogram
MRI	Magnetic Resonance Imaging
PEM	Positron emission mammography
PET	Positron emission tomography
PPV	Percentage of people with a positive test
RNN	Recurrent Neural Network
TN	True Negative
TP	True Positive
US	ULTRASOUND
WHO	World Health Organization

ABSTRACT

Breast cancer continues to pose a significant global health challenge, evident from the staggering statistics of 10 million deaths and 19.3 million new cases worldwide in 2018. Conventional testing methods are costly, prone to human error, and often yield inaccurate results. The classification of breast cancer based on mammography images presents its own set of challenges, exacerbated by shortages of trained radiologists and screening equipment. Ethiopia is particularly alarming due to its highest age-standardized death rate from breast cancer. With a projected 50% increase in cancer-related mortality by 2030, the importance of early detection and precise classification cannot be overstated. In addressing breast cancer detection in Ethiopia, the proposal suggests the implementation of a Convolutional Neural Networks (CNN) model, necessitating collaboration among academics, policymakers, healthcare practitioners, and IT experts. Previous research conducted in Ethiopia has demonstrated CNNs achieving accuracy rates of up to 88% using solely local datasets.

In this study, the accuracy of breast cancer classification was assessed using public, local, and combined mammography image datasets, employing two transfer learning architectures: ResNet50 and EfficientNetB0. The datasets were partitioned into three subsets: 70% for training, 15% for validation, and 15% for testing, totaling 2376 mammography images classified as benign or malignant. Applying EfficientNetB0 to the local, public, and merged image datasets yielded an accuracy of 78.43% with ROC curves of 0.81, 0.89, and 0.46 for 40 epochs with a learning rate of 0.0001respectively. Similarly, utilizing ResNet50 on the same datasets resulted in an accuracy of 78.43% with ROC curves of 0.81, 0.69, and 0.81 for 40 epochs with a learning rate of 0.0001respectively.

The study conducted a comparison between the transfer learning architectures of ResNet50 and EfficientNetB0 using mammography images obtained from local, merged, and publicly available datasets. Results from the analysis of the local dataset revealed an accuracy of 78.43% with ROC curves of 0.81, indicating that EfficientNetB0 outperformed ResNet50 in accurately classifying breast cancer as either benign or malignant.

Keywords: *Benign, Breast Cancer, Dataset, EfficentNetB0, Malignant, Mammogram and ResNet50*

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Breast cancer is a major cause of death among women globally, second only to lung cancer (Gezahegn, 2020) and (Debelee et al., 2020). According to the World Health Organization (WHO), 2.3 million women were diagnosed with breast cancer and 685,000 women died from the disease in 2020. By the end of 2020, there were 7.8 million women alive who had been diagnosed with breast cancer in the past 5 years. A study by 2030 (Nazir et al., 2022), found that one in every 19 women is at risk of developing breast cancer by 2030. Early detection of breast cancer has significantly improved survival rates (Shallu & Mehra, 2018). Breast cancer occurs due to abnormal cell division, forming tumors that can be malignant or benign. A biopsy is performed to differentiate between benign and malignant masses (Al Noumah et al., 2022). In Ethiopia, more than 80% of patients receive an advanced stage diagnosis, and there is no national program for BC screening (Ayele, 2021).

Breast cancer stands as a prominent global health challenge, with 19.3 million new cancer cases and 10 million cancer-related deaths reported in 2018. The prevalence of breast cancer is on the rise, particularly in underdeveloped nations where a significant portion of cases is identified at advanced stages. In developed countries like the USA, 2019 witnessed 48,100 new cases of DCIS (ductal carcinoma in situ) and 268,600 new cases of invasive breast cancer (BC), resulting in 41,760 women succumbing to the disease. While substantial strides have been made in breast cancer research and treatment over the past 45 years in high-income countries, with many women experiencing successful outcomes through procedures like lumpectomy, minimal lymph node surgery, and targeted therapy, low and middle-income countries bear a heavier burden of breast cancer mortality and encounter a diminished quality of life post-treatment. Despite advancements such as screening mammography, radiotherapy, and targeted therapies in the US since the 1970s, these breakthroughs remain largely inaccessible to a majority of women (Hoxha et al., 2022).

The incidence and mortality rates of breast cancer in East Africa are 33.0 and 17.9 per 100,000 women annually, respectively, when age-standardized. According to the World Health Organization (WHO) 2020 report, breast cancer has the highest age-standardized

mortality rate of 22.9 per 100,000 people in Ethiopia, making it the most common cancer. According to forecasts made by the WHO, the number of cancer deaths could increase by 50% to 15 million by 2030 worldwide. High-income countries diagnose over 70% of breast cancer patients in stages I and II, while only 20-50% in low- and middle-income countries receive early diagnosis, leading to limited treatment options and lower survival rates (Nazir et al., 2022). According to (Ilhan et al.,. (2021), high-risk factors for breast cancer include genetic mutations such as BRCA1 and BRCA2 genes, obesity, birth control, abnormal menstrual cycles, radiation therapy, and estrogen exposure. Modifiable factors include obesity, menopausal hormone use, and breastfeeding (Ayele et al., 2022).

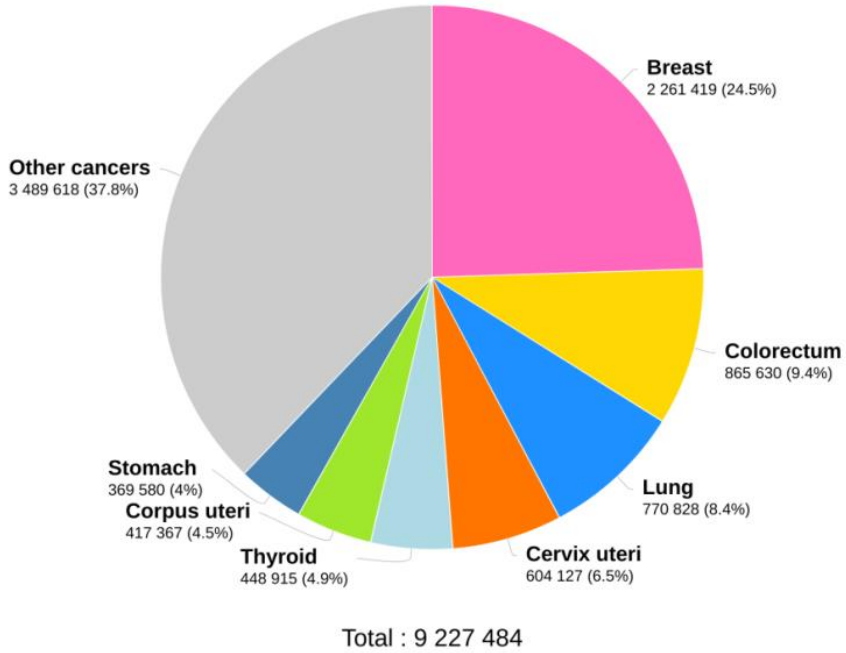


Figure 1: GLOBOCAN 2020: New Global Cancer Data

According to a study by Tesfaw et al.,. (2020), women in Ethiopia who are diagnosed with late-stage breast cancer often lack access to proper medication, painkillers, or palliative care. (Duche et al., 2021) report that the main risk factors for breast cancer in Ethiopia include lack of knowledge, cultural and religious beliefs, economic hardships, lack of transportation and healthcare, fear of surgical procedures, and lack of trust in medical care.

Among low-income nations, BC is the most common cancer to be diagnosed and the primary cause of cancer-related deaths among women (Ayele, 2021). Examining local risk factors for breast cancer prevention is crucial in Ethiopia due to the diverse female population with

varying cultural norms, economic conditions, and practices related to conception and lactation. Despite these challenges, the prevalence of the disease is higher in low-income countries like Ethiopia, where 60% of deaths from breast cancer occur. The shortage of doctors, oncologists, and pathologists is evident in Ethiopia, as reported by (Gezahegn, 2020) and (Tolessa et al., 2021), with an alarming ratio of one doctor for every 57,876 people nationwide and a substantial deficit of medical professionals, particularly in southwest and west-central Ethiopia, with estimates ranging between 200,000 and 300,000 people per doctor. Despite the extensive experience of medical professionals, cancer diagnosis remains a formidable task, marked by common mistakes and the potential for a delayed diagnosis that could significantly impact lives, as highlighted by (Rakhshan, 2023).

Radiologists employ various imaging techniques, such as digital breast tomosynthesis and screen-film mammography, to detect anomalies like masses, microcalcifications, and bilateral asymmetry indicative of breast cancer. Nevertheless, these methods have limitations, including the potential concealment of cancer by breast tissue overlaying, as pointed out by (Debelee et al., 2020). To identify and monitor breast cancer, a spectrum of techniques like mammography, MRI, ultrasound, PET, thermography, and surgical incision is utilized. The interpretation of data can pose challenges due to different cancer types presenting with similar clinical symptoms. Recognizing the growing role of deep learning in computer-assisted diagnostics, there is an increasing emphasis on automating diagnostic processes to alleviate the burden, as highlighted by (Ayalew et al., 2021).

While mammography is a widely employed breast imaging technique, it faces challenges, including intricate image interpretation and a shortage of radiologists. To address these issues, medical image analysis incorporates Deep Learning (DL) alongside the Computer-Aided Diagnosis (CAD) system. EfficentNetB0 architecture prevalent in deep learning, serve as effective models for image identification. The proposed method for breast cancer classification utilizes digital image processing techniques, automated from mammography images, enhancing the precision of patient classification at health centers and diminishing the time required for visual inspections by medical personnel (SM, 2016).

1.2.Motivation of the Study

In order to overcome the limitations of human procedures in identifying mammography images for breast cancer classification, the study uses deep learning methodology. Reducing the computational load on deep learning algorithms for managing both local and global datasets is a major goal. Furthermore, the goal is to build models for breast cancer case classification using deep learning image processing, which would enable medical practitioners to classify cases quickly and accurately. This expedited diagnosis procedure improves the efficacy and precision of classifying breast cancer cases, which is especially advantageous for medical facilities that do not have instant access to specialized medical knowledge.

1.3. Statement of the problem

Breast cancer affects women worldwide who are between the ages of 20 and 59. Traditional testing approaches are expensive, subject to human error, and have inaccurate interpretations. According to (Shiferaw et al., 2020) and (Bimrew Sendekie Belay, 2022) it is also challenging to classify breast cancer as benign or malignant based on mammography images due to a shortage of trained radiologists and lack of screening equipment's. A model based on Convolutional Neural Networks is proposed by (Khan et al., 2021) for the identification of breast cancer in Ethiopia, with the goal of mitigating the challenges associated with manual analysis. Researchers, policymakers, medical professionals, and IT developers must work together to automatically classify breast cancer in order to address the mentioned problems.

Ethiopian studies on the classification of breast cancer have demonstrated the efficacy of deep learning models. A CNN model used in an experiment (Bimrew Sendekie Belay, 2022) with a local dataset obtained from a Korean hospital achieved 88% accuracy, while a method developed by (Debelee et al., 2018) achieved 98.75% and 98.90% accuracy on public datasets MIAS and DDSM which might result in problems with artifacts, noise, and data imbalances.

The study's objective is to apply a deep learning-based model that uses datasets of mammography images from Korea Hospital and other online resources to classify breast cancer as benign or malignant.

1.4. Research questions

This study aims to investigate the following research inquiries within the proposed solution:
RQ1: Can locally organized datasets be used to distinguish between benign and malignant breast cancer?
RQ2: What is the difference in breast cancer classification accuracy between local and international image datasets?
RQ3: Which deep learning technique performs best in classifying breast cancer cases?
RQ4: What challenges come with classifying breast cancer using locally produced images?

1.5. Objectives

1.5.1. General Objective

The primary objective of this study is to develop a breast cancer classification model using deep learning approach.

1.5.2. Specific Objectives

The following specific objectives are accomplished to achieve the general objective of the study.

- To collect and prepare mammography breast images from hospitals.
- To improve performance by applying data augmentation techniques.
- To compare deep learning methods and select the best breast cancer classification architecture.
- To examine how locally gathered breast cancer images can be classified as malignant or benign using deep learning algorithms.
- To evaluate the accuracy of deep learning methods in classifying breast cancer utilizing local, global and merged image datasets.
- To investigate the constraints associated with using locally acquired images for breast cancer classification.

1.6. Scope and Limitation of the Study

Using data from female mammography taken between 2016 and 2023 from Myung Sung Comprehensive Specialized Hospital and maintained by the iView system and global datasets of breast cancer images, the research attempts to develop a deep learning model for the categorization of breast cancer.

The study has certain drawbacks. Among these limitations include the inability to account for variables like the stage of breast cancer, the classification of male breasts, and the recognition of X-ray and microscopic pictures. Moreover, the study may lack certain image datasets, cancer subtypes, machine learning strategies, or other components that could affect the effectiveness and accuracy of deep learning techniques.

1.7. Significance of the study

This study is important because it uses public and local datasets to create a system for classifying breast cancer. A precise and timely informational resource that facilitates manageable treatment options for radiologists, women, and families is the expected consequence. The research leverages both local and international image datasets with the identified deep learning-based techniques to advance deep learning methods for early diagnosis and treatment of breast cancer. It also intends to offer significant insights into the accuracy and performance of various algorithms. In addition, the study will assess the difficulties and restrictions related to classifying breast cancer using locally produced images.

1.8. Organization of the Thesis

The subsequent sections of the thesis follow this structure. The next chapter explores an extensive review of literature and related works, providing insights into breast cancer, as well as fundamental principles of machine learning and deep learning. Additionally, this chapter examines relevant studies conducted by various authors, with a specific focus on the utilization of deep learning techniques for breast cancer classification. Moving forward, the third chapter outlines the research methodology, elucidating the experimental approach in detail. Chapter four elaborates on the experiment conducted and its resulting outcomes. Subsequently, Chapter five summarizes the findings of the study, drawing conclusions. Finally, Chapter six outlines the contributions made by the research and offers recommendations for future endeavors.

CHAPTER TWO

2. LITERATURE REVIEW AND RELATED WORK

2.1.Literature Review

Within this chapter, an exploration of the literature and related works is presented. The discussion serves to elucidate the contributions of past scholars, identify existing research gaps, and provide an overview of pertinent topics. This encompasses a delineation of breast cancer, elucidation of various types and stages of the disease, an examination of digital image processing in the context of breast cancer detection, exploration of breast disease detection through Convolutional Neural Networks, and an inventory of related works within the domain. The chapter concludes with a concise summary of the literature review.

2.1.1. Breast anatomy

Recent advancements in mammography can be categorized into three types. The first is general digital mammography, utilizing electronic alternatives for x-ray film to convert x-rays into breast mammography images. The second involves computer-aided detection (CAD) systems, which analyze digitized mammograms for signs of malignancy. The third type is 3D mammography, also known as digital breast tomosynthesis (DBT), an advanced technology that stitches together multiple breast images from various perspectives to create a three-dimensional image set. It's noteworthy that ongoing research explores novel imaging methods like scintimammography (molecular breast imaging), electrical impedance imaging (EIT), positron emission mammography (PEM), and elastography (Rahel & Desta, 2020). Breast cancer, the most prevalent cancer in women, ranks fifth in mortality after lung, colon, liver, and stomach cancers (Bimrew Sendekie Belay, 2022).

The exocrine glands within the human body are responsible for producing breast milk, primarily composed of microscopic structures called lobes and fat. Each lobe consists of alveoli, surrounded by my epithelial cells linked to milk-releasing epithelial cells. Ducts, small tubes collecting milk from the lobes, are formed by my epithelial cells, creating a network of interconnected pipes that culminate in the nipple. Breast mammograms classify breast anatomy abnormalities as density, mass, or calcification, with the BI-RADS score grading mammography findings used by radiologists. Mammography, an X-ray imaging

process, is crucial for diagnosing breast cancer and is a follow-up tool when worrisome lumps are identified in clinical breast exams. It is essential to note that not all abnormalities detected are malignant (Bimrew Sendekie Belay, 2022).

Both males and females have paired breasts on the anterior thoracic wall, with females experiencing more prominent breast development after maturation. Female breasts contain mammary glands, an auxiliary gland of the female reproductive system, crucial for breastfeeding.

The female breast anatomy comprises various components, such as:

- Lobes: Each breast has 15 to 20 lobes or sections surrounding the nipple like spokes on a wheel.
- Glandular tissue (lobules): Small sections inside lobes containing bulb-like glands that produce milk.
- Milk (mammary) ducts: Small tubes carrying milk from glandular tissue (lobules) to the nipples.
- Nipples: Located in the center of the areola, each nipple has about nine milk ducts and nerves.
- Areolae: Circular dark-colored areas of skin surrounding the nipple, with Montgomery's glands secreting lubricating oil to protect against chafing during breastfeeding.
- Blood vessels: Circulate blood throughout the breasts, chest, and body.
- Lymph vessels: Part of the lymphatic system, these vessels transport lymph, aiding the body's immune system in fighting infection, connecting to lymph nodes under the armpits, in the chest, and elsewhere.
- Nerves: Nipples are highly sensitive to touch and arousal, containing hundreds of nerve endings.

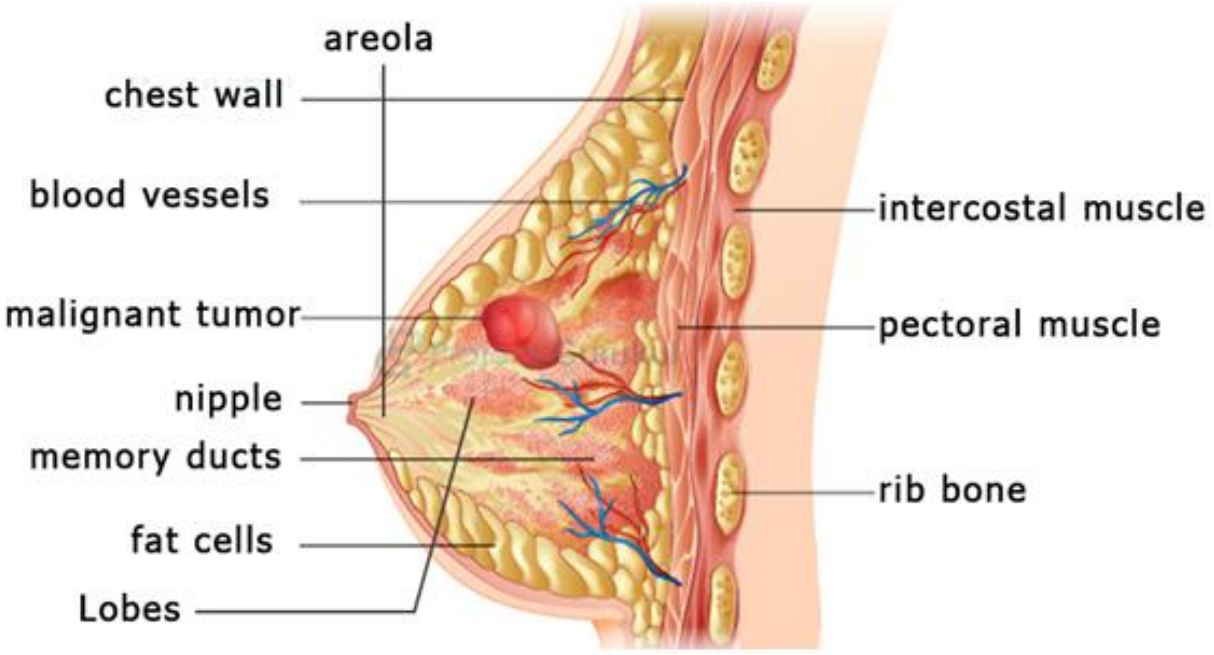


Figure 2: Anatomy of the Breast

The milk-producing secretory tissue in the breast's mammary gland consists of 12-15 lobes arranged radially, with ducts converging to open at the nipple's tip. Physiologically, breasts exhibit variations in shape and size among individuals, categorized as flat and shallow, hemispheroid, pear-shaped, and pendulous(Rahel & Desta, 2020).

2.1.2. Development of the Breast

The initiation of human breast tissue occurs around the sixth week of fetal development, originating along the lines of the armpits and extending to the groin by the ninth week of pregnancy. During this process, two breast buds form on the upper chest, only to regress back to the chest region. From each breast bud, columns of breast tissue grow inward, giving rise to individual sweat glands with ducts leading to the nipple. At birth, both male and female breasts are similar and undeveloped.

Breast development initiates during early puberty, marked by the increased prominence of the areola. In late puberty, the areola undergoes further changes, gaining fat, while glandular tissue in the breast continues to increase. The female breast comprises glandular tissue, acini, ducts, Cooper's ligament, and Montgomery's gland, while the male breast also exists. The synthesis of estrogen and progesterone hormones plays a key role in stimulating the growth of glandular breast tissue.

Several factors, including breast tissue volume, family history, age, weight fluctuations, skin thickness, hormonal influence, and pregnancy history, collectively contribute to determining a woman's breast size(Abebe, 2019).

2.1.3. Types of breast cancer

There are two types of breast cancer. The first one is Ductal Carcinoma In Situ (DCIS), which is the most common form of non-invasive breast cancer. It originates in the milk ducts of the breast and remains non-invasive as it does not spread into the surrounding breast tissue. While DCIS itself is not life-threatening, having it can elevate the risk of developing invasive breast cancer later in life. The second type comprises non-invasive breast cancers, which are confined to the milk ducts or lobules in the breast. These cancers have not extended into or invaded the normal breast tissue. Referred to as carcinoma in situ, non-invasive cancers are sometimes labeled as pre-cancers.

A tumor is an abnormal growth or mass of cells. It is considered benign when the cells are healthy but have undergone abnormal growth, forming a lump. If the cells are cancerous and have the potential for uncontrolled expansion, it is classified as a malignant tumor. A biopsy is conducted to determine the nature of the tumor, and a pathologist examines the sample under a microscope (Abebe, 2019).

There are two types of tumors. The first is a benign tumor, which is non-cancerous and cannot spread throughout the body. Benign tumors only grow locally and do not invade or metastasize. While they may cause concern if they affect nearby tissues or organs, they are generally less worrisome. Examples include lipomas and uterine fibroids, which may require surgical removal. Some benign tumors, like intestinal polyps, are considered precancerous and are removed to prevent malignancy. Once removed, benign tumors often do not return, but if they do, it is usually in the same location (Abebe, 2019).

The second type is a malignant tumor, which is cancerous and has the capacity for uncontrolled growth, metastasis, and invasion of surrounding tissues. Malignant cells can infiltrate nearby tissues, enter blood vessels, and spread to other parts of the body. Unlike benign tumors, malignant cells lack chemical adhesion molecules that keep them attached to

the original site of growth. Malignant tumors are associated with greater health risks and often require aggressive treatment (Abebe, 2019).

Breast cancer originates in the breast tissue, and without timely detection and treatment, it can progress to the lymph nodes in the armpit. Once breast cancer reaches the lymph nodes, it has the potential to metastasize to other organs, such as the liver or bones. Subsequently, tumors comprised of breast cancer cells may form in these organs. The characteristics of these cancers can be identified through a biopsy of the initial breast cancer tumor (Abebe, 2019).

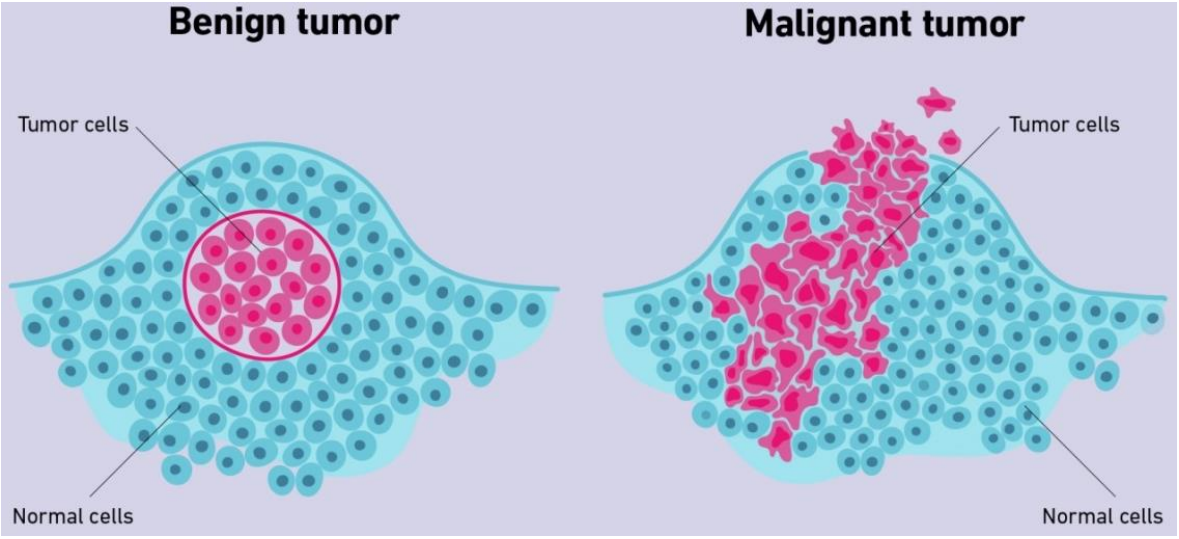


Figure 3: Benign vs Malignant Tumors

2.1.4. Breast cancer category

Doctors utilize the BI-RADS system to categorize abnormal findings, assigning scores from 0 to 6. Women aged 40 and older typically receive scores between 0 and 2, indicating normal results or benign abnormalities. Scores of 3 or above may prompt doctors and radiologists to recommend a follow-up appointment or a biopsy to determine the appropriate next steps.

- Category 0: A BI-RADS score of 0 indicates an incomplete test, where mammography images may have been challenging to interpret. Additional tests and comparisons with earlier images are needed for a definitive evaluation.
- Category 1: A score of 1 certifies that the mammography is negative, confirming the absence of cancer and equally dense breasts. Regular screenings are recommended to maintain vigilance.

- Category 2: A BI-RADS score of 2 signifies normal mammography results. While there are no signs of cancer, benign lumps or cysts may be present. Regular follow-up visits are advised, and future findings will be compared to the initial report.
- Category 3: A score of 3 suggests a 2 percent chance of cancer, despite likely normal mammography results. A follow-up visit within six months is recommended to confirm benign findings. Routine checkups help avoid unnecessary biopsies and support early cancer diagnoses.
- Category 4: Category 4 indicates an anomaly or suspicious finding with a 20 to 35 percent likelihood of cancer. Biopsy is necessary for a definitive diagnosis, and this category is further divided into three levels of suspicion: 4A (low likelihood), 4B (moderate likelihood), and 4C (high likelihood).
- Category 5: A score of 5 indicates a strong suspicion of cancer with a minimum 95% likelihood. Biopsy is strongly advised to confirm findings and determine the appropriate treatment course.
- Category 6: A BI-RADS score of 6 is assigned after a confirmed breast cancer diagnosis through biopsy. This category is used to monitor the cancer's response to interventions like chemotherapy, surgery, or radiation, using accompanying images for comparison.

2.1.5. Stages of Breast cancer

Breast cancer stages are categorized based on tumor size and the extent of its spread to lymph nodes or other body parts. The stages range from 0 to 4, with additional subcategories. The stage at the time of diagnosis significantly influences the prognosis. The following outlines each of these primary stages.

- Stage 0, also known as ductal carcinoma in situ, indicates that cancerous cells are confined to the ducts without spreading to adjacent tissues.
- Stage 1 signifies a tumor measuring up to 2 centimeters (cm) in diameter, with no lymph node involvement or small groups of cancer cells in the lymph nodes.
- At Stage 2, the tumor is either 2 cm across and has begun spreading to nearby nodes or ranges from 2 to 5 cm without lymph node involvement.
- Stage 3 involves a tumor up to 5 cm across with spread to multiple lymph nodes, or a tumor larger than 5 cm with involvement in a few lymph nodes.

- In Stage 4, cancer has metastasized to distant organs, commonly the bones, liver, brain, or lungs.

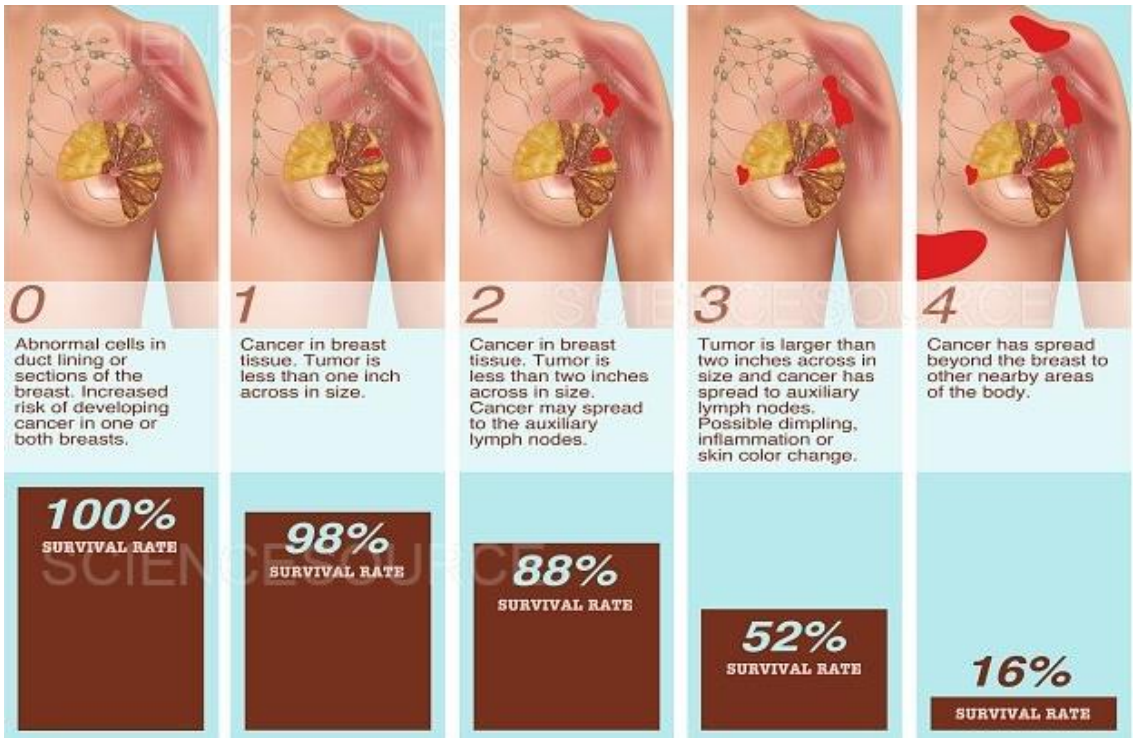


Figure 4: The Stages of Breast Cancer

2.1.6. Breast Cancer Diagnosis equipment's

Breast cancer is identified through a range of methods, with doctors frequently employing additional tests for detection and diagnosis. If necessary, women may be referred to a breast specialist or surgeon, though this referral does not necessarily indicate cancer or the immediate need for surgery. These specialists possess expertise in diagnosing breast-related issues. Refer to Table 3 (Ramadhani & First, 2020) for a summary of Breast Cancer Diagnosis equipment and various Medical Imaging Modalities.

Table 1: Various Medical Imaging Modalities

Type	Short description
Mammogram (Mg)	Mammograms are available in three formats: screen-film mammograms (SFMs), digital mammograms (DMs), and digital breast tomosynthesis (DBT). While SFMs and DMs are grayscale 2D images, DBT offers multiple 2D grayscale image frames resembling black-and-white videos.Top of Form
ULTRASOUND (US)	Ultrasound, commonly referred to as Sonograms, utilizes images in three configurations: basic 2D grayscale ultrasound images, ultrasound images supplemented with shear-wave elastography (SWE) color images, and ultrasound images combined with Nakagami color images.
Magnetic Resonance Imaging (MRI)	MRI is employed for breast cancer diagnosis using pre and post-contrast imaging, specifically Dynamic Contrast-enhanced (DCE). While post-contrast images are initially colored, they are typically converted to grayscale for input into Artificial Neural Networks (ANN).
Histopathology (HP) Images	HP Images refer to color images stained with H&E and are categorized into two groups: the whole slide image (WSI) and the image patch extracted from WSI by the pathologist.
Multimodality	Some studies employ a combination of two grayscale image modalities referred to as multimodality for breast cancer classification.
Full-field digital mammography (FFDM)	This imaging modality is considered clinically acceptable for population-based breast cancer screening. Additionally, Ultrasound (US) and Magnetic Resonance Imaging (MRI) are utilized as supplementary imaging modalities for mammography in specific female subgroups.
Contrast-Enhanced Digital Mammography (CEDM)	The recent advancement in digital mammography involves the intravenous injection of iodinated contrast agents alongside mammography examination, combining Full-Field Digital Mammography (FFDM) and Magnetic Resonance Imaging (MRI).

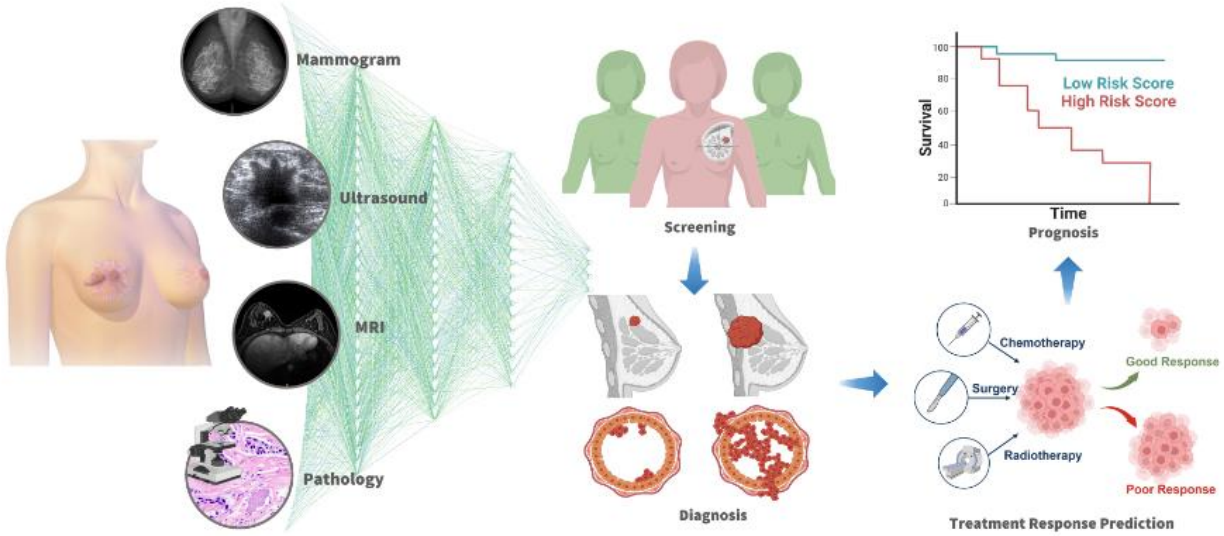


Figure 5: An examination a deep learning in breast cancer with its significant impact.

2.1.7. Mammogram

A specialized medical instrument known as mammography is crafted to capture X-ray images of the breast, aiming at the early detection of breast cancer. Digital mammography has emerged as a potent imaging modality for screening breast cancer at its initial stages (Debelee et al., 2020). By utilizing ionizing radiation and employing low-energy X-rays, this technique generates grayscale images of the female breast. These grayscale images are then scrutinized by radiologists or physicians to identify any anomalies or lumps in the breast tissue. Mammography proves particularly beneficial for investigating masses and uncovering breast cancer in its early phases. However, it is noteworthy that there exists a 10% probability of false-negative results, attributed to the potential resemblance of tumors to normal tissues in mammography images, leading to the risk of misinterpretation. The shift from early X-ray film-based mammograms to digital technology has become prevalent. In this digital approach, a two-dimensional (2D) image of breast tissue is electronically recorded and presented on a monitor. Digital mammography demonstrates heightened sensitivity, especially for women under 50 and those with dense breast tissue. Nevertheless, it encounters three notable limitations: an elevated occurrence of false positives resulting in increased recall rates, heightened instances of false negatives, and augmented exposure to radiation (Debelee et al., 2020). Despite these constraints, extensive follow-up studies have established the efficacy of

mammography, introduced in 1913, in significantly reducing mortality rates associated with breast cancer (Luo et al., 2023).

Figure 6: Mammography image Scanner

Mammography analysis serves as the primary diagnostic method for physicians. However, this approach is susceptible to bias and the impact of physician fatigue. Regrettably, mammography exhibits a low detection rate, with false-negative results ranging from 5% to 30%, influenced by factors such as lesion type, breast density, and the patient's age. To address this, low-dose radiography is implemented in mammography, providing a detailed view of the internal breast structure (Falsi-based et al., 2023).

Advances of Imaging Test in Diagnostic Breast Cancer

There are three main groups of recent developments in mammography. The first type, referred to as comprehensive digital mammography or general digital mammography, uses an electronic substitute for conventional X-ray film. This is the process of converting X-rays into a breast mammography image. In order to detect possible indications of cancer, computer-aided detection (CAD) systems evaluate digitalized mammograms. The third category includes digital breast tomosynthesis (DBT), often known as 3D mammography. DBT is a sophisticated technique that combines several photographs of the breast taken from different perspectives to produce a set of three-dimensional images. It is noteworthy to emphasize that current investigations are examining innovative imaging techniques, such as elastography, electrical impedance imaging (EIT), positron emission mammography (PEM), and scintimammography (molecular breast imaging) (Rahel & Desta, 2020).

Limitations of a Mammography

Mammography has limitations of its own and is not without problems, despite recent advancements in the field. It is not assured that every case of breast cancer will be fully detectable. The possibility of breast transplantation and the distinctive qualities of each woman's breast can lead to variability in mammography accuracy. It is possible to have both false-positive and false-negative results. When mammography is unable to detect cancer, the result is referred to as false-negative; conversely, a false-positive result suggests that cancer may exist even though it is not present. Because breast cancer is difficult to visualize, medical experts frequently rely on older pictures and data (Rahel & Desta, 2020).

2.1.8. Types of Breast Cancer Detection Technique

2.1.8.1. Manual Breast Cancer Detection Technique

Instead of using a computer or specific digital image processing equipment for slide microscopy, a physician or pathologist examines tissue samples under a light microscope as part of the manual breast cancer diagnosis approach. This method involves carefully examining the tissue sample under a microscope by a pathologist, a physician with specific expertise. There may be times when the results require validation by a medical team due to the subjective nature of the pathologist's assessments.

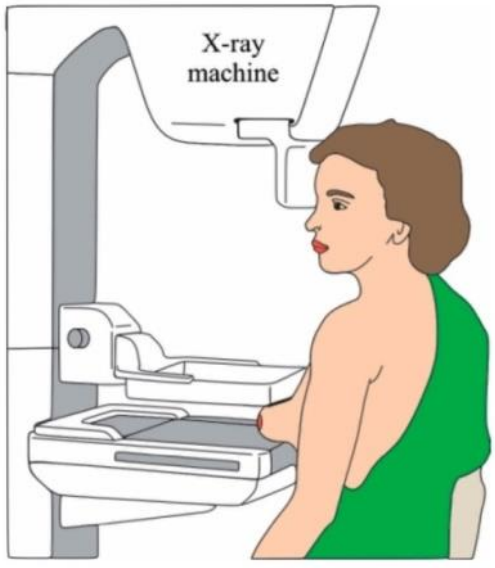


Figure 7: Manual Breast Cancer Detection Technique

2.1.8.2. Digital Breast Cancer Detection Technique

A digital image processing application system created for the purpose of detecting breast cancer from microscopic pictures is used in the automatic breast cancer detection approach. By using digital image processing techniques, this method does not require validation of the results by a group of physicians or a specialist in medicine. The system produces results in a number of categories after receiving a microscopic image that was taken with a microscope mounted on a camera.

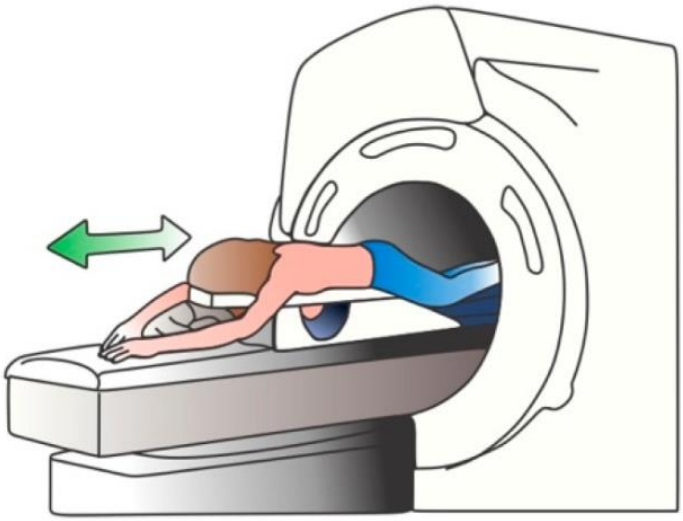


Figure 8: Automatic Breast Cancer Detection Technique

2.2.Machine learning

Machine learning is a branch of artificial intelligence (AI) dedicated to creating algorithms and models that enable computers to learn from data and make predictions or decisions without explicit programming. The ultimate aim of machine learning is to empower computers to adapt to new information, refine their performance over time, and glean insights from experience.

There exist various types of machine learning algorithms, each tailored to specific tasks and datasets. Supervised learning entails training models on labeled data, facilitating tasks like classification and regression. Unsupervised learning, on the other hand, involves training on unlabeled data, enabling tasks like clustering and anomaly detection. Semi-supervised learning amalgamates aspects of both supervised and unsupervised learning, typically utilizing a blend of labeled and unlabeled data. Reinforcement learning revolves around

training agents to make decisions by interacting with environments, receiving feedback in the form of rewards or penalties, and optimizing their behavior to maximize long-term rewards.

Machine learning algorithms further branch out based on their underlying principles and methodologies. This includes Decision Trees, Random Forests, Support Vector Machines (SVM), Neural Networks, and Gradient Boosting Machines (GBM). These algorithms necessitate datasets for training, typically divided into training, validation, and test sets. The model learns patterns and correlations from the training data, with its performance assessed on the validation and test sets to ensure robustness and generalization to unseen data.

2.3.Deep learning

Deep learning, a subset of machine learning, employs representation learning and artificial neural networks to equip computers with the capability to perform tasks traditionally executed by humans. These models can utilize supervised, semi-supervised, or unsupervised learning techniques to extract knowledge from data. Various architectures fall under the umbrella of deep learning, including convolutional neural networks, recurrent neural networks, deep neural networks, deep belief networks, and deep reinforcement learning. These architectures find application in diverse fields such as bioinformatics, drug development, medical image analysis, computer vision, speech recognition, natural language processing, machine translation, material inspection, and board game programming. Notably, in these applications, deep learning models often surpass or equal older approaches in terms of performance.

Deep learning serves as a methodology empowering computer models to categorize objects within images, encompassing tasks like identifying breast cancer. Interestingly, deep learning systems at times achieve higher accuracy than human experts. These models undergo training using multi-layered neural network architectures and extensive datasets with labeled information. Since most deep learning methods leverage neural networks, they are commonly and interchangeably referred to as deep neural networks. The term 'deep' signifies the inclusion of numerous hidden layers within the neural network, with deep networks accommodating up to 150 hidden layers—significantly more than the two or three layers commonly found in conventional neural networks.

Table 2: Features of Benign vs Malignant Tumors

Features	**Benign**	**Malignant**
Growing	Slow Growing	Fast- Growing
Capsulated	Yes	Non- Capsulated
Invasive	Non- Invasive	Invasive & Infiltrate
Metastasis	Non	Yes
Shape	Smooth /oval / lobulated/ regular	Nodular / Stellate / Irregular
The pains	Painful	Painless
Skin	No skin dimpling	skin dimpling
Nipple	No Nipple retraction	Nipple retraction
Prognosis	Good	Bad
Damage to human body	Relatively smaller	Relatively bigger

Utilizing a deep learning network facilitates the precise classification of mammogram images into benign or malignant categories. The foundation of this breast cancer classification deep learning model lies in the deployment of convolutional neural networks (CNNs). Engineered specifically for pattern detection, artificial neural networks, including CNNs, undergo training to discern features within mammography images. In the training phase, extensive datasets of mammography images are fed into the deep learning model, meticulously labeled to indicate benign or malignant tumors. Consequently, the model acquires the ability to recognize the distinct characteristics associated with each lesion class. Following the training phase, the deep learning model becomes adept at accurately categorizing mammography images as benign or malignant. Evaluating the model's performance on a new set of mammography images serves as a benchmark for its accuracy. This capability significantly contributes to the early detection of breast cancer, a critical factor for effective therapeutic interventions.

2.3.1. EfficientNet

Researchers at Google revealed EfficientNet in 2019, a CNN architecture that achieves exceptional accuracy in image classification applications while minimizing processing costs and parameter limitations. Compound scaling ensures effectiveness and efficiency by methodically scaling network width, depth, and resolution. Three strategies—efficient scaling,

new mobile inverted bottleneck convolution block, and uniform scaling utilizing a compound coefficient—make it easier to modify the architecture for changing resource restrictions. Seven models make up the EfficientNet family of CNN architectures, which has demonstrated superior performance in image classification tasks using the ImageNet dataset. The computational cost of the largest model, B7, is higher than that of the lowest, B0. Semantic segmentation and object detection have also made use of these models.(Prodan & Paraschiv, 2023).

2.3.1.1. Single dimension Scaling

In order to increase performance, a convolutional neural network can be scaled by changing its dimensions. The number of layers, or depth, breadth, or channels in a convolutional layer, and the resolution, or resolution of the input images, are the dimensions of a convolutional neural network. Wider networks are faster to train and better at capturing fine-grained characteristics, while deeper networks are probably more performant. Larger pictures improve accuracy by reducing the number of FLOPS overhead.

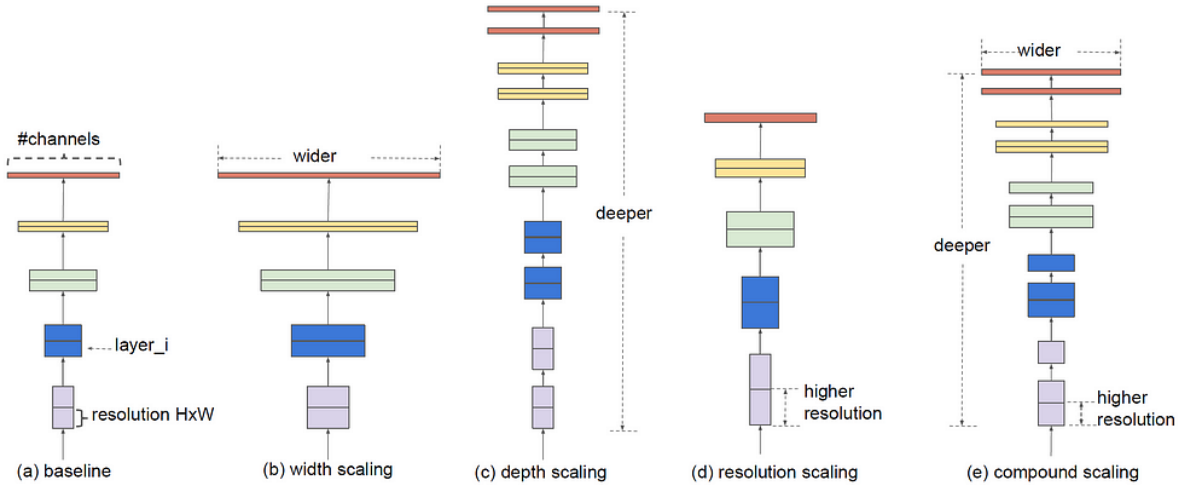


Figure 9: EfficentNetB0 Single Dimension Scaling

2.3.1.2. Compound scaling

The authors of EfficientNet recommend starting with a baseline network (N) and attempting to increase its length (L), width (C), and resolution (W,H) without altering the baseline architecture. This approach differs from the conventional method of attempting to find the optimal layer architecture to obtain a better accuracy. Finding the optimal coefficients for width (w), depth (d), and resolution (r) that optimize network accuracy while adhering to

resource restrictions (memory and number of feasible operations (FLOPS)) is the definition of the optimization issue.

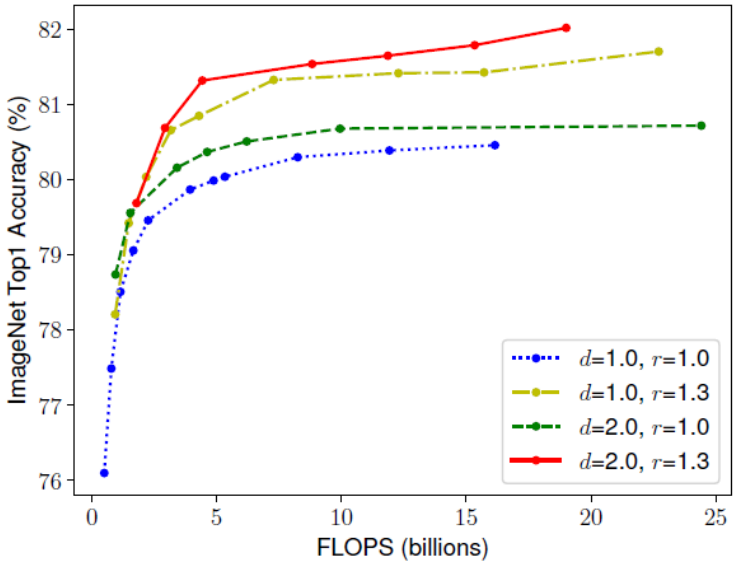


Figure 10: Scaling Network Width for Different Baseline Networks.

$$\max_{d,w,r} \text{Accuracy}\left(N(d,w,r)\right) \qquad \text{Equation 1: Compound scaling}$$

$$\text{Memory}(N) \leq target\_memory$$

$$\text{FLOPS}(N) \leq target\_flops$$

Furthermore, the proposed to constrain that all layers must be scaled evenly using a fixed ratio in order to further minimize the search space <L,C,W,H>. Accordingly, the network's dimensions are described as follows:

depth" $d = \alpha^{\Phi}$ Equation 2: Defining the dimensions of the network

width" $w = \beta^{\Phi}$

resolution: $\gamma = \beta^{\Phi}$

Where Φ is known as the compound coefficient and α,β, and γ are searchable constants using a tiny grid. The user specifies Φ as a coefficient to regulate the quantity of resources that are available. Whereas these resources are assigned to the depth, width, and resolution of the network, respectively, by α, β, and γ.

2.3.2. EfficentNetB0 architecture

Compound scaling, as previously mentioned, expands the network's width, depth, and resolution without changing the operations occurring within its layers. It is crucial to have a robust baseline network as a result. The authors developed EfficientNet-B0, a baseline

network with a mobile size that uses a multi-objective neural architecture to maximize accuracy and FLOPS. The architecture of the model is as follows; Mnas-Net was a major influence.

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Figure 11: EfficientNet-B0 baseline network

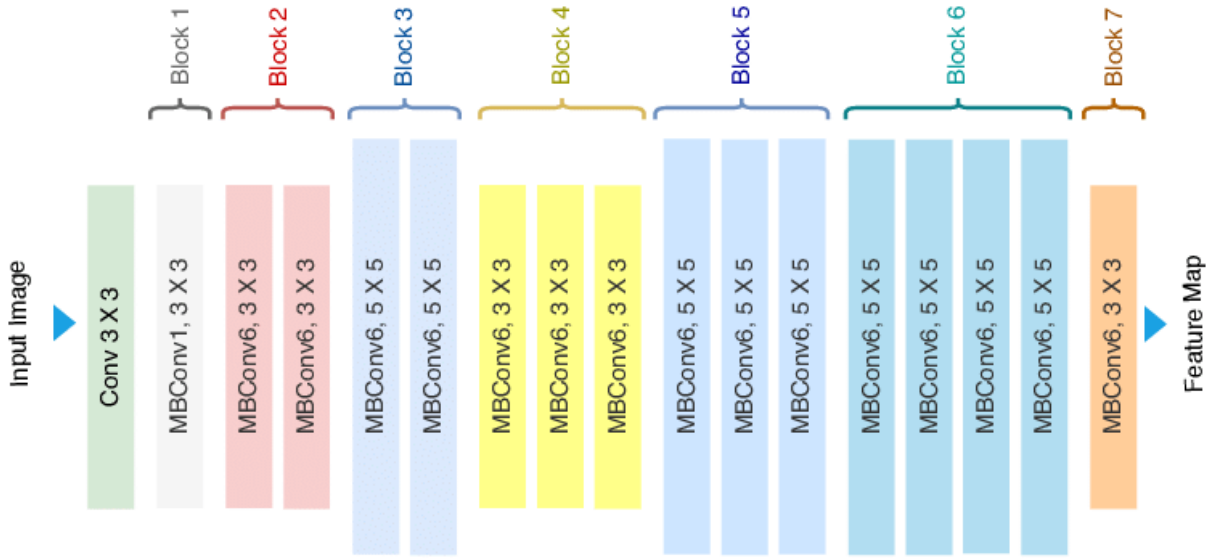


Figure 12: EfficientNet-B0 architecture

The mobile inverted bottleneck MBConv, also known as the inverted residual block with an extra SE (Squeeze and Excitation) block, is the fundamental component of this design. The explanation of these two blocks follows.

The compound scaling method was implemented by the authors from the baseline network, EfficientNet-B0. To discover the parameters α,β, and γ, they first set Φ=1 and used a grid

search with the constraint $\alpha.\beta^2.\gamma^2 \approx 2$ using the equations from the previous section. The findings showed that $\alpha=1.2$, $\beta=1.1$, and $\gamma=1.15$.

Subsequently, the established parameters were locked, and the network's dimension equation was employed to derive a family of neural networks, dubbed EfficientNet-B1 through B7.

2.3.3. EfficientNet Performance

Various benchmarks have been employed to assess the effectiveness and efficiency of EfficientNet. Initially, various scaling techniques were used to scale EfficientNet-B0.

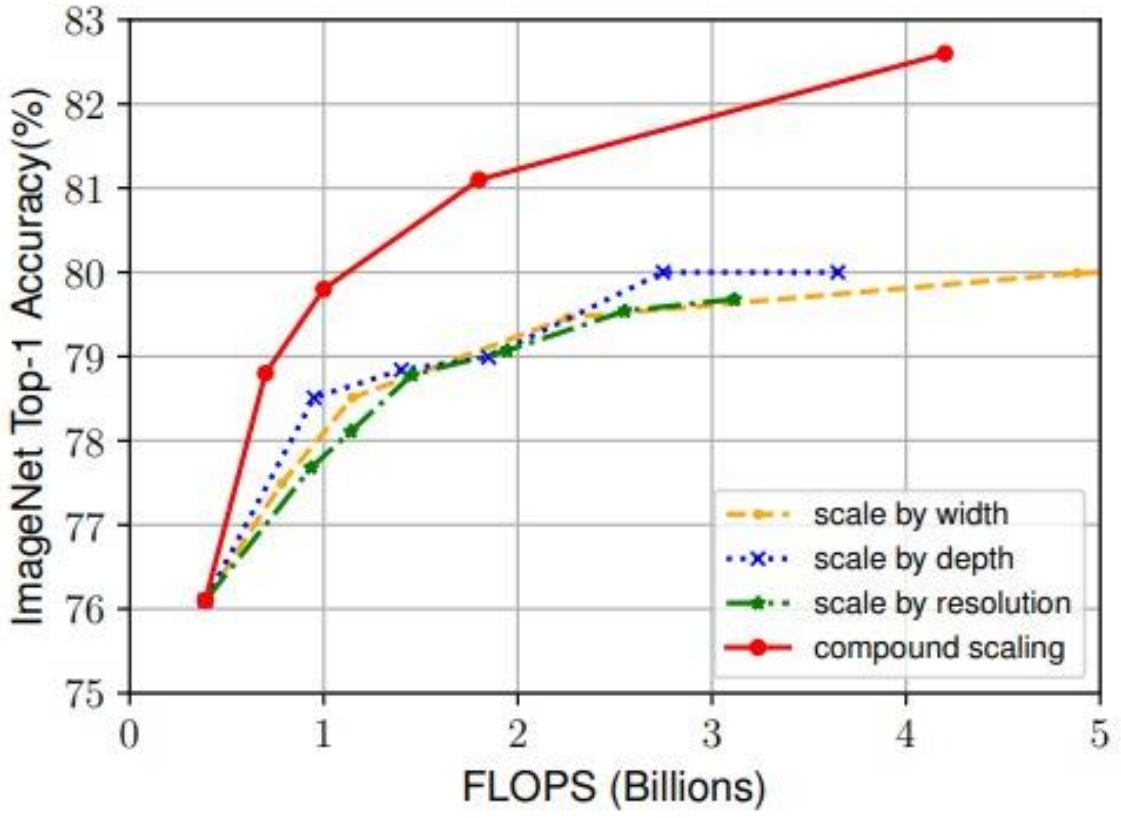


Figure 13: Scaling up EfficientNet-B0 with different methods.

The compound scaling method outperformed the other methods as shown in the figure. Then the EfficientNet family has been tested on ImageNet dataset and it outperformed state of the art models in terms of efficiency and accuracy.

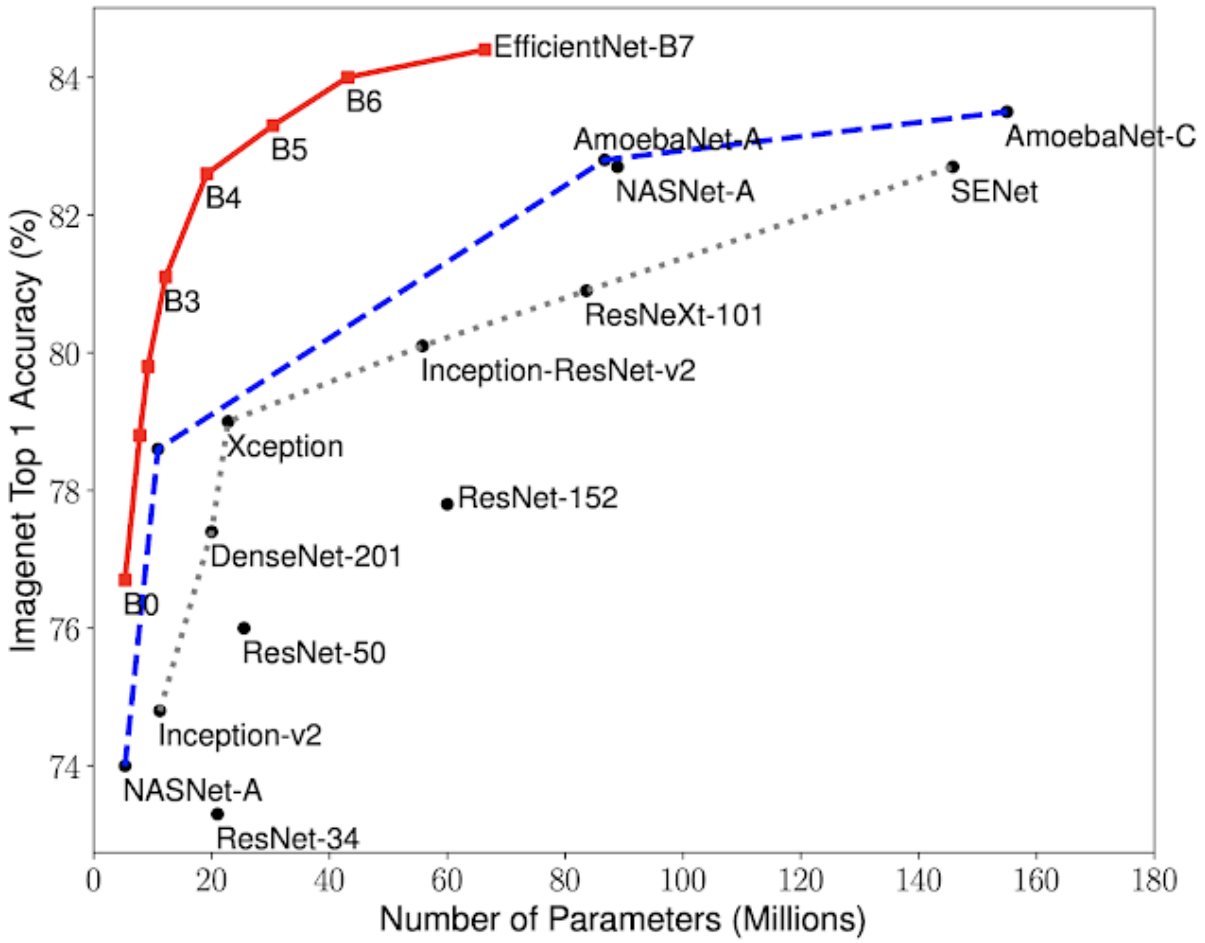


Figure 14: Comparison of EfficientNet family state of the art models

2.3.4. Residual Network, is a convolutional neural network (CNN)

ResNet is a convolutional neural network (CNN) architecture designed for image classification tasks. It was introduced by Kaiming He et al. in 2015 and is known for its use of residual blocks, which help train deep neural networks without the vanishing gradient problem. This problem occurs when gradients become too small to update weights in earlier layers during training. ResNet addresses this issue by introducing skip connections, or shortcut connections, that circumvent one or more layers, allowing gradients to flow directly through the network. The residual block, consisting of two convolutional layers with batch normalization and ReLU activation functions, is the fundamental building block of ResNet. Its architectures come in various variants, including ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, each with varying depth and complexity. Deeper variants like ResNet-50, ResNet-101, and ResNet-152 have shown superior performance on image classification benchmarks like ImageNet. ResNet is a robust CNN architecture that overcomes

challenges in training deep networks by introducing residual connections, ensuring high performance and accuracy.

2.3.5. Image classification with EfficientNet

EfficientNet is a convolutional neural network structure crafted to enhance accuracy and efficiency by scaling the network's depth, width, and resolution in a systematic manner. At its core lies EfficientNetB0, the foundational variant of this design.

Here's a streamlined breakdown of how image classification operates with EfficientNetB0:

- Input Image: The process starts with an input image, whose size may vary but is typically adjusted to a fixed size suitable for the network.
- Preprocessing: Prior to feeding the image into the network, preprocessing tasks like normalization (scaling pixel values within a certain range) and data augmentation (randomly altering images to augment diversity in training data) may occur. This enhances the model's generalization and robustness.
- Convolutional Layers: The image traverses a sequence of convolutional layers, which extract features at varying levels of abstraction. In EfficientNet, these layers' depth and width are adjusted according to a predetermined formula, achieving a balance between accuracy and efficiency.
- Pooling Layers: At intervals, pooling layers (like max pooling) are employed to reduce the spatial dimensions of feature maps while preserving crucial information. This aids in diminishing computation and managing overfitting.
- Fully Connected Layers: Following several convolution and pooling layers, feature maps are flattened and passed through fully connected layers. These layers execute the final classification, associating extracted features with output classes.
- Softmax Activation: In the final layer, a softmax activation function typically generates class probabilities. This transformation converts raw scores from the network into probabilities, signifying the likelihood of the input image belonging to each class.
- Loss Calculation: Predicted probabilities are juxtaposed with ground truth labels using a loss function (e.g., categorical cross-entropy). During training, the objective is to minimize this loss, often achieved through optimization algorithms like stochastic gradient descent (SGD) or its variants.

- Training: The complete network undergoes end-to-end training utilizing a labeled dataset. Throughout training, the network fine-tunes its parameters (weights and biases) to minimize the loss function, thereby honing its ability to classify images accurately.
- Inference: Once trained, the model can be deployed for inference on new, unseen images. The input image is fed into the trained network, and the output class with the highest probability is considered the predicted label for the image.
- EfficientNetB0, with its streamlined architecture, offers an optimal blend of model size, accuracy, and computational efficiency, rendering it suitable for diverse image classification tasks, particularly on devices with limited resources.

2.4.Related Works

To comprehensively grasp the overarching approach, methodologies, and outcomes of prior research, this section meticulously examines seminal publications relevant to the classification system for breast cancer. The literature extensively explores early cancer diagnosis and prediction, with a focus on various research papers employing techniques such as CNN, ANN, GAN, EfficentNet, and Deep Learning.

Based on research (Tan & Le, 2019), a novel scaling technique utilizing a basic compound coefficient is suggested for model scaling in Convolutional Neural Networks (ConvNets). ResNet and MobileNets can be scaled up with success using this technique. The authors then create EfficientNets, a new baseline network that outperforms earlier ConvNets in terms of accuracy and efficiency. EfficientNet-B7 performs with 91.7% accuracy on transfer learning datasets and 84.4% top-1/97.1% top-5 accuracy on ImageNet.

A deep learning-based convolutional neural network (ConvNet) that lowers human error in breast cancer diagnosis is shown in the study (Mahmood et al., 2022). To identify worrisome areas in mammograms as benign or malignant, the model makes use of task-specific characteristics and classification tasks. In terms of classifying breast masses on mammograms, the model obtained high sensitivity of 0.99, test accuracy of 0.97, training accuracy of 0.98, and an AUC of 0.99. This approach helps physicians quickly calculate mammograms, diagnose breast masses, schedule treatments, and monitor the course of their patients' illnesses.

Using mammography pictures from the Myung Sung Comprehensive Specialized Hospital repository, the research (Bimrew Sendekie Belay, 2022) attempts to create a deep learning model for the detection and classification of breast cancer. Using a Convolutional Neural Network (CNN) and pre-trained models, the methodology produced the maximum accuracy of 88% for InceptionV3. Insufficient data for certain classes, however, suggests that future efforts should concentrate on gathering more information for every class.

In order to reduce the difficulties involved with manual analysis, the study (Khan et al., 2021) suggests a Convolutional Neural Network (CNN)-based model for the detection of breast masses in Ethiopia. For effective feature extraction, the model combines the Region of Interest (ROI) and Region Proposal Network (RPN) components from the quicker R-CNN. With a detection accuracy of 91.86%, sensitivity of 94.67%, and AUC-ROC of 92.2%, it effectively distinguishes between benign and malignant abnormalities in mammography pictures.

The paper (Debelee et al., 2018) presents an automated feature extraction technique for breast cancer diagnosis and classification based on Convolutional Neural Networks (CNNs). The method combines Principal Component Analysis (PCA) with K-Nearest Neighbors (KNN) to classify mammograms into normal and abnormal categories, thereby reducing feature dimensionality. With accuracies of 98.75% and 98.90% on the Imaging Image Analysis Society (MIAS) and Digital Database for Screening Mammography (DDSM) datasets, respectively, the experimental results show promising outcomes.

Insights into patient mammography density and pathology are provided by the research (杜彬 & Amaliyyah, 2021), which presents a deep learning model for Multi-Label Breast Cancer Screening. The model yields a 13.25% f1_score, 53% hamming loss, 36.7% coverage error, and only 12.5% precise matches, indicating subpar performance in feature extraction and classification. The paper suggests adding more breast datasets to the analysis, refining the loss function, and optimizing feature extractors in order to improve the model.

The study (Rahel & Desta, 2020) proposes a preprocessing technique for mammography that removes background artifacts and pectoral muscles in order to enhance image quality and facilitate Region of Interest (ROI) extraction. Through thresholding and morphological techniques, the technique suppresses background artifacts, segments pectoral muscles, and enhances image quality by applying a median filter and Contrast Limited Adaptive Histogram Equalization (CLAHE). Utilizing a mini-MIAS database, the method was assessed and was able to segment pectoral muscles with 97% accuracy. This approach is regarded as suitable for early detection of breast cancer.

The research (Aljuaid et al., 2022)made use of the publicly available BreakHis dataset, which comprises 7909 samples from 82 different patients. The average accuracies in image-based binary classification (benign or malignant) and multi-class classification (eight classes) for pre-trained neural networks ResNet18, Inception-V3Net, and ShuffleNet were 97.81%, 96.07%, and 95.79%, respectively. The results point to the necessity of considering additional reliable datasets and offer recommendations for enhancing the model's breast cancer detection accuracy.

Using the assessed Wisconsin Breast Cancer Datasets (WBCD), the study's (Aslam & Cui, 2020) recommended CAD (Computer-Aided Diagnosis) technique for diagnosing breast cancer combines a Deep Convolutional Neural Network with a Softmax classifier. Surprisingly, despite a flaw in the way legacy hardware was implemented, the suggested classifier managed to attain high accuracy scores of 100% and 99.1% for two different datasets.

Using Convolutional Neural Network (CNN) Models with the BCDCNN technique applied to the mini-MIAS dataset, the researchers (Qasim & Ouda, 2020) achieved an outstanding accuracy of 99.4%. The study used a dataset comprising of 322 mammography pictures, including 52 malignant and 270 normal cases.

Utilizing the Wisconsin Breast Cancer Dataset (WBCD) from the UCI repository, experiments were conducted as part of the research. (Karthik et al., 2018) Recursive feature elimination (RFE) was used for feature selection, and a deep neural network (DNN) was used

as the classifier model. Even with limitations on algorithm training time and no computationally better GPU-integrated system, the performance metrics (accuracy, sensitivity, specificity, and recall) attained were 98.62%, which is higher than previous state-of-the-art techniques.

The recommendation is to use CNN Generative Adversarial Network (GAN), which is 10 times quicker than CNN, and CNN as two deep learning technologies in the study(Guan & Loew, 2019). The suggested method extracts features from input mammograms using a pre-trained VGG-16 model, then applies GAN for picture augmentation. A neural network (NN)-classifier is then trained using these features that were extracted. With the use of the DDSM database, a consistent average validation accuracy of 91.48 percent was attained for the distinction between benign and malignant cases.

Within the research (Al-Dhabyani et al., 2019), The suggested method uses two different deep learning approaches—a Convolutional Neural Network (CNN) approach and a Transfer Learning approach—for the categorization of breast cancer. By integrating conventional techniques with GAN-based augmentation, the researchers were able to attain a higher degree of accuracy (99 percent). They utilized local datasets consisting of 780 photos, of which 133 are normal, 437 benign, and 210 malignant.

Table 3: Summary of related works

Ref.	Title and Authors	Dataset	Models and algorithm	Performance / Accuracy	Gaps /Limitations
1	Title: Association of risk factors and breast cancer among women treated at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia: a case–control study Authors: Fatuma Hassen ,1,2 Fikre Enquselassie,2	BreakHis 7909 breast cancer histopathology image training–testing data (90%–10%, 80%– 20%, and 70%–30%)	pre-trained deep CNN models: VGG16, VGG19, and ResNet50	92.60% accuracy, 95.65% area under ROC curve (AUC), and 95.95% accuracy precision score (APS) for 90–10 training–testing data splitting	• Biasing • Over-fitting and under-fitting.

	Ahmed Ali,2 Adamu Addissie,2 Girma Taye,2 Aster Tsegaye,1 Mathewos Assefa3				
4	Title: Breast Cancer Classification using Deep Convolutional Neural Network Authors: Muhammad Aqeel Aslam1,a Aslam and Daxiang Cui	Wisconsin Breast Cancer Datasets (WBCD)	deep convolutional neural network (DCNN) with Softmax classifier and architectures: LeakyRELU, max-pooling layers,	100% and 99.1% for two different datasets. Dataset 1: 699 samples and dataset2: 683 samples	needs substantial effort on legacy hard for best performance
8	Title: Breast Cancer Detection using Artificial Neural Networks Authors: Md Haris Uddin Sharif	UCI machine learning repository	ANN	98.24%	Needs high performance GPU-incorporated systems
9	Title: Breast Cancer Detection Based on Deep Learning Technique Authors: Nur Syahmi Ismail and Cheab Sovuthy	IRMA (Image Retrieval in Medical Application)	CNN classification using VGG16 and ResNet50 model networks	94% and 91.7%	Over fitting
5	Title: A Review Comparative Mammography Image Analysis on Modified CNN Deep Learning Method, Authors: Ramadhani, Siti First, Siti Ramadhani	Wisconsin Breast Cancer Datasets (WBCD)	DCNN with Softmax classifier	99.1%	fault in the implementation of legacy hardware

CHAPTER THREE

3. MATERIALS AND METHODS

3.1.Methodology overview

This chapter provides insights into the collection, classification, and processing of images, along with the creation of a classification model using the employed modeling methodologies. Subsequently, the performance of the constructed model is assessed using specified evaluation measures.

3.2.Research Flow

This study utilizes experimental methodologies to identify diseases in mammogram images, following a systematic three-step process. In the initial phase, the research defines the problem domain through a thorough literature review, establishes clear thesis objectives, and prepares the data for the research prototype. The dataset, obtained from the Myung Sung Comprehensive Specialized Hospital in Ethiopia, is meticulously labeled by medical professionals and categorized into training, validation, and testing classes. The second step involves preprocessing the data's. In the third phase, the thesis is implemented by applying the designed model with appropriate tools and techniques.

The model undergoes training and testing using the relevant data, with continuous performance assessment throughout the training phase. The evaluation stage identifies the best-performing model, which is then compared with previously trained models. The primary goal of the study is to establish a correlation between the presences of diseases in mammogram images.

This research adopts an experimental research approach for breast cancer classification within a deep learning environment of EfficentetB0 architecture, encompassing the following components.

- Observation: Carefully examining a phenomenon and exploring the factors contributing to it.
- Induction: Developing hypotheses to serve as broad explanations for the observed phenomenon.
- Deduction: Devising experiments specifically crafted to assess the formulated hypotheses.

- Testing: Executing procedures to evaluate hypotheses and gather pertinent data.
- Evaluation: Analyzing the gathered data and constructing a comprehensive theory.

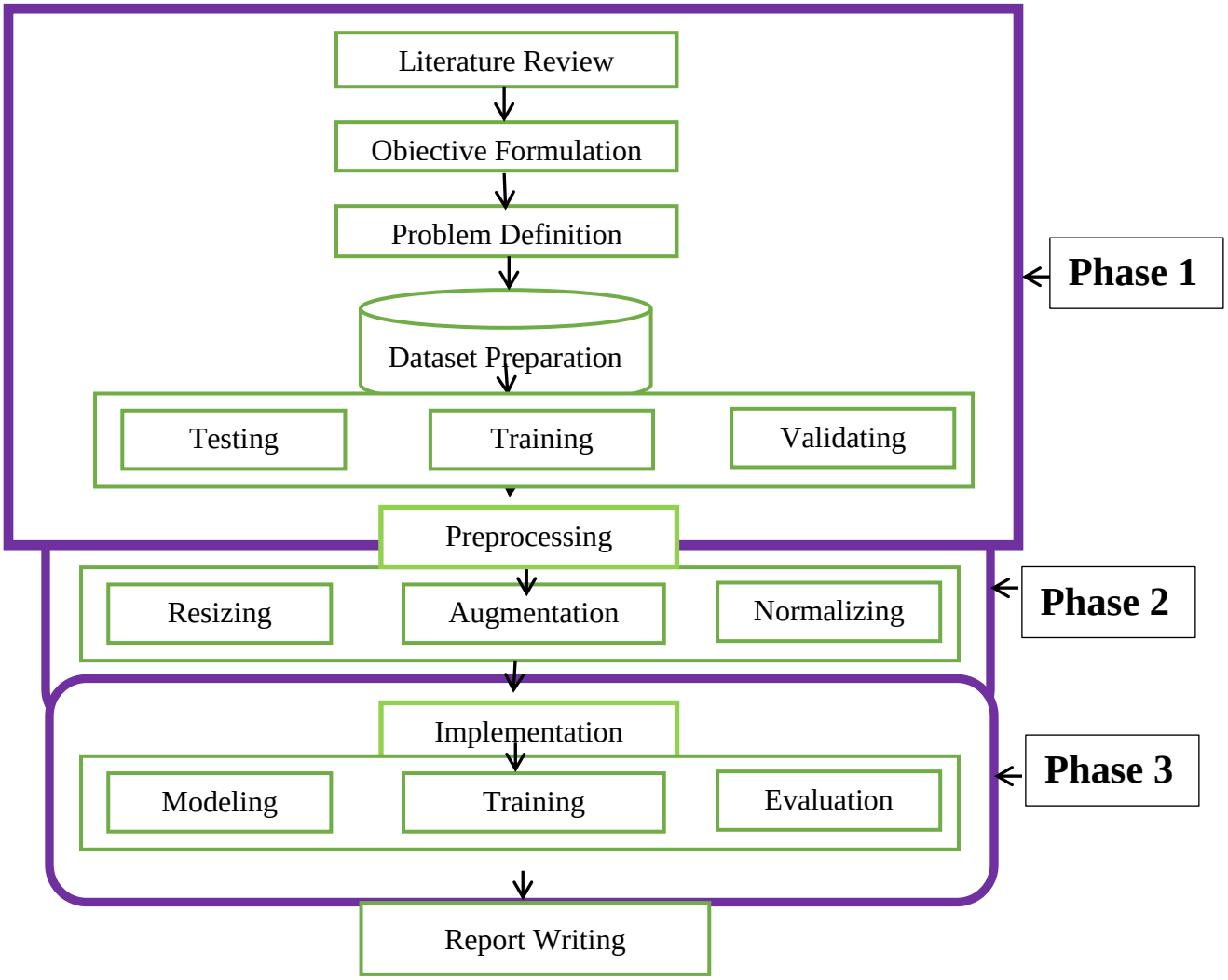


Figure 15: Research Flow

3.3.Preprocessing and Dataset Preparation

In order to train a neural network model for the classification of breast cancer, this thesis uses deep learning techniques. The main dataset consists of mammography imaging data that was acquired from Myung Sung Comprehensive Specialized Hospital in association with medical specialists and the publicly available image dataset. There are three subsets of the dataset: test, validation, and training. Model performance is evaluated and parameters are refined on the validation subset, while the test subset is used for post-training model evaluation. An equal number of images from each class are used for testing and validation in order to avoid overfitting and guarantee impartial weight updates.

The process of converting analog images into digital formats and carrying out several procedures to produce improved images or extract pertinent data is known as digital image processing. Preprocessing is crucial for normalizing data and eliminating duplicate or unclear information from mammography screening images, which are more difficult to understand than other medical imaging modalities, when it comes to deep learning investigations.

A strong foundation for successful deep learning models is provided by the preprocessing stage, which is essential for improving image quality and removing unnecessary parts. It's critical to divide the data into separate sets for training, validating, and testing when utilizing photos to train deep learning models. 70% of the data are used for training, 15% are used for validation, and 15% are used for testing in a split that is frequently utilized.

3.4. Augmentation

Data augmentation serves as an attractive strategy to mitigate overfitting, enhance model generalization, and improve overall performance. Overfitting occurs in CNNs when models excessively grasp the details of training data but struggle to generalize effectively to unseen data, leading to suboptimal predictions for testing data. This issue commonly arises when the size of the training dataset is relatively small compared to the number of model parameters that require learning.

Data augmentation involves artificially generating new sample images by applying transformations, such as flipping and rotation, to the existing data. Typical data augmentation techniques for mammography images include horizontal flipping, rotations at various angles (90, 180, and 270 degrees), jittering, and random scaling. These augmentation methods produce diverse training samples that account for the potential variations in tumor orientations and sizes. Importantly, augmentation techniques do not alter the fundamental pathology of the masses (Abdelhafiz et al., 2019).

Data augmentation, as used in the EfficientNetB0 algorithm for image classification tasks, is the process of introducing different random changes to the training images to expand the size and diversity of the dataset. During training, this strategy is used to improve the model's handling of fresh, unseen images and to increase its capacity for generalization.

The types of data augmentation commonly used in EfficientNetB0 include:

1. Random Rotation: A specific degree of random rotation is applied to the training images. This aids in the model's learning of object recognition from various orientations and angles.
2. Scaling: Random size scaling is applied to the training photos. This makes the model more resilient to changes in object sizes and aids in its ability to identify objects at various scales.
3. Flipping: Horizontally flipped training images are used. By doing this, the model becomes more adept at handling images of varying orientations and learns to be invariant to left-right flips.

Data augmentation adds to the quantity of training data that the model has access to by implementing these arbitrary changes on the training pictures. This enhances the model's generalization to fresh, unseen images and helps minimize overfitting. Improved robustness to changes in object appearance, including varied rotations, scales, and orientations, is another benefit of this optimization.

Effectively adding variation to the training dataset through data augmentation allows the model to acquire more resilient and comprehensive features, making it a crucial strategy in the EfficientNetB0 algorithm. The overall performance and accuracy of the model in picture classification tasks are enhanced by this augmentation process.

In summary, by combining various preprocessing methods, medical images are made suitable for classification using the EfficientNet algorithm, resulting in excellent classification task accuracy and efficiency.

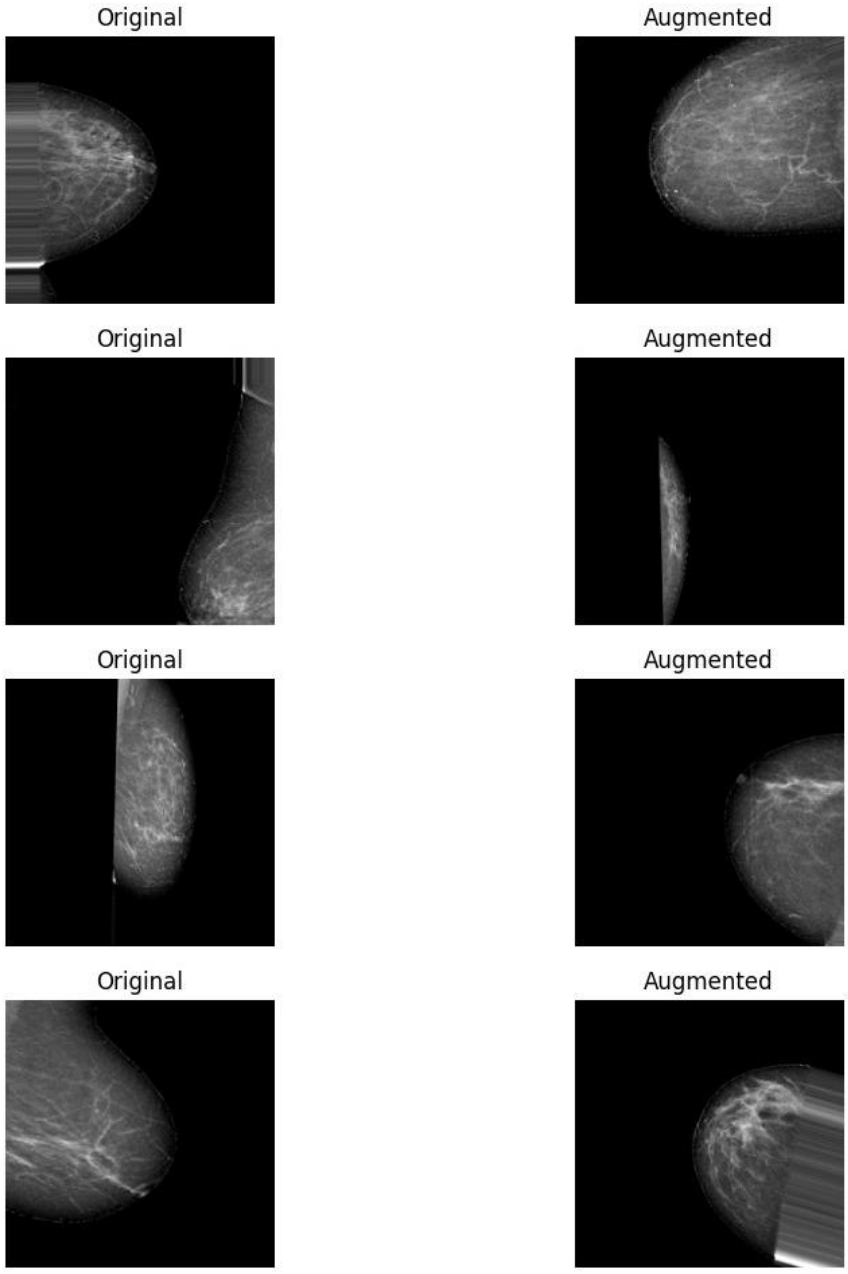


Figure 16: Original and Augmented images

3.5. The proposed architecture

The system design commences with the preparation and labeling of training, validation and testing datasets. In this initial stage, images undergo data normalization, cropping, and categorization into benign and malignant diseases. The preprocessed images exhibit dimensions of 224x224x3, highlighting the region of interest effectively. The stacked layers

extract essential features from each image, and subsequent classification is carried out based on these extracted features for model creation.

Throughout the training phase, the model extracts pertinent features from breast images, and the classifier identifies patterns indicative of benign and malignant breast conditions. Model performance is evaluated using the validation dataset during the training process. In essence, once the model is trained, the validation set is employed to assess its overall performance. The model ultimately provides class predictions, representing the probability of an image belonging to either the benign or malignant class based on the training data.

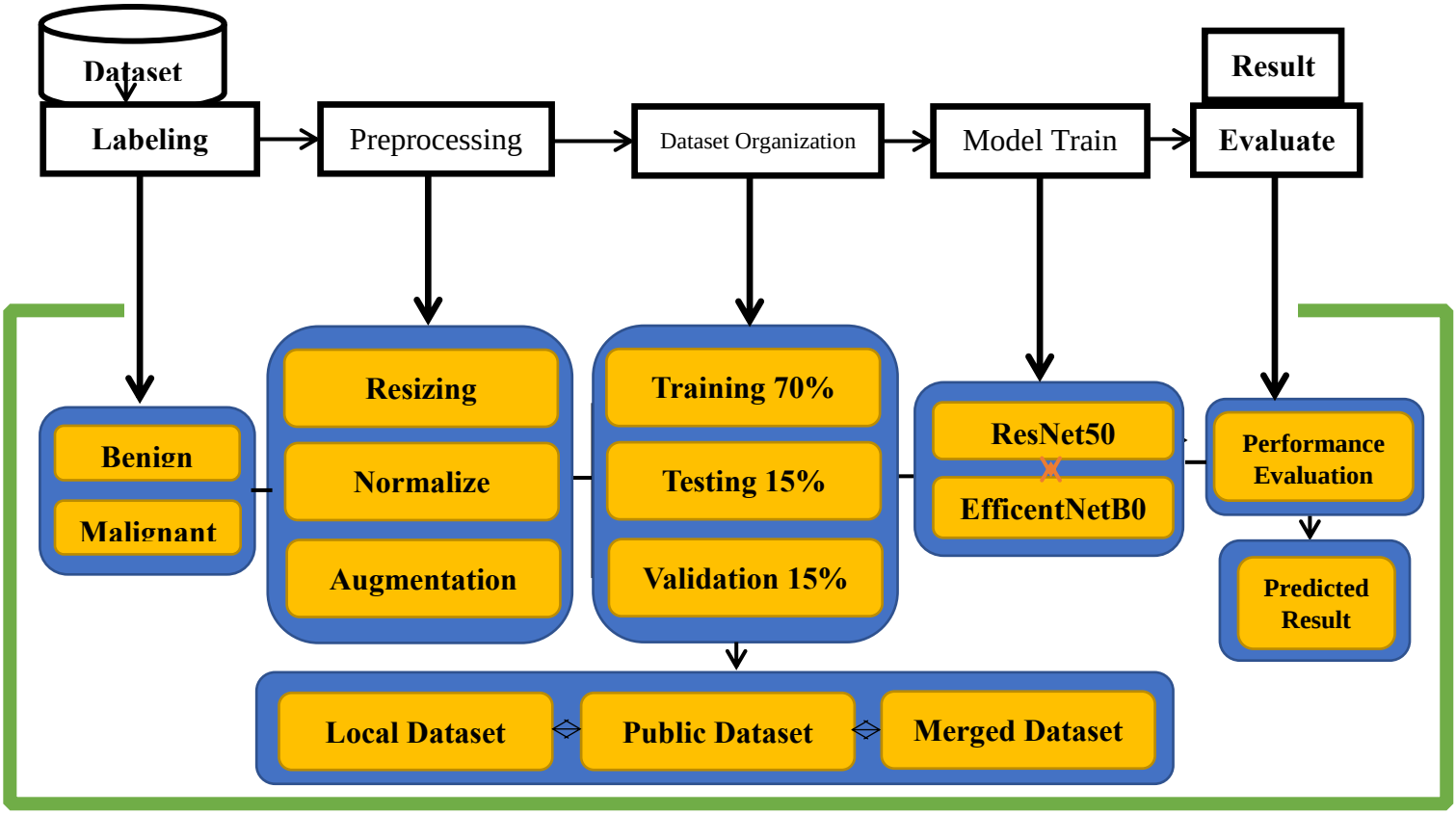


Figure 17: The proposed Architecture for Breast Cancer Classification

3.6. Working Environment

The operational setting encompasses the necessary hardware and software resources essential for executing this task effectively and achieving favorable results.

3.6.1. Hardware Utilities

The laptop was utilized to carry out various aspects of this project, including model training and testing, as well as diagram creation. It possesses the following specifications: Enhanced

Graphics Capability: Upgraded AMD Radeon(TM) Vega 8 Graphics Processor: AMD Ryzen 5 2500U with Radeon Vega Mobile Gfx (8 CPUs), operating at approximately 2.0GHz Solid-State Drive (SSD): Configured storage capacity of 625GB Random Access Memory (RAM): 8 GB installed memory.

3.6.2. Software Utilities

To execute the study, we will employ a range of software components, including: Windows 11 Pro 64-bit Anaconda3 distribution with machine learning libraries such as TensorFlow, OpenCV, Keras, PyTorch, NumPy, Pandas, Matplotlib, Scikit-learn, along with additional utilities and applications like Google Colab, Microsoft Office Word and Visio, and Mendeley.

3.7.Evaluation methods

Studying a model is crucial to determine its effectiveness and performance. In computational tasks like classification and detection, evaluation metrics such as accuracy, precision, recall, and F1-score are utilized to assess the assignment of instances to their respective classes. The confusion matrix, particularly the False Negative value, is instrumental in calculating these metrics, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

3.7.1. Confusion matrix

The confusion matrix serves as a tool to assess the effectiveness of a classifier model in classification tasks by summarizing the predicted classifications against the actual classifications in a tabular format. It provides insights into how well a model can accurately classify instances into different categories or classes.

These metrics are computed based on categorization results, revealing the model's performance across each class.

In the specific context of breast cancer classification using the model, It is trained to predict whether an image exhibits malignant (cancerous) or benign (non-cancerous) features in breast tissue. In this scenario, the confusion matrix typically takes the form of a 2x2 matrix since there are two classes (benign and malignant) that the model aims to classify.

The confusion matrix comprises four cells, each representing different prediction outcomes:

- True Positive (TP): Instances correctly predicted as malignant.
- False Positive (FP): Instances predicted as malignant but are actually benign (also known as Type I error).

- False Negative (FN): Instances predicted as benign but are actually malignant (also known as Type II error).
- True Negative (TN): Instances correctly predicted as benign.

These components of the confusion matrix are instrumental in assessing model performance metrics like accuracy, precision, recall, and F1-score. For instance, accuracy gauges the proportion of correctly classified instances (TP + TN) out of all instances, while precision measures the ratio of true positives to all instances predicted as malignant (TP + FP).

Thus, the confusion matrix serves as a valuable tool to evaluate the accuracy, reliability, and efficacy of CNN-based breast cancer classification systems. Below is an illustration of a confusion matrix for a binary classification model:

Table 4: The confusion matrix

Classes		**Actual**	
		Benign	Malignant
Predicted	Benign	True Positive (TP)	False Positive (FP)
	Malignant	False Negative (FN)	True Negative (TN)

Table 4 illustrates the individual cells of the confusion matrix for binary classes. Here's the breakdown:

TP (True Positive): Positive observations correctly predicted as positive.

FP (False Positive): Positive observations incorrectly predicted as negative.

TN (True Negative): Negative observations correctly predicted as negative.

FN (False Negative): Negative observations incorrectly predicted as positive.

Accuracy

Accuracy represents the percentage of correct predictions made by the model and is a commonly used metric for evaluating classification models. It indicates how often the classifier correctly predicts the class labels. Accuracy is particularly relevant when there are

an equal number of observations in each class, and all predictions carry equal importance. The value of accuracy ranges between 0 and 1, where a value of 1 signifies that all samples in the dataset are correctly classified, while a value of 0 indicates that none of the samples are classified correctly. The accuracy metric is calculated using the following equation:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Equation 3: Accuracy Formula

This formula determines the proportion of correct predictions out of the total number of predictions made by the model.

Precision

Precision, also known as positive predictive value, quantifies the accuracy of positive predictions made by the model. It measures the ratio of true positive predictions to the total number of instances predicted as positive. In other words, precision answers the question: "Out of the items predicted as positive, how many are actually positive?" In the context of our problem, precision indicates how accurately the classifier identifies benign samples. The precision metric is computed using the following formula:

$$Precision = \frac{(TP)}{(TP + FP)}$$

Equation 4: Precision Formula

This metric is more important in cases where we are more concerned with finding the maximum number of positive classes.

Recall

Recall, also known as sensitivity, measures the proportion of true positive cases that the model correctly predicts out of all actual positive cases. In our scenario, recall assesses the classifier's ability to identify all existing malicious instances as 'Malicious'. The recall metric is computed using the following formula:

$$Recall = \frac{(TP)}{(TP + FP)}$$

Equation 5: Recall Formula

F1- score

The F1-score, represented by the following equation, is the harmonic mean of precision and recall values. A high F1-score, close to 1, indicates a well-performing model, while a low

score, closer to 0, signifies poor predictive capability. This metric is particularly crucial for assessing models with unbalanced classes. The F1-score of a model is computed as:

$$F1\text{- score} = \frac{(2 * ((precision * recall))}{((precision + recall)\ )}$$

Equation 6: F1-Score Formula

3.7.2. Receiver Operating Characteristic (ROC) Curve:

In image classification tasks, including those using EfficientNetB0 architecture, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) score are important evaluation metrics, especially in binary classification scenarios. Here's an explanation of ROC AUC score in the context of image classification:

Receiver Operating Characteristic (ROC) curve

- The ROC curve is a graphical representation of the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) for different threshold values.
- The true positive rate (TPR) is the proportion of correctly classified positive instances (e.g., images belonging to the positive class) out of all actual positive instances. It is calculated as TP / (TP + FN), where TP is the number of true positives and FN is the number of false negatives.
- The false positive rate (FPR) is the proportion of incorrectly classified negative instances (e.g., images belonging to the negative class) out of all actual negative instances. It is calculated as FP / (FP + TN), where FP is the number of false positives and TN is the number of true negatives.
- The ROC curve plots TPR against FPR at various threshold settings, typically ranging from 0 to 1. Each point on the curve represents a different threshold value.

Area Under the Curve (AUC):

- The Area Under the Curve (AUC) score quantifies the overall performance of a binary classification model based on the ROC curve.
- AUC measures the area under the ROC curve, ranging from 0 to 1. A higher AUC value indicates better model performance.
- AUC provides a single scalar value that summarizes the model's ability to distinguish between positive and negative instances across all possible threshold values.

- An AUC score of 0.5 suggests that the model performs no better than random guessing, while an AUC score of 1.0 indicates perfect discrimination between positive and negative instances.

In the context of image classification using EfficientNetB0, the ROC AUC score provides valuable insights into the model's ability to classify images correctly, particularly in binary classification tasks where distinguishing between two classes (e.g., benign and malignant) is crucial. A higher ROC AUC score implies that the model effectively separates positive and negative instances, demonstrating its discriminative power and overall performance.

CHAPTER FOUR

4. PROPOSED SYSTEM

4.1.Experimental Setup

The experiments were conducted on a personal laptop with AMD Ryzen 5 2500U processor, Radeon Vega Mobile Gfx, and Windows 11 Pro 64-bit Operating System. Deep learning frameworks like Python, TensorFlow, Keras, Google Colab, and Microsoft Office Word and Excel were used.

4.2.Dataset organization

This study uses image datasets categorized as benign and malignant, divided into training, testing, and validation sets. The training set teaches the model to recognize patterns, while the validation set evaluates the model's generalization to unseen data. The number of epochs ranges from 1 to 40, and selecting the appropriate number is crucial to prevent underfitting or overfitting. To improve model accuracy, tactics like increasing data, expanding the network, and extending epochs can be implemented. The study also presents findings on breast cancer classification using mammogram imaging using a locally collected dataset sourced from Myung Sung Comprehensive Specialized Hospital comprises 2376 mammography breast cancer images, with 1865 labeled as benign and 511 as malignant. To assess the model's classification performance, the dataset was divided into training, validation, and testing subsets, and three data split configurations were employed to analyze the impact of data split size on network performance.

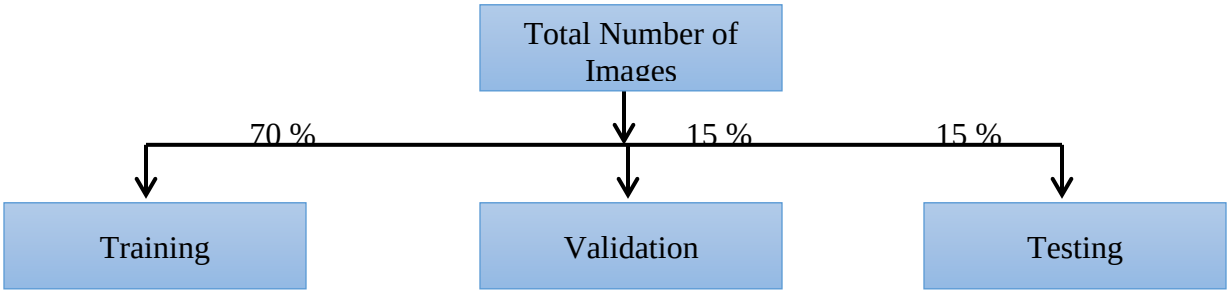


Figure 18: Train, test and validate split ratio on the dataset

Under fitting and Overfitting

The primary issue known as underfitting arises when the model's performance on the training set falls short of the expected error of an ideal model. This indicates a high bias in the model.

To tackle underfitting, it's recommended to extend the training duration to allow the model to adequately optimize its parameter values.

In contrast, overfitting occurs when the model excels on the training dataset but fares poorly on the test dataset, indicating high variance. This suggests the model struggles to generalize insights from the training data. To address overfitting, one strategy is to acquire more data and employ regularization techniques. This ensures the model is exposed to a sufficient number of training examples to effectively learn patterns. Additionally, excessive complexity of the neural network can contribute to overfitting, making regularization techniques essential for achieving superior results.

In the domain of deep learning, various methods combat overfitting:

Dropout randomly deactivates certain neurons during training to reduce the model's reliance on specific inputs, thus enhancing robustness. Early stopping prevents the model from excessively memorizing the training dataset, thereby mitigating overfitting.

Data augmentation enriches the diversity of the training dataset by applying transformations such as rotation, flipping, or scaling to existing data samples, facilitating a more comprehensive learning process.

Table 5: Local Dataset organization

Category	**Benign**	**Malignant**	**Total**
Training	1305	357	1662
Validation	280	77	357
Testing	280	77	357
Total	1865	511	2376

Table 6: Public Dataset organization

Category	**Benign**	**Malignant**	**Total**
Training	1305	357	**1662**
Validation	280	77	**357**
Testing	280	77	**357**
Total	1865	511	2376

```
# Number of augmented images to display
num_augmented_images = 5

# Generate augmented images
augmented_images, _ = train_data.next()

# Display the original and augmented images
fig, axes = plt.subplots(num_augmented_images, 2, figsize=(10, 15))
fig.suptitle('Original and Augmented Images', fontsize=16)

for i in range(num_augmented_images):
    # Display original image
    axes[i, 0].imshow(augmented_images[i])
    axes[i, 0].set_title('Original')
    axes[i, 0].axis('off')

    # Display augmented image
    augmented_img, _ = train_data.next()
    axes[i, 1].imshow(augmented_img[0])
    axes[i, 1].set_title('Augmented')
    axes[i, 1].axis('off')

plt.show()
```

Figure 19: Loading and displaying Sample Data

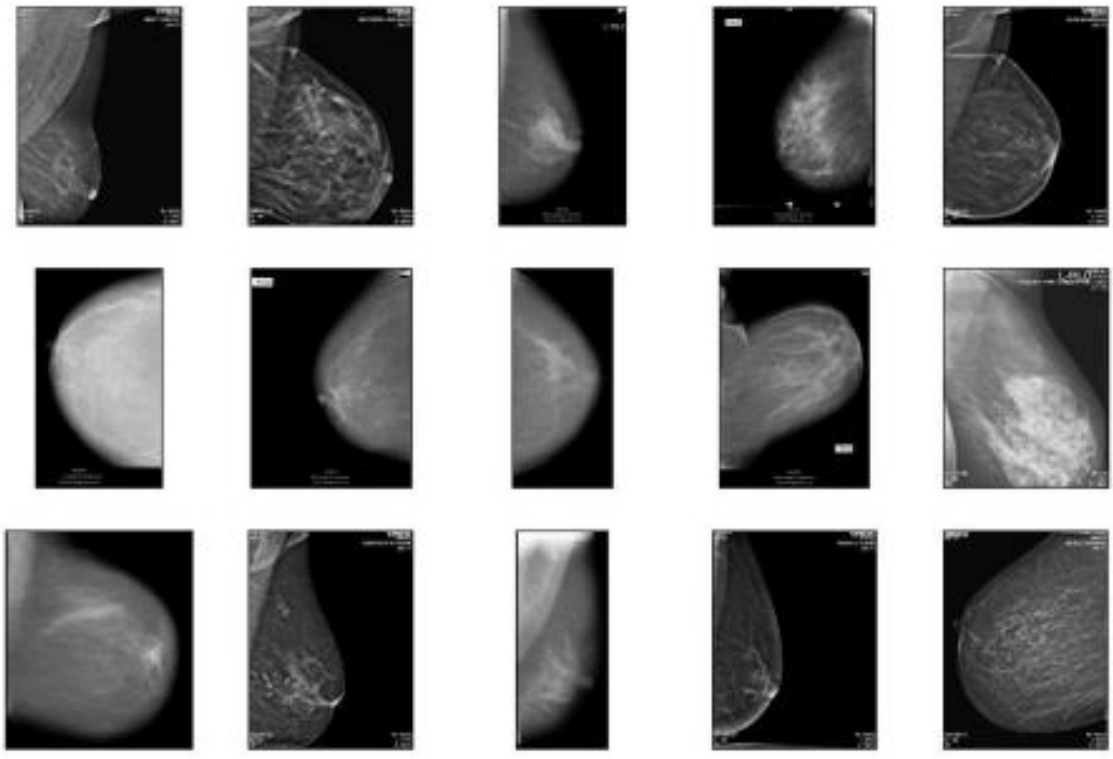


Figure 20: Sample Breast Images from Myung Sung Comprehensive Specialized Hospital

This code snippet displays original and augmented images generated by the data augmentation process. It specifies the number of augmented images to display, generates a batch of augmented images, and displays the original and augmented images in a grid of subplots. The title for the entire figure is set, and the loop iterates over each row in the grid, displaying the original image in the first column and its corresponding augmented image in the second column. The axis labels for both subplots are turned off. The plot is then displayed, showing the original and augmented image pairs. This code segment provides a visual comparison between original and augmented images, illustrating the effects of data augmentation techniques applied during training. It helps in understanding how the augmentation process diversifies the training data to improve model generalization and performance.

4.3.Library Used

We used various Python libraries to perform breast cancer classification using Convolutional Neural Networks (CNNs). Some of the libraries we used in our experiment are pandas, numpy, tensorflow, seaborn, Scikit-learn with all features, matplotlib and more. The following code shows how to import these libraries.

```
# Import necessary libraries
import os
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score, roc_curve,
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.applications import ResNet50
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint, EarlyStopping
```

Figure 21: Libraries Used

4.4.Importing dataset

We used KERAS utility to load the data for the experiment. We used all the samples: 1662 for training, 357 for validation, and 357 for testing. The output layer size can be adjusted to match the number of classes the user needs. Figure 18 shows that the model has more than 9M parameters seen on figure 18. However, this simple, elegant, and easy-to-use model also has some challenges. The number of parameters is too high for embedded devices. The CNN

model is also very slow to train for large datasets. The dataset size is 6.81 GB, which requires a lot of disk space on google drive.

```
# Define paths for dataset
train_dir = '/content/drive/MyDrive/BC_Dataset_Experment/Korean_Hospital_MCM/Train'
val_dir = '/content/drive/MyDrive/BC_Dataset_Experment/Korean_Hospital_MCM/Val'
test_dir = '/content/drive/MyDrive/BC_Dataset_Experment/Korean_Hospital_MCM/Test'
```

Figure 22: Importing Dataset

4.5.Preprocessing

Originally the image size was 2304w x 3072h x 72 R x 8 bits and grayscale mammogram image. But we have adjusted the input 224, 224, 3 means the image with grayscale are inserted into the model with filter 32 and filter size 3*3with stride 2 and the image fed into convolutional layer with 200*200 output is max pooled and the 200 *200 are inserted into the second layer (200*200) are max pooled with second max pooling layer then 200*200*32 output fed into the last convolutional layer 200*200*64 output are max pooled and 50*50*64 are flattened to display the output of 6 class. Binary activation function used of full connected layer to select and display one of the classes.

```
# Data preprocessing and augmentation
train_datagen = ImageDataGenerator(
    rescale=1./255,
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest')

test_datagen = ImageDataGenerator(rescale=1./255)


# Load and augment training data
batch_size = 32
train_data = train_datagen.flow_from_directory(
    train_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary')

 Found 1662 images belonging to 2 classes.
```

```
# Load validation and testing data without augmentation
val_data = test_datagen.flow_from_directory(
    val_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=False)

test_data = test_datagen.flow_from_directory(
    test_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=False)

 Found 357 images belonging to 2 classes.
 Found 357 images belonging to 2 classes.
```

Figure 23: Data preprocessing and augmentation

This code segment outlines the process of preparing and augmenting image data for training a neural network model using the Keras `ImageDataGenerator` class. It includes the image data augmentation configuration, which involves defining an image data generator for training data with various augmentation parameters, such as rescale, rotation, width and height shift, shear transformation, zoom, horizontal flip, and fill mode. The image data generators for training, validation, and testing are defined, with the batch size specified. The training data is loaded and augmented from the specified directory, resizing images to the target size, dividing them into batches of size 32, and setting the class mode to 'binary' for binary classification tasks. The validation and testing data are loaded from the specified directories, resizing images, dividing them into batches, and setting the class mode to 'binary'. These generators do not apply augmentation and shuffle the data during batching. In summary, this code segment sets up data pipelines for training, validation, and testing neural network models, with data augmentation applied only to the training dataset, ensuring the model is trained on diverse input data variations while maintaining the validation and testing data for accurate evaluation.

4.6.Modeling

In machine learning, modeling refers to the process of creating a mathematical representation of a real-world system or phenomenon. The goal of modeling is to use data to create a predictive algorithm that can be used to make predictions or inform decisions about new data.

The modeling process consists of three main steps: selecting a suitable algorithm, training the algorithm using relevant data, and evaluating its performance using test data. A machine learning model is typically created by feeding large amounts of data into an algorithm, which then uses this data to learn patterns or relationships between variables. Once the algorithm has been trained, it can be used to make predictions or classifications about new data. Overall, modeling is an essential component of machine learning, as it enables algorithms to make accurate predictions and decisions based on data.

```
# Define the base model with EfficientNetB0
base_model = tf.keras.applications.EfficientNetB0(weights='imagenet', include_top=False, input

# Freeze the base model layers
for layer in base_model.layers:
    layer.trainable = False

# Create the classification model
model = tf.keras.Sequential([
    base_model,
    tf.keras.layers.GlobalAveragePooling2D(),
    tf.keras.layers.Dense(512, activation='relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(256, activation='relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])

# Fine-tune the model
for layer in base_model.layers[-20:]:
    layer.trainable = True

Downloading data from https://storage.googleapis.com/keras-applications/efficientnetb0_not
16705208/16705208 [==============================] - 0s 0us/step
```

Figure 24: Image classification model using EfficientNetB0

This code snippet outlines a deep neural network architecture for fine-tuning a pre-trained ResNet50 model on a classification task. It consists of several parts, including the input shape specification, base model definition, freezing base model layers, and classification model construction. The input shape specifies the shape of the input images, which are expected to have a height and width of 224 pixels and three color channels (red, green, and blue). The base model is initialized on the ImageNet dataset, excluding the top classification layer, and serves as a feature extractor. The model freezes all layers to prevent weight updates during training, allowing the model to leverage knowledge learned from ImageNet without overfitting to the new task. The classification model is constructed by stacking layers

sequentially, with the base model serving as the first layer. A global average pooling layer is added to reduce the spatial dimensions of feature maps, while a dense layer with 512 units and ReLU activation function is added for feature transformation. A dropout layer is added to prevent overfitting by randomly setting a fraction of input units to zero during training. The output layer is added with a single unit and sigmoid activation function, suitable for binary classification tasks. In summary, this architecture combines the feature extraction capabilities of the pre-trained ResNet50 model with additional dense layers for fine-tuning on specific classification tasks, aiming to achieve higher accuracy and better generalization.

model.summary() prints a summary of the model architecture, showing the number of parameters and output shapes for each layer.

```
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 efficientnetb0 (Functional  (None, 7, 7, 1280)        4049571
 )

 global_average_pooling2d (  (None, 1280)              0
 GlobalAveragePooling2D)

 dense (Dense)               (None, 512)               655872

 dropout (Dropout)           (None, 512)               0

 dense_1 (Dense)             (None, 256)               131328

 dropout_1 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 128)               32896

 dense_3 (Dense)             (None, 1)                 129

=================================================================
Total params: 4869796 (18.58 MB)
Trainable params: 3347089 (12.77 MB)
Non-trainable params: 1522707 (5.81 MB)
_________________________________________________________________
```

Figure 25: Summary of EfficientNetB0 model

The model.summary() method is used to print a summary of the neural network model architecture .

The summary includes the following information:

- The layer type and name.
- The output shape of each layer.
- The number of parameters in each layer.

In this code, the model consists of five layers:

- An EfficientNetB0 layer with an output shape of (None, 7, 7, 1280) and 4,049,564 parameters.
- A Flatten layer with an output shape of (None, 62,720) and no parameters.
- A Dense layer with an output shape of (None, 256) and 16,056,576 parameters.
- A Dropout layer with an output shape of (None, 256) and no parameters.
- A final Dense layer with an output shape of (None, 1) and 257 parameters.

The code defines a convolutional neural network (CNN) model using the Keras library. A CNN is a type of deep learning model that can learn to extract features from images and perform tasks such as image classification, object detection, face recognition, etc. The Keras library is a high-level API that allows you to easily build and train deep learning models in Python. The function create_model() returns a model object that contains the following layers:

The model can be used for binary image classification tasks, such as distinguishing between benign and malignant, or between normal and abnormal. To train the model, you need to provide labeled images as input and output pairs, and use an optimizer and a loss function to update the model parameters based on the prediction errors.

4.7.Compiling the model

Compiling in python is the process of converting the source code into a code object that can be executed by the python interpreter. The compile() function is a built-in function that can perform this process for a given source code, filename, and mode. The source code can be a string, a bytes object, or an abstract syntax tree (AST) object. The filename is the name of the file that contains the source code, or any string if the source code does not come from a file.

```
# Compile the model
model.compile(optimizer=Adam(learning_rate=0.0001),
              loss='binary_crossentropy',
              metrics=['accuracy'])
```

Figure 26: Code for compile

The above code is configuring the model for training using the compile method. The arguments are:

The above code is written in Python and uses the Keras library to compile a machine learning model. The compile () method is used to configure the learning process of the model.

- The optimizer parameter specifies the optimization algorithm used to update the weights of the model during training. In this case, the Adam optimizer is used with a learning rate of 0.0001.
- The loss parameter specifies the loss function used to measure how well the model fits the training data. In this case, binary cross-entropy is used as the loss function.
- The metrics parameter specifies a list of metrics to be evaluated by the model during training and testing. In this case, three metrics are specified: accuracy, precision, and recall.
- Accuracy measures how often the model makes correct predictions, precision measures how many of the predicted positive cases are actually positive, and recall measures how many of the actual positive cases are correctly predicted by the model .
- By using these metrics, we can evaluate how well our model is performing and make adjustments as necessary.

4.8.Training the model

Training the model in python is the process of adjusting the model parameters to minimize the loss function and maximize the accuracy on a given dataset which are splited into training, validating and testing sets with a ratio of train 70%, test 15%, and validate 15%, where the training set is used to fit the model and the testing set is used to evaluate the model performance.

```
# Train the model
history = model.fit(train_data,
                    epochs=40,
                    validation_data=val_data,
                    callbacks=[EarlyStopping(patience=3)])
```

Figure 27: Train the Model

We use the Keras library to train a machine learning model. The fit() method is used to train the model on the training data.The train_data parameter specifies the training data for the model. The val_data parameter specifies the validation data for the model. The epochs parameter specifies the number of times the model will be trained on the entire dataset. In this case, the model will be trained for given epochs . The callbacks parameter specifies a list of functions that will be called during training. In this case, an instance of EarlyStopping is

passed as a callback function with a patience of 3. After training, the history of the training process is stored in the history variable.

```
Epoch 1/40
52/52 [==============================] - 667s 13s/step - loss: 0.5538 - accuracy: 0.7611 -
Epoch 2/40
52/52 [==============================] - 29s 558ms/step - loss: 0.5349 - accuracy: 0.7852
Epoch 3/40
52/52 [==============================] - 28s 542ms/step - loss: 0.5334 - accuracy: 0.7852
Epoch 4/40
52/52 [==============================] - 28s 544ms/step - loss: 0.5274 - accuracy: 0.7852
Epoch 5/40
52/52 [==============================] - 29s 550ms/step - loss: 0.5303 - accuracy: 0.7852
Epoch 6/40
52/52 [==============================] - 28s 547ms/step - loss: 0.5327 - accuracy: 0.7852
Epoch 7/40
52/52 [==============================] - 29s 562ms/step - loss: 0.5288 - accuracy: 0.7852
Epoch 8/40
52/52 [==============================] - 28s 542ms/step - loss: 0.5273 - accuracy: 0.7852
Epoch 9/40
52/52 [==============================] - 28s 545ms/step - loss: 0.5332 - accuracy: 0.7852
```

Figure 28: Train the model (train 70%, test 15%, and validate 15%) with EfficentNetB0

4.9.Model Evaluation

Model evaluation involves analyzing the performance and efficacy of a machine learning model on a specific dataset. It is a crucial step in any machine learning project as it helps in selecting the most appropriate model from a pool of options, fine-tuning the model's hyperparameters, and assessing its capability to generalize to unseen data. In Python, there are diverse methods for evaluating a model, depending on factors like the type of problem (classification, regression, clustering), the characteristics of the data (balanced, imbalanced, multiclass), and the evaluation approach (single or multiple metrics, train-test split or cross-validation). The model evaluation process utilized in our experiment is illustrated in Figure 25.

```
# Calculate precision, recall, F1-score, sensitivity, specificity, accuracy
precision = precision_score(test_data.classes, y_pred_labels)
recall = recall_score(test_data.classes, y_pred_labels)
f1score = f1_score(test_data.classes, y_pred_labels)
sensitivity = recall
specificity = recall_score(test_data.classes, y_pred_labels, pos_label=0)
accuracy = accuracy_score(test_data.classes, y_pred_labels)

print("Precision    :", precision)
print("Recall       :", recall)
print("F1-Score     :", f1score)
print("Sensitivity  :", sensitivity)
print("Specificity  :", specificity)
print("Accuracy     :", accuracy)
```

```
Precision    : 0.0
Recall       : 0.0
F1-Score     : 0.0
Sensitivity  : 0.0
Specificity  : 1.0
Accuracy     : 0.7804878048780488
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-defined and
  _warn_prf(average, modifier, msg_start, len(result))
```

```
# Evaluate the model
loss, accuracy = model.evaluate(test_data)
print("Test Loss            :", loss)
print("Test Accuracy        :", accuracy)
```

Figure 29: Evaluating the performance of a model

This code snippet is utilizing the ImageDataGenerator class from Keras to perform data augmentation on image datasets. Here's a breakdown of each part:

- ImageDataGenerator Configuration: The ImageDataGenerator is initialized with various augmentation parameters such as rotation range, width shift range, height shift range, shear range, zoom range, horizontal flip, and fill mode. These parameters control how the input images are transformed during training.
- Load and Augment Training Data: The flow_from_directory method is used to load training data from the specified directory (train_dir). The target_size parameter is set to resize the images to the specified input shape, and batch_size determines the number of images in each batch. The class_mode is set to 'binary' since it's a binary classification problem. During training, the data generator applies the specified augmentation techniques to the images in real-time.
- Load Validation and Testing Data: Similarly, validation and testing data are loaded using the flow_from_directory method from their respective directories (val_dir and test_dir). However, for validation and testing data, augmentation is not applied (shuffle=False), ensuring consistency in evaluation. The target_size, batch_size, and class_mode parameters are set similarly to the training data.

In summary, this code sets up a data pipeline for loading and augmenting image datasets for training a deep learning model. The ImageDataGenerator is used to perform various

transformations on the input images during training, helping to increase the diversity and robustness of the training data. The augmented data is then fed into the model for training, while the validation and testing data remain unchanged for accurate evaluation.

4.9.1. Generate Confusion Matrix

A confusion matrix is a table that is commonly used to evaluate the performance of a classification model, especially in image classification tasks using EfficentNetBo in Python. In image classification, a confusion matrix provides a detailed summary of how well the model is performing by showing the counts of true positive, true negative, false positive, and false negative predictions. The rows represent the actual classes of the images, while the columns represent the predicted classes.

To create a confusion matrix based on image classification using EfficentNetBo in Python, we can follow these steps:

1. Load trained EfficentNetBo and the test dataset.
2. Iterate through the test dataset and predict the class of each image using the EfficentNetBo model.
3. Capture the true labels of the images.
4. Create an empty confusion matrix with rows and columns representing the different classes.
5. For each predicted class and the true class, increment the respective cell in the confusion matrix.
6. Display the confusion matrix to evaluate the performance of the model.

```
# Create confusion matrix
cm = pd.DataFrame(confusion_mat, index=class_labels.values(), columns=class_labels.values())

plt.figure(figsize=(8, 6))
plt.title("Confusion Matrix")
sns.heatmap(cm, annot=True, cmap="Blues", fmt="d")
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.show()
```

This code segment is used to visualize the confusion matrix, a crucial tool for evaluating the performance of a classification model. Here's what each part does:

- Creation of DataFrame (cm): The confusion_mat variable holds the confusion matrix values, typically computed during the evaluation of a classification model. This matrix

is represented as a 2D array, where rows represent the true labels and columns represent the predicted labels. The pd.DataFrame() function from the pandas library is used to create a DataFrame (cm) from the confusion matrix array. The index and columns parameters are set to class_labels.values(), which are the class labels (e.g., "Benign" and "Malignant"). This ensures that the rows and columns of the DataFrame are labeled with the corresponding class names.

- Visualization with Seaborn Heatmap: After creating the DataFrame, the code proceeds to visualize the confusion matrix using a heatmap. First, a Matplotlib figure is created with a specific size using plt.figure(figsize=(8, 6)). The title of the heatmap is set to "Confusion Matrix" using plt.title(). Then, the sns.heatmap() function from the Seaborn library is called to create the heatmap. The cm DataFrame is passed as input, and annot=True adds numerical annotations to each cell of the heatmap, displaying the actual counts of true positive, false positive, true negative, and false negative predictions. The cmap="Blues" parameter specifies the color map used for the heatmap (in this case, shades of blue), and fmt="d" formats the annotations as integers. Finally, the x-axis label ("Predicted Label") and y-axis label ("True Label") are set using plt.xlabel() and plt.ylabel() functions, respectively.
- Display: The heatmap visualization is displayed using plt.show().

Overall, this code segment provides a clear and informative visualization of the confusion matrix, making it easier to interpret the performance of a classification model in terms of true positives, false positives, true negatives, and false negatives for each class. Finally, the plt.show() function is called to display the plot.

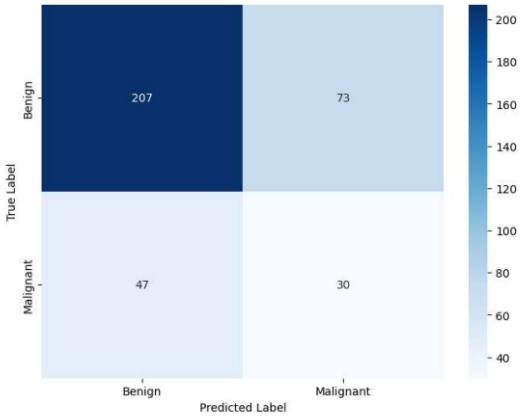


Figure 30: Confusion matrices for the experimentation of the EfficentNetB0 architecture

4.9.2. Predicting the results

In image classification, the prediction process entails utilizing a pre-trained model to extract features from an input image, utilizing these features to forecast the image's class or label, and subsequently assessing the predictive performance. Various libraries and frameworks in Python, including Tensorflow, Keras, and Scikit-learn, streamline the implementation of image classification models and the prediction process.

```
# Display predicted images with labels
y_pred = model.predict(test_data)
y_pred_labels = np.round(y_pred)
class_labels = {0: 'Benign', 1: 'Malignant'}
sample_images, true_labels = next(test_data)
plt.figure(figsize=(10, 10))
for i in range(9):
    plt.subplot(3, 3, i+1)
    plt.imshow(sample_images[i])
    plt.title(f"True: {class_labels[int(true_labels[i])]}, Predicted: {class_labels[int(y_pred
    plt.axis('off')
plt.show()
```

Figure 31: Predicting results as Benign or Malignant

This code snippet is used to make predictions on a test dataset using a trained model and visualize the results alongside the true labels of the samples. Let's break down each part:

- Model Prediction:

  - `y_pred = model.predict(test_data)`: This line generates predictions for the test dataset (`test_data`) using the trained model (`model`). It returns the predicted probabilities for each sample in the test dataset.

- Round Predictions:

  - `y_pred_labels = np.round(y_pred)`: The predicted probabilities obtained from the model are rounded to either 0 or 1, representing the predicted class labels. This rounding operation is necessary since the model predicts probabilities, and for binary classification, we need discrete class labels.

- Class Labels Definition:

  - `class_labels = {0: 'Benign', 1: 'Malignant'}`: This dictionary defines the mapping between class indices (0 and 1) and their corresponding class labels ('Benign' and 'Malignant').

- Visualization:
  - `sample_images, true_labels = next(test_data)`: Retrieves a batch of sample images and their true labels from the test dataset generator (`test_data`).

- `plt.figure(figsize=(10, 10))`: Creates a new figure with a specified size for plotting the sample images and their corresponding predictions.
- The subsequent loop iterates over the first nine samples in the batch:
- `plt.subplot(3, 3, i+1)`: Specifies the subplot position for each sample image in a 3x3 grid.
- `plt.imshow(sample_images[i])`: Displays the current sample image.
- `plt.title(...)`: Sets the title of the subplot, indicating the true label and the predicted label for the current sample.
- `plt.axis('off')`: Turns off the axis labels for better visualization.
- `plt.show()`: Displays the figure containing the sample images with their true and predicted labels.

Overall, this code segment enables the visual inspection of model predictions on a subset of the test dataset, allowing users to assess the model's performance and gain insights into its behavior.

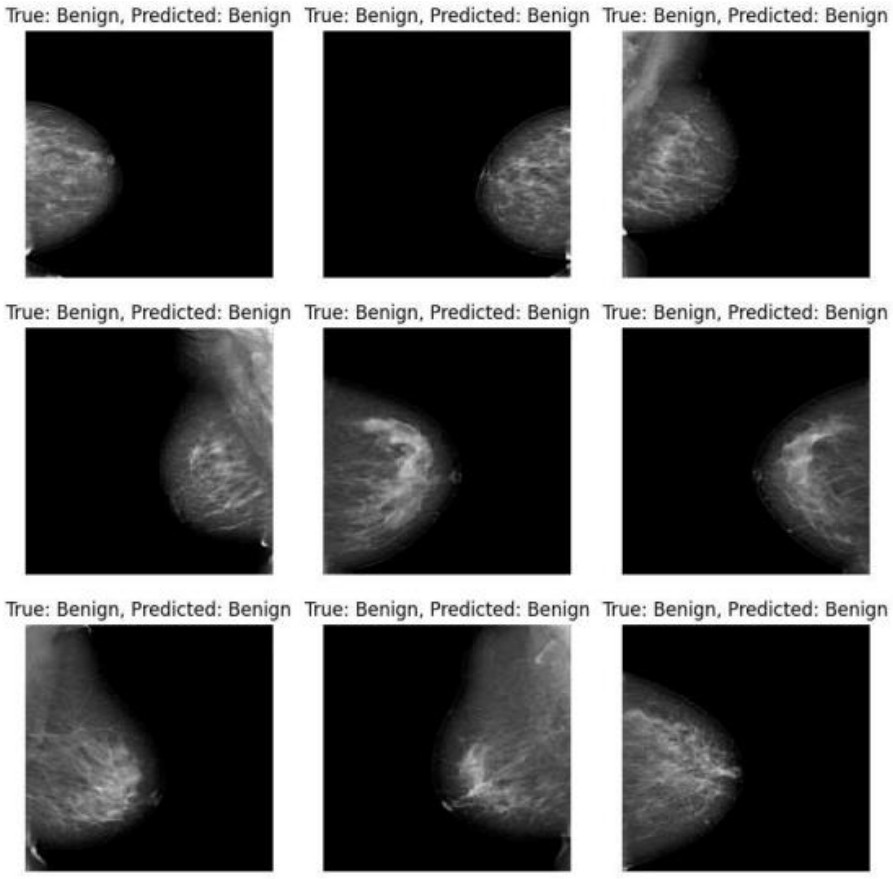
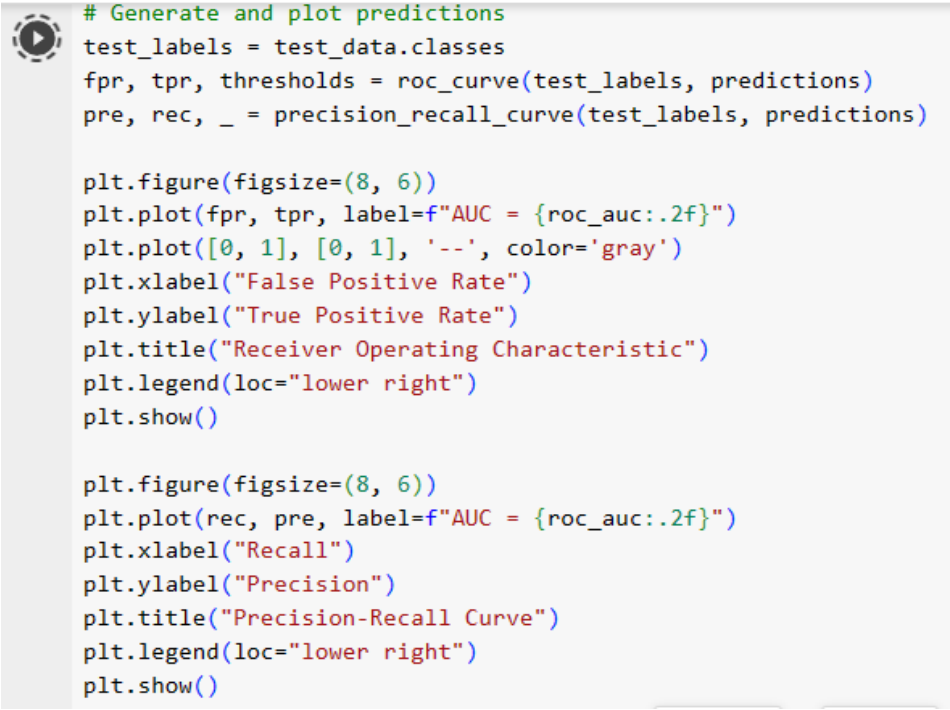


Figure 32: Results with prediction

4.9.3. Receiver Operating Characteristic (ROC)

The ROC curve is a graphical representation of the performance of a binary classification model as the threshold for classification is varied. It provides valuable insights into the model's ability to correctly classify positive instances and its susceptibility to false positives. The AUC summarizes the overall performance of the model, and the curve can be used to select an appropriate threshold based on specific requirements.



```
# Generate and plot predictions
test_labels = test_data.classes
fpr, tpr, thresholds = roc_curve(test_labels, predictions)
pre, rec, _ = precision_recall_curve(test_labels, predictions)

plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, label=f"AUC = {roc_auc:.2f}")
plt.plot([0, 1], [0, 1], '--', color='gray')
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("Receiver Operating Characteristic")
plt.legend(loc="lower right")
plt.show()

plt.figure(figsize=(8, 6))
plt.plot(rec, pre, label=f"AUC = {roc_auc:.2f}")
plt.xlabel("Recall")
plt.ylabel("Precision")
plt.title("Precision-Recall Curve")
plt.legend(loc="lower right")
plt.show()
```

Figure 33: Receiver Operating Characteristic (ROC)

The code is used to plot the Receiver Operating Characteristic (ROC) curve and Precision-Recall (PR) curve for a binary classification model. The ROC curve shows the relationship between the true positive rate (TPR) and false positive rate (FPR) at different classification thresholds. The PR curve shows the relationship between precision and recall at different classification thresholds. The code first computes the FPR, TPR, and thresholds for the ROC curve using roc_curve function from sklearn.metrics. Similarly, it computes precision, recall, and thresholds for the PR curve using precision_recall_curve function from sklearn.metrics.

Then, it plots the ROC curve and PR curve using matplotlib.pyplot. The ROC curve is plotted with FPR on the x-axis and TPR on the y-axis. The PR curve is plotted with recall on the x-axis and precision on the y-axis. The area under the ROC curve (AUC-ROC) and area under the PR curve (AUC-PR) are also calculated and displayed in the legend of each plot.

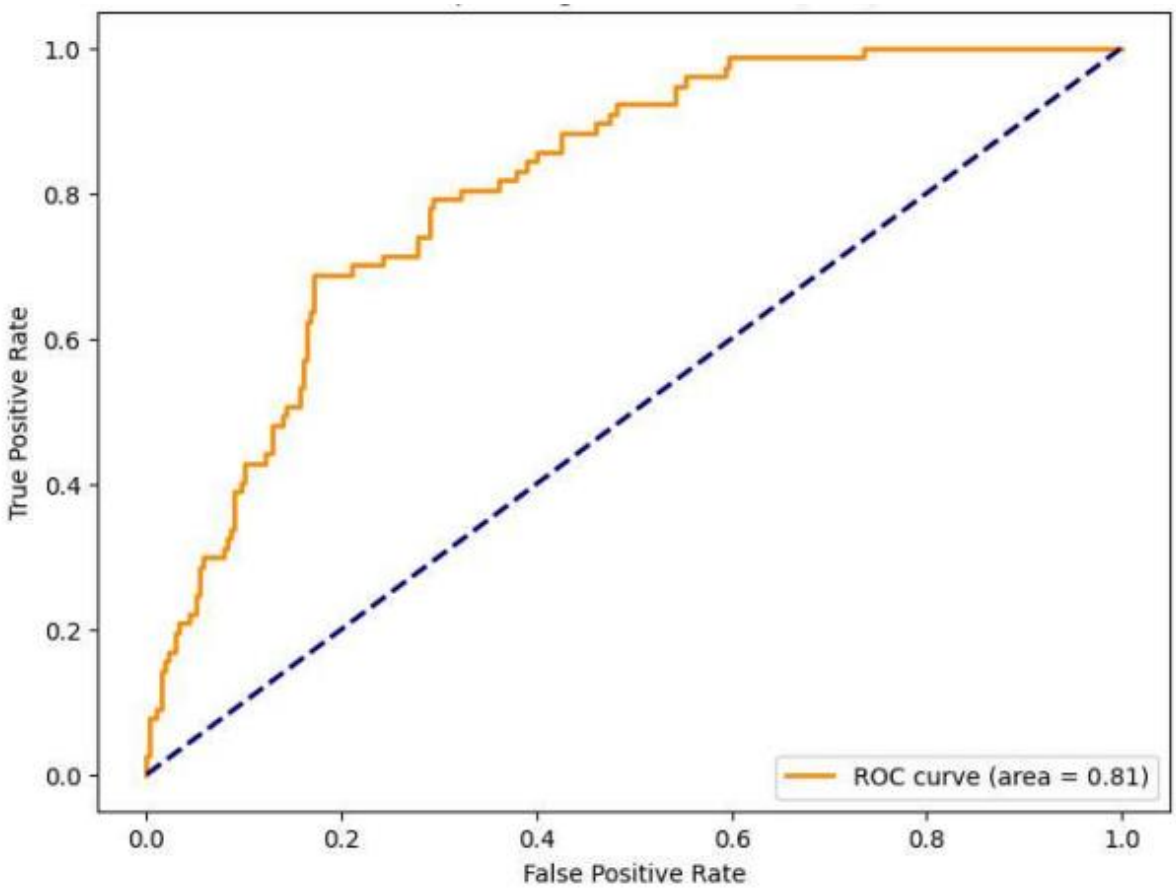


Figure 34: Receiver Operating Characteristic (ROC) Curve

4.9.4. The loss and accuracy curves

The loss and accuracy curves are plots that show the performance of a machine learning model over time during the training phase. The loss curve represents the model's error or how well it is able to make predictions. It is usually calculated as the difference between the predicted value and the actual value. The lower the loss, the better the model's performance. The loss curve typically starts high and gradually decreases as the model learns and adjusts its parameters to make better predictions. The accuracy curve, on the other hand, represents how well the model is able to classify or predict correctly. It is calculated as the ratio of the number of correct predictions to the total number of predictions made by the model. The higher the accuracy, the better the model's performance. The accuracy curve usually starts low and increases as the model learns and improves its prediction abilities. The result of the code is showed on figure 28.

```
# Plot the loss and accuracy curves
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.title('Loss Curves')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(history.history['accuracy'], label='Training Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.title('Accuracy Curves')
plt.legend()

plt.tight_layout()
plt.show()
```

Figure 35: The loss and accuracy curves

The code is used to plot the loss and accuracy curves of a model. The history object is used as input to the plot function. The history object is a record of training loss values and metrics values at successive epochs, as well as validation loss values and validation metrics values (if applicable). The code creates a figure with two subplots, one for the loss curves and one for the accuracy curves. The first subplot shows the training and validation loss curves, while the second subplot shows the training and validation accuracy curves. The x-axis of both subplots represents the number of epochs, while the y-axis represents the loss or accuracy value. The plt.subplot function is used to create two subplots in a single figure. The first argument specifies the number of rows in the grid, the second argument specifies the number of columns in the grid, and the third argument specifies which subplot to activate. The plt.plot function is used to plot the loss or accuracy values against the number of epochs. The label parameter specifies the label for each curve, which is used in the legend. The plt.xlabel, plt.ylabel, and plt.title functions are used to set the labels for the x-axis, y-axis, and title of each subplot, respectively. Finally, plt.tight_layout() is called to adjust the spacing between subplots so that they do not overlap. And then plt.show() is called to display the figure.

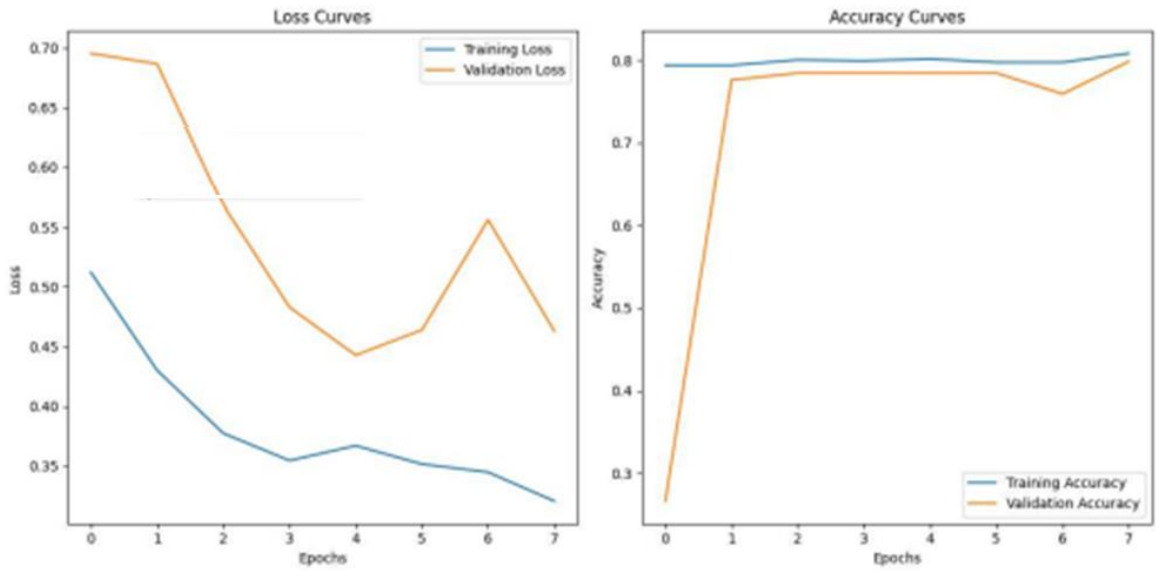


Figure 36: The loss and accuracy curves

4.9.5. Classification report

The classification report in Python is a tool that helps in evaluating the performance of a classification model. A classification report provides various metrics like precision, recall, F1-score, and support for each class in the dataset. These metrics are computed for each class individually and also as an average across all classes.

```
# Generate evaluation metrics
classification_rep = classification_report(test_data.classes, predicted_labels, target_names=class_labels.values())
confusion_mat = confusion_matrix(test_data.classes, predicted_labels)
roc_auc = roc_auc_score(test_data.classes, predictions)

print("Classification Report:")
print(classification_rep)
print("ROC AUC Score:", roc_auc)
```

Figure 37: Generate evaluation metrics

This code snippet is used to evaluate the performance of a classification model. Here's a breakdown of each part:

`classification_rep = classification_report(test_data.classes, predicted_labels, target_names=class_labels.values())`: This line generates a classification report using the `classification_report` function from the scikit-learn library. It takes three main arguments:

`test_data.classes`: The true class labels of the test data.

`predicted_labels`: The predicted class labels generated by the model.

`target_names=class_labels.values()`: An optional argument specifying the names of the target classes. It helps in labeling the rows of the classification report.

`confusion_mat = confusion_matrix(test_data.classes, predicted_labels)`: This line computes the confusion matrix using the `confusion_matrix` function from scikit-learn. It takes two main arguments:

`test_data.classes`: The true class labels of the test data.

`predicted_labels`: The predicted class labels generated by the model.

`roc_auc = roc_auc_score(test_data.classes, predictions)`: This line calculates the ROC AUC (Receiver Operating Characteristic Area Under the Curve) score using the `roc_auc_score` function from scikit-learn. It takes two main arguments:

`test_data.classes`: The true class labels of the test data.

`predictions`: The probability estimates or predicted scores of the positive class generated by the model.

`print("Classification Report:")`: This line prints a header indicating that the classification report will be displayed.

`print(classification_rep)`: This line prints the classification report generated in the first step. The classification report provides metrics such as precision, recall, F1-score, and support for each class.

`print("ROC AUC Score:", roc_auc)`: This line prints the ROC AUC score calculated in the third step. The ROC AUC score quantifies the model's ability to distinguish between classes, with a score closer to 1 indicating better performance.

Overall, this code snippet is used to assess the classification model's performance by providing a detailed report of various metrics such as precision, recall, and F1-score, along with the confusion matrix and the ROC AUC score.. Finally, the results are printed using the print function.

Classification Report:

	precision	recall	f1-score	support
0	0.78	1.00	0.88	280
1	0.00	0.00	0.00	77
accuracy			0.78	357
macro avg	0.39	0.50	0.44	357
weighted avg	0.62	0.78	0.69	357

Figure 38: Classification report

CHAPTER FIVE

5. RESULTS AND DISCUSSIONS

5.1. RESULTS

Performance evaluation metrics, such as precision, recall, and F1 score, were computed separately for each class, with the average scores across both classes used to evaluate the performance of each experiment. The training and testing processes for both the EfficientNetB0 and ResNet transfer learning architectures on the local dataset typically took place. Additionally, real-time data augmentation and preprocessing of mammography batches were executed using Python, leveraging the OpenCV package from Keras. Optimal hyperparameters, including batch size, learning rate, and optimization function, were meticulously selected throughout the experimentation phase.

5.2. Hyper parameters

An epoch is a hyper parameter indicating the number of times the learning algorithm will work through the entire training dataset, allowing each sample to update its internal model parameters. The parameters used in the study are as follows.

Table 7 Parameters used in the study

Parameters	**Values**
Batch Size	32
Learning Rate	0.0001
Epochs	1,5,10,20-40 ..100
Dropout	0.5
Activation function	Softmax
Optimization function	Adam
Training Split	0.7
Testing Split	0.15
Validating Split	0.15
Transfer Learning Architectures	EfficentNetB0 and ResNet50

5.1. Discussion

The research aimed to develop an experimental test design and a model specifically designed for breast cancer classification to address the research questions.

RQ1: Can locally organized datasets be used to distinguish between benign and malignant breast cancer?

Deep learning techniques have shown promise in distinguishing between benign and malignant breast cancer using locally curated datasets. These techniques use intricate neural network architectures to extract discriminative features from input images, enabling precise case classification. The inclusion of locally curated datasets enhances model performance by providing varied and representative samples reflecting the target population's diversity. Research in this domain holds potential to improve breast cancer screening and treatment efforts. The EfficientNetB0 transfer method is selected, which demonstrated accuracy in differentiating between benign and malignant tumors. With a learning rate of 0.0001 and 40 epochs, the study achieved an accuracy of 78.43% and a ROC AUC Score of 0.81 using local image dataset in classifying benign and malignant tumors. The findings demonstrate the effectiveness of locally maintained images collections in accurate breast cancer classification.

RQ2: What is the difference in breast cancer classification accuracy between local and international image datasets?

Due to differences in data collecting techniques, imaging technology, and annotation processes, the accuracy of the classification of breast cancer might differ between local and international image datasets. International datasets include a greater variety of instances, but local datasets may provide more representative images unique to the population. Differences between local and international settings might also be attributed to disparities in resources allotted for model development and validation. Variations exist in the accuracy of breast cancer classification according to variables such patient demographics, data quality, and sample size. Because they include a larger range of patient demographics and illness manifestations, international databases could provide superior accuracy. However, the precision of classification models may be impacted by variations in data gathering techniques and equipment.

In this study, two transfer learning architectures, ResNet50 and EfficientNetB0, were employed to assess the accuracy of breast cancer classification using public, local, and combined mammography image datasets. The datasets were split into 70% for training, 15% for validation, and 15% for testing, resulting in a total of 2376 mammogram images categorized as benign or malignant.

We achieved an accuracy of 78.43% with ROC curves of 0.81, 0.89, and 0.46 for epochs 40 using EfficientNetB0 on local, public, and merged image datasets, respectively. Additionally, utilizing ResNet50 on the same datasets, we obtained an accuracy of 78.43% with ROC curves of 0.69, 0.97, and 0.81 for epoch 40, respectively. According to our experiment using local dataset yields best accuracy in terms of ROC curves.

RQ3: Which deep learning technique performs best in classifying breast cancer cases?

EfficientNetB0 and ResNet50 are popular convolutional neural network (CNN) architectures used for image classification tasks, including breast cancer classification. EfficientNetB0 uses compound scaling to improve efficiency and accuracy, while ResNet50 uses skip connections to bypass layers. EfficientNetB0 achieves higher accuracy with fewer parameters, while ResNet50 has a larger number of parameters, potentially leading to overfitting. Both models perform well on local datasets, with EfficientNetB0 requiring fewer resources and shorter training times. However, selecting the optimal deep learning approach requires rigorous experimentation and validation. Based on our experiment we select EfficentNetB0.

RQ4: What challenges come with classifying breast cancer using locally produced images?

Challenges including data quality, privacy, and ethical issues arise when classifying breast cancer using images supplied locally. While annotating takes time and expertise, imbalances in data distribution might induce biases. Deep learning algorithm performance might be restricted by limited computer resources. Strong validation processes may be necessary due to the complexity of classification jobs caused by local datasets. In order to guarantee data accuracy and representativeness, collaboration between academics, healthcare professionals, and policymakers is necessary. Insufficient diversity in datasets and insufficient number of radiologists with training might potentially impede precise diagnosis in Ethiopia.

CHAPTER SIX

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

Breast cancer represents a significant global health challenge, with escalating incidence rates posing a serious threat. Often, delayed diagnosis due to lack of awareness contributes to heightened mortality rates associated with the disease. In this context, computer-aided diagnosis emerges as a critical tool, offering precise diagnostic capabilities accessible to individuals from diverse backgrounds. While it cannot replace the expertise of medical practitioners, a CAD system substantially supports healthcare professionals in making informed decisions by analyzing patient data. This helps mitigate diagnostic errors stemming from limited expertise or inadequate report analysis, thereby enhancing modern medical practice. However, the efficacy of such systems hinges on the clarity of the training model. Despite exhibiting improved performance compared to its predecessors, the algorithm's training time presents a limitation, particularly in deeply trained neural networks. Nevertheless, in systems equipped with GPUs, this process is expedited relative to conventional hardware setups, resulting in a computationally superior testing and processing environment for users.

The study effectively tackled all research inquiries outlined in the Results and Discussion section. Ultimately, by harnessing local image datasets and leveraging EfficientNetB0 for breast cancer classification, the study yielded promising outcomes in the precise detection and diagnosis of breast cancer. Furthermore, the adoption of EfficientNetB0 holds potential in assisting medical practitioners in making accurate early-stage diagnoses, thereby potentially bolstering the survival rates of breast cancer patients.

6.2. Recommendations

Breast cancer classification is a critical endeavor, and the advancement of computer-aided techniques for precise detection of malignant tumors holds immense potential to enhance diagnostic and treatment outcomes. However, there exists a notable dearth of diverse and inclusive datasets, particularly in relation to underrepresented demographics like Black women. Here are several potential avenues for future research in breast cancer classification utilizing EfficientNetB0 with locally sourced image datasets:

6.3.Future Work

Transfer learning techniques expedite the training process of EfficientNetB0 models by leveraging pre-trained models and fine-tuning them on smaller datasets. Exploring the fusion of multi-modal data, which integrates features from diverse imaging modalities such as mammography and MRI, holds promise for enhancing classification accuracy. It is crucial to assess the effectiveness of various data augmentation techniques in EfficientNetB0 model training, as not all methods are universally applicable. Additionally, the integration of histopathology images and genomic data has the potential to improve the accuracy of breast cancer detection. Investigating the suitability of EfficientNetB0 models for classification tasks that incorporate these additional data sources, while ensuring balanced datasets that adequately represent both White and Black women, is essential. Understanding the impact of these factors on model accuracy is critical for advancing the precision of breast cancer detection.

REFERENCE

Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: Advances, challenges and applications. *BMC Bioinformatics*, *20*(Suppl 11). https://doi.org/10.1186/s12859-019-2823-4

Abebe, M. S. (2019). *S-transform Based Detection and Classification of Breast Cancer in Ultrasound Images*.

Al-Dhabyani, W., Fahmy, A., Gomaa, M., & Khaled, H. (2019). Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *International Journal of Advanced Computer Science and Applications*, *10*(5), 618–627. https://doi.org/10.14569/ijacsa.2019.0100579

Al Noumah, W., Jafar, A., & Al Joumaa, K. (2022). Using parallel pre-trained types of DCNN model to predict breast cancer with color normalization. *BMC Research Notes*, *15*(1), 1–6. https://doi.org/10.1186/s13104-021-05902-3

Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine*, *223*, 106951. https://doi.org/10.1016/J.CMPB.2022.106951

Aslam, M. A., & Cui, D. (2020). Breast Cancer Classification using Deep Convolutional Neural Network. *Journal of Physics: Conference Series*, *1584*(1). https://doi.org/10.1088/1742-6596/1584/1/012005

Ayalew, K., Ayalew, Y., Alemu, T., & Gebru, T. (2021). Knowledge and Experience of Women with Breast Cancer Receiving Chemotherapy in Selected Public Hospitals, Addis Ababa, Ethiopia. *Journal of Cancer Treatment and Research*, *9*(3), 56. https://doi.org/10.11648/j.jctr.20210903.13

Ayele, W., Führer, A., Braun, G. A., Formazin, F., Wienke, A., Taylor, L., Unverzagt, S., Addissie, A., & Kantelhardt, E. J. (2022). Breast cancer morbidity and mortality in rural Ethiopia: data from 788 verbal autopsies. *BMC Women's Health*, *22*(1). https://doi.org/10.1186/s12905-022-01672-7

Bimrew Sendekie Belay. (2022). No Titleהכי קשה לראות את מה שבאמת לנגד העינים. *הארץ*, *8.5.2017*, 2003–2005.

Debelee, T. G., Amirian, M., Ibenthal, A., Palm, G., & Schwenker, F. (2018). Classification

of mammograms using convolutional neural network based feature extraction. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, *244*, 89–98. https://doi.org/10.1007/978-3-319-95153-9_9

Debelee, T. G., Schwenker, F., Ibenthal, A., & Yohannes, D. (2020). Survey of deep learning in breast cancer image analysis. *Evolving Systems*, *11*(1), 143–163. https://doi.org/10.1007/s12530-019-09297-2

Duche, H., Tsegay, A. T., & Tamirat, K. S. (2021). Identifying risk factors of breast cancer among women attending selected hospitals of addis ababa city: Hospital-based unmatched case-control study. *Breast Cancer: Targets and Therapy*, *13*, 189–197. https://doi.org/10.2147/BCTT.S293867

Falsi-based, E. C. R., Alrashidi, H., Tariq, U., & Cha, J. (2023). *BC 2 NetRF : Breast Cancer Classification from Mammogram Images Using Enhanced Deep Learning Features and Features Selection*.

Gezahegn, Y. G. (2020). *B c d u c - n n. 027*(2018), 1–8.

Guan, S., & Loew, M. (2019). *Using generative adversarial networks and transfer learning for breast cancer detection by convolutional neural networks*. 48. https://doi.org/10.1117/12.2512671

Hoxha, I., Islami, D. A., Uwizeye, G., Forbes, V., & Chamberlin, M. D. (2022). Forty-Five Years of Research and Progress in Breast Cancer: Progress for Some, Disparities for Most. *JCO Global Oncology*, *8*, 1–5. https://doi.org/10.1200/go.21.00424

Ilhan, U., Uyar, K., & Iseri, E. I. (2021). Breast Cancer Classification Using Deep Learning. *Advances in Intelligent Systems and Computing*, *1306*, 709–714. https://doi.org/10.1007/978-3-030-64058-3_88

Karthik, S., Srinivasa Perumal, R., & Chandra Mouli, P. V. S. S. R. (2018). Breast cancer classification using deep neural networks. *Knowledge Computing and Its Applications: Knowledge Manipulation and Processing Techniques: Volume 1*, 227–241. https://doi.org/10.1007/978-981-10-6680-1_12

Khan, M. H. M., Boodoo-Jahangeer, N., Dullull, W., Nathire, S., Gao, X., Sinha, G. R., & Nagwanshi, K. K. (2021). Multi- class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). *PLoS ONE*, *16*(8 August 2021), 1–15.

https://doi.org/10.1371/journal.pone.0256500

Luo, L., Wang, X., Lin, Y., Ma, X., Tan, A., Chan, R., Vardhanabhuti, V., & Cw, W. (2023). *Deep Learning in Breast Cancer Imaging : A Decade of Progress and Future Directions*.

Mahmood, T., Li, J., Id, Y. P., Akhtar, F., & Rehman, M. U. (2022). *Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach*. 1–25. https://doi.org/10.1371/journal.pone.0263126

Nazir, M. S., Khan, U. G., Mohiyuddin, A., Saleh, M., Reshan, A., Shaikh, A., Rizwan, M., & Davidekova, M. (2022). *Research Article A Novel CNN-Inception-V4-Based Hybrid Approach for Classification of Breast Cancer in Mammogram Images*. *2022*.

Prodan, M., & Paraschiv, E. (2023). *applied sciences Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection*.

Qasim, K. R., & Ouda, A. J. (2020). An accurate breast cancer detection system based on deep learning cnn. *Medico-Legal Update*, *20*(1), 984–990. https://doi.org/10.3750/v20/i1/2020/mlu/194726

Rahel, D., & Desta, R. (2020). *[138] PREPROCESSING OF MAMMOGRAPHY IMAGE FOR 2020.pdf*.

Rakhshan, P. (2023). *BREAST CANCER DETECTION BASED ON CNN AND FEDERATED LEARNING USING EMBEDDED DEVICES*.

Ramadhani, S., & First, S. R. (2020). A Review Comparative Mammography Image Analysis on Modified CNN Deep Learning Method. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, *3*(1), 1–10.

Shallu, & Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express*, *4*(4), 247–254. https://doi.org/10.1016/j.icte.2018.10.007

Shiferaw, W. S., Aynalem, Y. A., Akalu, T. Y., & Demelew, T. M. (2020). Incidence and Predictors of Recurrence among Breast Cancer Patients in Black Lion Specialized Hospital Adult Oncology Unit, Addis Ababa, Ethiopia: Retrospective Follow-up Study with Survival Analysis. *Journal of Cancer Prevention*, *25*(2), 111–118. https://doi.org/10.15430/jcp.2020.25.2.111

SM, A. (2016). Trends of Breast Cancer in Ethiopia. *International Journal of Cancer Research and Molecular Mechanisms (ISSN 2381-3318)*, *2*(1), 2–6.

https://doi.org/10.16966/2381-3318.121

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *36th International Conference on Machine Learning, ICML 2019*, *2019-June*, 10691–10700.

Tesfaw, A., Alebachew, W., & Tiruneh, M. (2020). Why women with breast cancer presented late to health care facility in North-west Ethiopia? A qualitative study. *PLoS ONE*, *15*(12 December), 1–15. https://doi.org/10.1371/journal.pone.0243551

The Oncologist - 2021 - Ayele - Breast Awareness Self-Reported Abnormalities and Breast Cancer in Rural Ethiopia A.pdf. (n.d.).

Tolessa, L., Sendo, E. G., Dinegde, N. G., & Desalew, A. (2021). Risk factors associated with breast cancer among women in addis ababa, ethiopia: Unmatched case–control study. *International Journal of Women's Health*, *13*, 101–110. https://doi.org/10.2147/IJWH.S292588

杜彬陶沙 卢静 **李媛媛** 马磊磊 **王翠翠**, & Amaliyyah, R. (2021). *No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title*. *February*, 6.

APPENDICES

Recommendation Letter

ASTU, P.O. Box 1888 ☎ 022-110-0073/25 Fax: 0251-022-112-01-50 E-mail: tekluurgessa2009@gmail.com

ADAMA SCIENCE & TECHNOLOGY UNIVERSITY

አዳማ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ

ቀን ፡ መጋቢ 0 2 2015

ቁጥር፡ ፖ/21.4/ኤ 20/3933/15

ለ፡ ኮሪያ ሆስፒታል

ለ፡ አርቴፊሻል ኢንተለጀንስ ተቋም

<u>አዲስ አበባ</u>

ኤሌክትሪካል ምህንድስና እና ኮምፒዩቲንግ ት/ቤት ዲን

ዶ/ር ተክሉ ኡርጌሳ አበበ

ጉዳዩ፡- ትብብር ስለመጠየቅ ይሆናል።

በአዳማ ሳይንስና ቴክኖሎጂ ዩኒቨርሲቲ በኤሌክትሪካል ምህንድስና እና ኮምፒዩቲንግ ት/ቤት ስር የሚገኘው የኮምፒውተር ሳይንስና ምህንድስና ት/ክፍል የድህረ ምረቃ ተማሪ የሆኑት ተማሪ ግርማ አበበ የመመረቂያ ጽሑፋቸውን " Breast Cancer Classification Using Deep Learning " በሚል ርዕስ ለመስራት መረጃ እያሰባሰቡ ይገኛሉ።

በመሆኑም የመመረቂያ ፅሑፋቸውን በተመለከተ ለሚያደርጉት የመረጃ አሰባሰብ አስፈላጊው ትብብር እንዲደረግላቸው እየጠየቅን ጥናታቸውን በኃላፊነት እና በስነ-ምግባራዊ መንገድ እንደሚሰሩና ማንኛውንም መረጃ ለዚህ ጥናታዊ ፅሑፍ ብቻ የሚውል መሆኑን እየገለፅን ከዚህ ውጭ ለሌላ ጉዳይ ቢያውሉት በህግ እንደሚጠየቁ እየገለፅን ለሚደረግላቸው ትብብር በቅድሚያ እናመሰግናለን።

ከሰላምታ ጋር

ዶ/ር ተክሉ ኡርጌሳ አበበ
ኤሌክትሪካል ምህንድስና እና
ኮምፒዩቲንግ ት/ቤት ዲን

<u>ግልባጭ</u>፡

- ለአካዳሚክ ጉዳዮች ም/ፕሬዝዳንት ጽ/ቤት
- ለአካዳሚክ ጉዳዮች ጀኔራል ዳይሬክተር
- ለት/ቤቱ ዲን
- ለት/ቤቱ አካ/ጉ/ተ/ዲን
- ለኮምፒውተር ሳይንስና ምህንድስና ት/ክፍል

  አ.ሳ.ቴ.ዩ

  <u>አዳማ</u>

We are dedicated to innovation knowledge