KWAME NKRUMAH UNIVERSITY OF SCIENCE & TECHNOLOGY, KUMASI
COLLEGE OF ENGINEERING
DEPARTMENT OF TELECOMMUNICATION ENGINEERING

BREAST CANCER DETECTION USING AN AI POWERED WEB APP

By

TETTEH Prince Dawson
APPIAH Eunice Owusuwaa
DOE-KIINNII Bubuney Korku Kpesu

A THESIS SUBMITTED TO THE DEPARTMENT OF TELECOMMUNICATION ENGINEERING, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF B.Sc. TELECOMMUNICATION ENGINEERING

AUGUST 2024

DECLARATION

We hereby declare that this submission is our own work towards the BSc. Telecommunication Engineering degree and that, to the best of our knowledge, it contains no material previously submitted by another person, nor material which has been accepted for the award of any other degree of this or any other university, except where due acknowledgement has been made in the thesis write-up.

ACKNOWLEDGEMENT

Words alone cannot sufficiently express our sincere and heartfelt gratitude to all who have helped in one way or the other to see to the successful completion of our undergraduate program in this highly esteemed institution.

We also want to thank our project supervisor, Prof. Jerry John Kponyo, for his stupendous effort and guidance given us during the course of our final year project, we are most grateful.

ABSTRACT

This thesis explores the development of an AI-powered web application for the early detection of breast cancer, aimed at enhancing radiology diagnostics and reducing misdiagnosis rates. The research focuses on leveraging machine learning (ML) models, particularly convolutional neural networks (CNNs), to classify breast tissue images into benign or malignant categories. The model was trained using the Digital Database for Screening Mammography (DDSM), employing VGG19 architecture, and achieved an accuracy of 97.92%. This result significantly surpasses benchmarks, highlighting the model's reliability.

The web application integrates a React.js front-end with a FastAPI back-end, enabling seamless interaction between users and the ML model. Features include patient data management, mammogram uploads, and real-time classification results. The tool enhances diagnostic efficiency, reduces radiologist error rates, and promotes early-stage cancer detection.

Key contributions include the model's superior performance, the practical application of AI in medical imaging, and recommendations for further research, including federated learning to address privacy concerns and extending the application to other cancer types. This work underscores the potential of AI to revolutionize healthcare and improve patient outcomes.

Table of Contents

List Of Figures

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Overview

This chapter provides an overview of the thesis, outlining the background, problem statement, objectives, contribution and structure of the thesis.

This chapter provides an overview of the thesis, outlining the background, problem statement, objectives, contribution and structure of the thesis.

1.2 Background

Breast cancer is a disease in which cells in a person's breast tissues change and divide uncontrolled typically resulting in a lump or mass. It is reported as the leading cause of death especially in women in developing countries and the second leading cause of death after lung cancer in the developed countries hence it poses a menace to society. Breast cancer at its latter stage is fatal hence its early diagnose is crucial for patient's survival. Breast cancer screening and detection is done in medical imaging by machine assisted systems for the initial finding of the cancer and to enhance the skill of radiologists.

Machine assisted systems are not only the latest development in medical imaging for the initial finding of breast cancer but also enhance the diagnostic skill of radiologists. The most common tools employed for breast cancer diagnostics are mammography, tomography, Breast Ultrasound (BUS), MRI, CT scans and even more deeply PET is advised. Typically, the breast is counted as the oversensitive organ of the human body, so only some of these procedures are advised, which depends upon the patient's condition and the tumour status. Mammography is considered a low-cost and secure procedure at an early stage of breast cancer, but it is ineffective in dense breast of young female. BUS procedure is considered supportive to mammograms to prevent needless biopsy. Several datasets of breast imaging are publicly available which are DDSM, MIAS, WBCD, BCDR and NBIA etc. Following image acquisition, various operations of pre-processing are performed before segmentation such as pectoral muscle removal and artefacts removal etc. The process of segmentation is the most important step of the machine-assisted system for enhancing accuracy and to reduce false positive of the existence of abnormality. Numerous studies recommended Gray Level Co-occurrence Matrix (GLCM) method to describe texture-based features. Similarly, Local Binary Pattern (LBP) is another remarkable mechanism used for texture extraction to isolate benign masses from malignant ones.

The diagnosis of breast cancer greatly depends upon the classification performance. Several machine learning approaches such as neural networks, decision trees, K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Ensemble classifiers are applied for the training and testing of features to distinguish the objects into a malignant or benign class. The issue with many groups is addressed with the aid of a decision tree classifier and the prediction of subtypes of breast cancer using the same or lesser number of genes is 100 % accurate. The use of machine learning techniques is a breakthrough in life sciences particularly the use of deep learning architectures that have generated encouraging results. Currently, CNN has attracted researchers for breast tumour detection and classification.

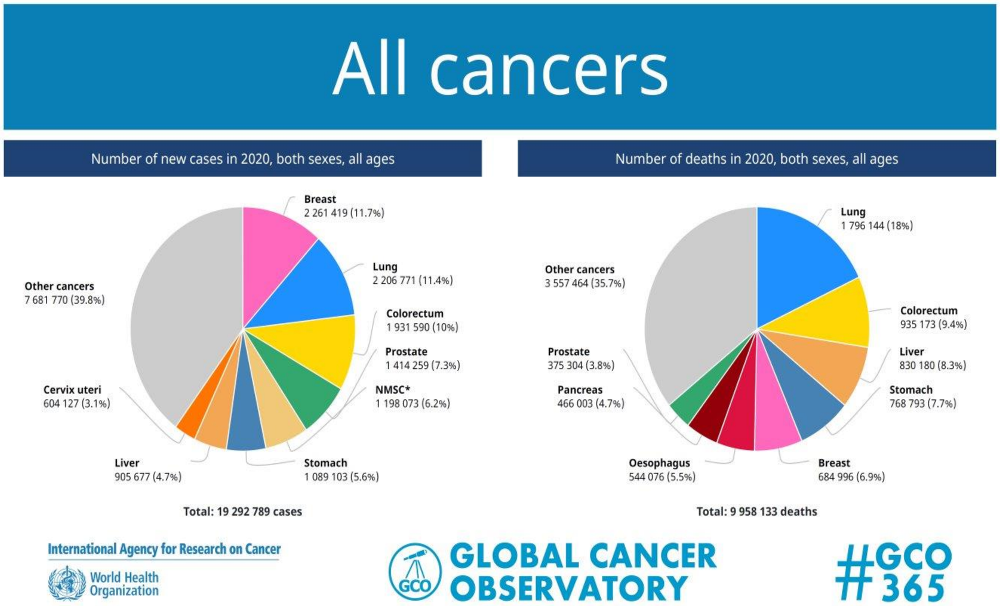


Figure 1: Shows the Global Cancer Observatory Chart for 2020

1.3 Problem Statement

The complexity of breast cancer diagnosis is compounded by factors such as subtle variations in tissue appearance, overlapping features between benign and malignant lesions, and differences in individual radiologists' experience and expertise. Studies have shown that even experienced radiologists can misdiagnose breast cancer, with reported diagnostic error rates ranging from 20% to 50%. These errors can result from human limitations, fatigue, image quality, and the inherent challenges in distinguishing between different types of breast lesions.

Given the significant impact on patient health and well-being, there is a need to reduce misdiagnosis rates in breast cancer detection. Improving radiologists' diagnostic accuracy through advanced tools such as artificial intelligence (AI)-assisted diagnostic.

1.4 Research Objectives

1.4.1 General Objectives

The main objective is to develop an AI Powered Web Application for breast cancer detection aimed at enhancing radiology diagnosis and treatment.

1.4.2Specific Objectives

This will be achieved by following the following specific objectives:

1. Train a selected model with stage one Breast Cancer images
2. Build a web application
3. Interface the Machine Learning model with the web application.

1.5 Justification of Study

Breast cancer remains one of the most prevalent and life-threatening cancers among women worldwide. Early detection and accurate diagnosis are critical in reducing mortality and improving patient health. Machine learning (ML) models have emerged as powerful tools in healthcare, offering promising solutions to enhance breast cancer detection. The importance of developing ML models for this purpose lies in their potential to provide early detection, increase diagnostic accuracy, handle large datasets, support personalized treatment, reduce healthcare costs, and assist resource-limited settings.

1.6 Significance of Study

By developing this tool, the rate at which radiologists make mistakes will be reduced significantly and patient’s life will be saved.

CHAPTER TWO

LITERATURE REVIEW

2.1 OVERVIEW

Here we discuss some related theoretical perspective that guides the research and some papers that have been reviewed.

2.2 TRADITIONSL BREAST CANCER DETECTION

Global statistics have demonstrated that breast cancer is the most frequently diagnosed invasive cancer and the leading cause of cancer death among female patients. Survival following a diagnosis of breast cancer is grossly determined by the stage of the disease at the time of initial diagnosis, highlighting the importance of early detection. Improving early diagnosis will require a multi-faceted approach to optimizing the use of currently available imaging modalities and investigating new methods of detection. The application of microwave technologies in medical diagnostics is an emerging field of research, with breast cancer detection seeing the most significant progress in the last twenty years. In this review, the application of current conventional imaging modalities is discussed, and recurrent shortcomings highlighted. Microwave imaging is rapid and inexpensive. If the preliminary results of its diagnostic capacity are substantiated, microwave technology may offer a non-ionizing, non-invasive, and painless adjunct or stand-alone modality that could possibly be implemented in routine diagnostic breast care. In this era of a rising incidence of breast cancer, ensuring diagnosis at the earliest possible stage requires further improvement in the capabilities of current diagnostic modalities and the development of novel imaging systems. Here current conventional imaging applications are reviewed, such as mammography, digital breast tomosynthesis, ultrasonography, and magnetic resonance imaging. These modalities exploit a variety of properties of biological tissues to form an image.

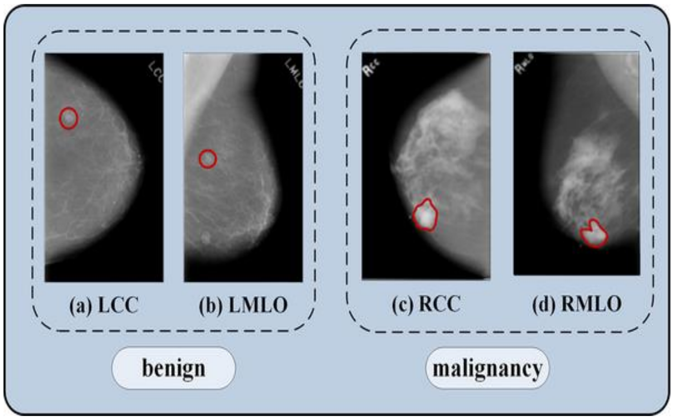


Figure 2.1: Shows a benign and malignant sample of a mammogram.

Mammography and digital breast tomosynthesis exploit the attenuation of x-rays as they pass through breast tissue measured by the attenuation coefficient of the tissue;

Magnetic resonance imaging uses radio waves, magnetic field gradients, and contrast agents to excite and measure the location of hydrogen atoms;

Ultrasound exploits differences in acoustic impedance between tissue types as sound waves propagate in the breast.

The potential of microwave breast imaging, a non-ionizing imaging modality that represents a promising method of breast cancer detection, is then discussed. Microwave breast imaging exploits the dielectric properties of biological tissues to form images, potentially providing complementary information to conventional modalities. While it is not expected that the acoustic, x-ray attenuation and dielectric properties are completely independent (all of the x-ray attenuation, acoustic impedance, and dielectric properties are influenced by water content), it is still expected that the dielectric properties would be influenced by other factors such as fractions of bound water and ion concentrations, suggesting that microwave breast imaging may be a useful adjunct imaging modality.

2.2.1 METHODS OF THE TRADITIONAL DETECTION

Mammography

Although the use of radiography to differentiate benign from cancerous specimen tissue was recorded as far back as 1913, its use to clinically evaluate the symptomatic patient was delayed until the 1930′s, when Stafford Warren, who is credited with the invention of the mammogram, pioneered a stereoscopic technique for mammography. To this day, mammography remains the gold standard investigation of symptomatic women aged 40 and over and for the whole population breast cancer screening. While it offers a cost-effective method of investigating breast cancer, its effectiveness is severely hindered by several limitations. For most women, breast compression is an uncomfortable experience. In a study of 954 patients, Keemers-Gels et al. reported that as many as 79% of patients undergoing breast cancer screening found mammography to mild to severely painful. Additionally, this study established that the pain associated with breast compression during mammography was the main deterrent for women who indicated that they would not attend for further screening. Mammography is also limited in identifying tumours that present without a characteristic mass (as is frequently the case with invasive lobular breast cancer), without calcification (as can occur in entities like non-calcified ductal carcinoma), and in breasts of higher density. The sensitivity of screening mammography, which aims to detect pre-clinical breast cancer in asymptomatic women, is estimated to be between 68 and 90%, with the lower margin of this range applicable to mammographically dense breasts, as is common in younger aged women. Even in women with lower breast density (> 75 years), the sensitivity for screening mammography is 88.4%. With diagnostic mammography, which involves the evaluation of patients with symptoms and signs suggestive of breast cancer, the sensitivity is slightly higher, at 93%.

Randomized Controlled Trials (RCTs)

Uninterrupted controversy continues to surround screening mammography and the true impact it has on breast cancer. While early breast cancer detection is undoubtedly beneficial for the patient, the benefits of screening appear to be less than as first thought. On the one hand, the use of this modality in organized screening programs remains the only screening test proven to reduce breast cancer mortality supported by randomized trials and subsequent meta-analyses. In the randomized control trials (RCTs) and meta-analysis (Table 1) conducted on screening mammography deemed to be of sound methodologic quality, there was a reduction in breast cancer mortality between 20% to 45% in female participants aged 40 to 70 years. Modern mammography screening cites a similar reduction in breast cancer mortality of 28%.

However, these RCTs have been subject to criticism, with concerns raised that they are irrelevant to the modern era of breast cancer management as they were conducted prior to the introduction of taxane adjuvant therapy and the standardized acceptance of adjuvant hormone treatment for oestrogen receptor-positive disease. This has led some authors to doubt that the benefits of screening mammography would persist under present conditions.

In a seminal publication in 2012, Bleyer et al. interrogated the US Surveillance, Epidemiology, and End Results data to examine trends of early- and late-stage breast cancer among women 40 years of age or older over a 30-year period. In this study, the group demonstrated that screening mammography resulted in a doubling in the annual incidence of early-stage breast cancer that is detected. However, it was calculated that breast cancer was over diagnosed in up to 31% of all cases, accounting for 1.3 million women over the past 30-year study period. Furthermore, this group demonstrated that screening mammography only marginally reduced the rate of incidence of late-stage breast cancer and concluded that screening was having a minor impact on breast cancer mortality overall.

MRI Guided Breast Biopsy

Physical, mammography, and other exams often detect lumps or abnormalities in the breast. However, these tests cannot always tell whether a growth is benign or cancerous.

Doctors use breast biopsy to remove a small amount of tissue from a suspicious area for lab analysis. The doctor may perform a biopsy surgically. More commonly, a radiologist will use a less invasive procedure that involves a hollow needle and image-guidance. Image-guided needle biopsy does not remove the entire lesion. Instead, it obtains a small sample of abnormality for further analysis. Image-guided biopsy uses ultrasound, MRI, or mammography imaging guidance to take samples of an abnormality.

In MRI-guided breast biopsy, magnetic resonance imaging is used to help guide the radiologist's instruments to the site of the abnormal growth.

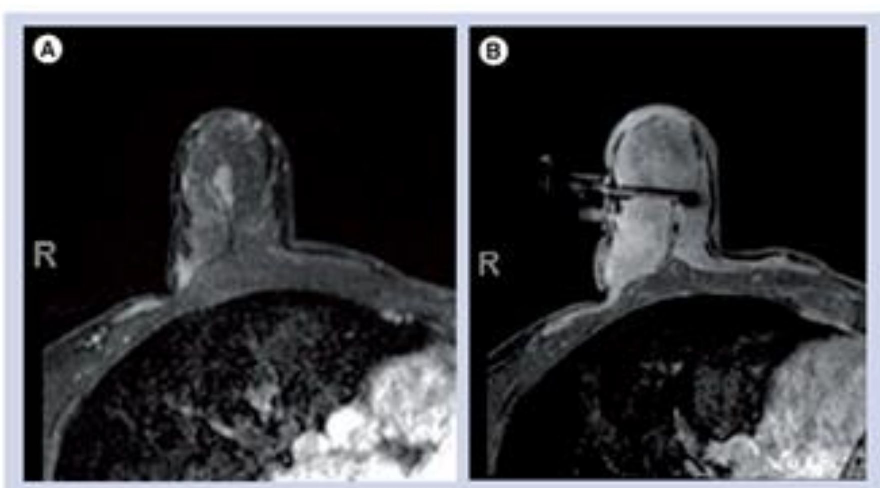


Figure 2.2 : Shows MR-guided breast biopsy. (A) Post contrast-enhanced MR axial image demonstrating enhancing lesions and (B) lesion obscured by background enhancement and obturator.

An MRI-guided breast biopsy is most helpful when MR imaging shows a breast abnormality such as:

- a suspicious mass not identified by other imaging techniques
- an area of distortion
- an area of abnormal tissue change

Doctors use MRI guidance in biopsy procedures that use:

- vacuum-assisted device (VAD), which uses a vacuum powered instrument to collect multiple tissue samples during one needle insertion.
- wire localization, in which a guide wire is placed into the suspicious area to help the surgeon locate the lesion for surgical biopsy. MRI is often able to tell the difference between diseased tissue and normal tissue better than x-ray, CT, and ultrasound.

Using MRI guidance to calculate the position of the abnormal tissue and to verify the placement of the needle, the radiologist inserts the biopsy needle through the skin, advances it into the lesion and removes tissue samples. If a surgical biopsy is being performed, MRI may be used to guide a wire into the mass to help the surgeon locate the area for excision.

Breast Sonography

Breast sonography offers a non-ionizing, high resolution, low cost, highly sensitive (81.7%) and specific (88%) instrument for investigating symptomatic patients. Using a handheld transducer, breast sonography has a definitive role in establishing the relationships between glandular, fat, and fibrous components of breast tissue based on the different acoustic impedances of these tissues. This is difficult to elucidate by mammography alone, as the fibrous and glandular components of the breast have similar X-ray attenuation coefficients. While initially employed as a diagnostic adjunct to further assess mammographic findings and palpable abnormalities, technical advances, including improved spatial and contrast resolution, and higher-megahertz (MHz) transducers have facilitated the use of US to characterize solid masses and provide uncomplicated image guidance for needle biopsy. Breast sonography is the modality of choice for further investigation of palpable breast findings that are not clearly benign and mammographic screen-detected abnormalities. Furthermore, breast sonography is advised as an initial investigation prior to mammography in women younger than 30 years or in female patients that are lactating or pregnant. This is recommended both to prevent unnecessary radiation exposure to a population of patients who carry a low incidence of breast cancer and, as women under 30 years of age, generally have a relatively higher breast tissue density.

Prospects for Microwave Breast Imaging

Despite its limitations, mammography remains the gold standard against which new imaging technologies are measured. The adjunct of sonography and, more recently, tomosynthesis into clinical practice have improved breast cancer detection, while the increasing use of breast MRI is having a favourable impact when used in the appropriate setting. Although technological advances have brought about improvements in imaging quality and developments in modalities, there remains a chasm between current and optimal breast cancer detection levels, with approximately 2.4–6% of patients still being identified with de novo stage-IV breast cancer at the time of initial diagnosis. While mortality rates are declining, the incidence of this multifaceted disease continues to increase with survival dependent on the stage at diagnosis. Microwave imaging is rapid and inexpensive and may offer an additional adjunct to mammography, or perhaps a stand-alone modality in the future. If the exciting preliminary results of several operating systems are substantiated, the non-ionizing, non-invasive, and painless characteristics of this modality will popularize its implementation in routine diagnostic breast care. While several study populations to date have been too small to determine clinical efficacy accurately, the favourable preliminary results from the larger trials such as the DC and MARIA® have indicated that microwave imaging may have an integrated role in breast imaging in the near future. From a clinical perspective, further clinical trials will require a significantly higher number of patients to validate the true potential of the systems. Ideally, these trials should have a variety of breast pathology to represent the heterogeneity of breast disease, including invasive ductal carcinoma, invasive lobular carcinoma, fibroadenoma, and cystic breast disease.

As we approach an age where the translation of microwave breast imaging systems into the clinical setting has become a reality, it will be imperative that the findings from microwave breast imaging studies are interrogated by breast radiologists with an aim to determine consensus opinions about multiple factors. These will include issues like reporting methods and terminology for describing findings, image presentation for the radiologist (i.e., 3D representation or standard anatomical planes such as coronal, axial and sagittal), imaging features (e.g., clusters/ foci of signal) and limits of signal that may suggest malignancy or benign disease. Furthermore, it will be crucial that the overseeing physician (radiologist), will play a role in developing standard operating procedures to limit any psychological or physical side-effects of MBI on a patient, considering the sensitivity of cancer imaging.

2.3 MACHINE LEARNING IN BREAST CANCER DETECTION

AI-powered software can automate interpretation of breast mammograms, ultrasounds, and MRI scans to get patients their results faster. AI techniques can help radiologists identify breast cancer that would have otherwise been undetectable in its early stages.

The software can store and evaluate large datasets of images and identify patterns and abnormalities that human radiologists might miss. It typically highlights potential problem areas in an image and assesses any likely malignancies. Machine learning (ML) plays a crucial role in breast cancer detection by improving the accuracy and efficiency of diagnosis. Here's a detailed explanation of how ML is used:

1. **Image Analysis:** ML algorithms are trained on large datasets of mammography images to learn patterns and features associated with breast cancer.

2. **Feature Extraction:** Algorithms extract relevant features from images, such as texture, shape, and density, to help identify potential tumours.

3. **Classification:** ML models classify images as malignant or benign based on extracted features.

4. **Risk Assessment:** ML algorithms analyse patient data, medical history, and image features to predict breast cancer risk.

5. **Segmentation:** ML helps segment tumours from surrounding tissue, allowing for more accurate diagnosis and treatment planning.

6. Detection of Micro calcifications: ML algorithms detect micro calcifications, which can be an early indicator of breast cancer.

7. Enhancement of Image Quality: ML improves image quality, reducing noise and artifacts, to facilitate more accurate diagnosis.

8. Computer-Aided Detection (CAD) Systems: ML-powered CAD systems assist radiologists in identifying potential breast cancer cases.

9. Deep Learning: Techniques like convolutional neural networks (CNNs) and transfer learning improve the accuracy of breast cancer detection.

10. Continuous Learning: ML models learn from new data, adapting to improve detection accuracy over time.

2.3.1 GENERATIVE ADVERSARIAL NETWORK FOR DETECTION IN BREAST CANCER DOMAIN

GAN was employed by Guan and Loew as a new mammographic image generator from the DDSM datasets, while CNN was utilized as GAN's discriminator. Compared to other image-augmentation methods, GAN performed better. Strategies such as under-sampling, over-sampling, and feature selection can deal with the adverse effects that occur from imbalanced source data. In recent years, novel data-augmentation strategies such as Generative Adversarial Networks (GANs) have been used to artificially generate additional data. Generally, GANs are employed to image data and comprise two sub-networks: Generator and Discriminator. The role of the Generator is to generate synthetic samples, whereas the Discriminator is designed to discriminate between fake and real samples. In other words, the function of the Generator is to produce samples with features that the Discriminator cannot separate from real samples, thus enriching the original dataset. Compared to other generative approaches, GANs have a higher computational speed and enhanced sample quality. Therefore, GANs are considered superior to other methods.

Additionally, GANs show a lower possibility of overfitting classifier risk and are less vulnerable to the impacts of non-pertinent sample features. Data augmentation using generative models are highly effective because only a specific patch of the entire sample needs to be augmented. The applicability of GANs in mammograms has potential for many reasons. For instance, GANs can overcome the unavailability of significant original datasets. Additionally, public datasets comprise only a small proportion of malignant samples in the general population. Another reason for the applicability of GAN is that they are advantageous and may improve cancer detection that could be used in screening. For example, generating synthetic images that can be used to train and validate cancer detection algorithms, identifying anomalies in images that may not show obvious signs of cancer, and generating personalized cancer screening images for each patient. These methods have been shown to be effective in detecting breast cancer, prostate cancer, and other types of cancer. Overall, GANs have the potential to significantly improve the accuracy and effectiveness of breast cancer screening.

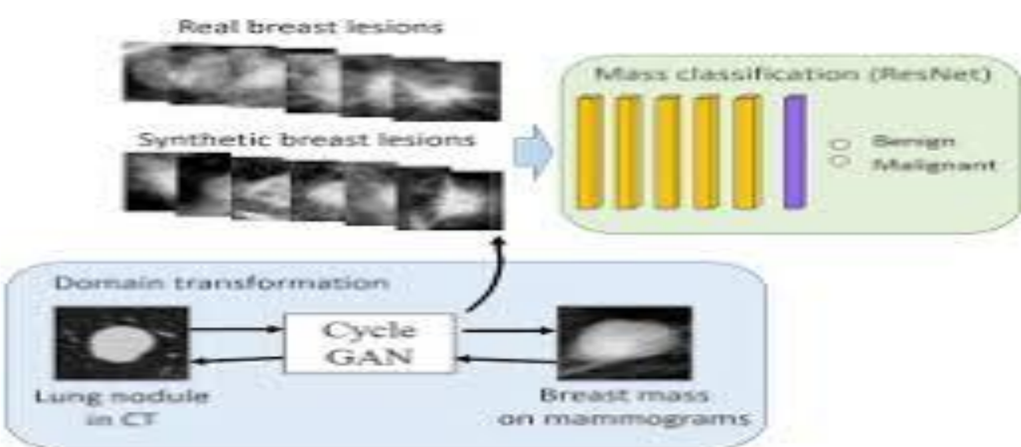


Figure 2.3: Shows the operation of GANs in breast cancer screening.

2.3.2 CONVOLUTIONAL NEURAL NETWORK

Convolution neural network has also been used in order to classify breast cancer histology images in order to provide early detection of breast cancer at low cost and high accuracy. A feed forward neural network can also be devised to statistically diagnose breast cancer with a very high accuracy. CNN model is employed to extract various features based on the validated gene expression data in order to detect clinical outcomes in breast cancer. Some authors use CNN to detect the mitosis process for the invasive breast cancer diagnosis based on histopathological imaging. A large number of studies have proved that CNN shows superior performance in breast cancer diagnosis. The proposed system uses CNNs to detect breast cancer from breast tissue images. The architecture of a CNN has 3 main layers, the convolutional layer, pooling layer, and fully connected layers.

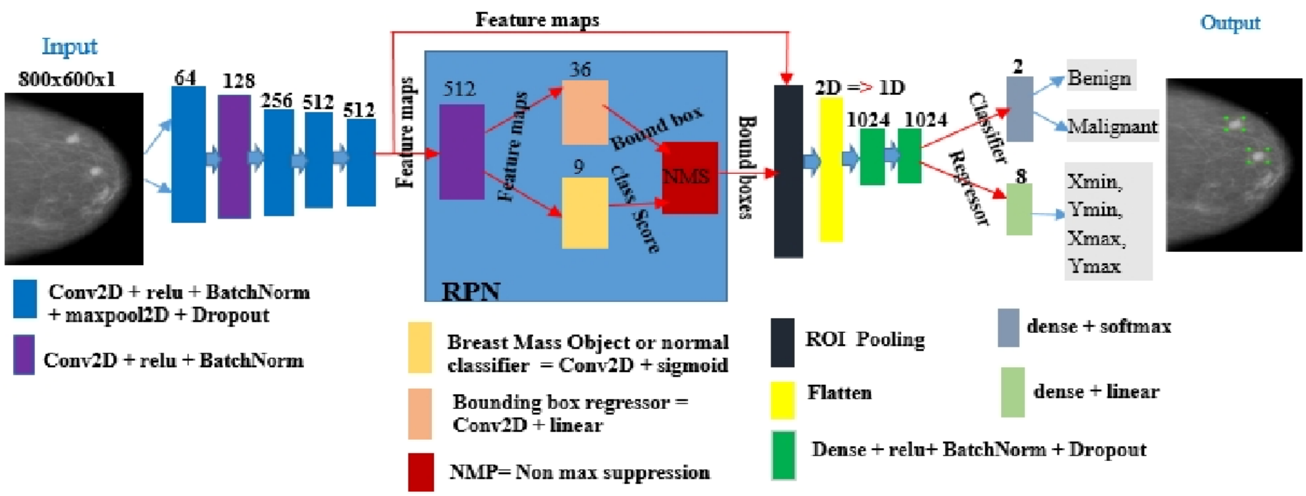


Figure 2.4: Shows the architecture of the CNN model.

Convolutional Layer:

This layer extracts features from the breast tissue images using filters or kernels. The filters scan the image, performing a dot product to generate feature maps. These feature maps highlight patterns such as edges, textures, and shapes in the image. In breast cancer detection, the convolutional layer helps identify features like tumour margins, calcifications, and architectural distortions.

Pooling Layer:

This layer down-samples the feature maps generated by the convolutional layer. It reduces spatial dimensions, retaining important information while decreasing the number of parameters. Pooling helps the network become invariant to small translations and distortions in the image. In breast cancer detection, pooling aids in identifying larger structures like tumours and glandular tissue.

Fully Connected Layer:

This layer flattens the output from the pooling layer and connects every neuron to every other neuron. It's responsible for making predictions based on the features extracted and down-sampled. In breast cancer detection, the fully connected layer uses the extracted features to classify images as malignant or benign. It outputs probabilities for each class, allowing for diagnosis and risk assessment.

2.4 REVIEW OF EXISTING LITERATURE

2.4.1 Introduction to the Literature Review

The examination of existing literature is an essential component of this thesis, presenting an intricate analysis of current research and progressions in breast cancer detection, with a specific focus on early diagnosis of breast cancer at stage one. The objective of this segment is to scrutinize the various approaches made in breast cancer detection using artificial intelligence and machine learning.

2.4.2 Review of Studies on Breast Cancer Detection

Several studies showed the use of deep learning algorithms in breast cancer detection, highlighting its benefits or impacts and challenges. One study by Shrinivas D Desai et al (2020) [1] employed the use of Deep Convolution GANs for generating synthetic images. This study also investigates data augmentation techniques for limited labelled medical image dataset. Image analysis was done using ImageJ Visual Turing Test for evaluating the GAN-generated images' appearance. This research found that combining Deep Convolution GAN generated synthetic images with original training dataset catered for the supply of training data. However, it also noted concerns about the insufficiency of available labelled medical images and the scarcity of training data.

Another study by Alexander Rakhlin et al (2018) [2] investigated feature extraction using Deep CNNs. This method employed the use of pre-trained CNNs on ImageNet and strong data augmentation techniques. This study found that fused model accuracy was higher than the individual constituents and standard deviation of ensemble is twice as low as individual models. However, its limitation was that diversity and generalizability of the results may be limited because the dataset used is an extension of a previous dataset.

In a similar vein, Tariq Mahmood et al (2021) [3] researched on a deep learning based convolutional neural network that learns features automatically for breast lesion classification was used. Based on the experimental outcome, it is perceived that the suggested architecture performed marginally better compared to traditional deep learning schemes. This study also highlighted the challenges in early screening for breast cancer abnormalities and also the high false positive ratio for mass detection due to handcrafted features.

Furthermore, H Sami (2021) [4] employed SVM with linear and polynomial kernels for breast lesion prediction and utilized open open-source datasets from the University of Manitoba for training. The result was that SVM algorithm with polynomial kernels gives promising predictions for breast cancer detection. This study highlighted that the lack of information on the size and diversity of the dataset used for training and testing the machine learning algorithm.

A study by N Siddique (2020) [5] employed base U-net, Residual block, Attention gate, Dense block, Inception block, Modified parallel U-net for medical image segmentation. This study emphasised U-net as a powerful tool for medical image segmentation and it's invaluable for medical image analysis. However, it also highlighted the computational power constraints for deep learning techniques.

2.4.3 Recommendations for Future Developments

The subsequent recommendations put forth for the advancement of breast cancer detection based on the limitations identified:

Sufficiency of labelled medical images to solve the issue of scarcity of training data to prevent generalizability and limited diversity of results. Adequate information on the size and diversity of the dataset used for training and testing the machine learning algorithm.

Enhancement in feature extraction and examination to cause a decline in the high false-positive ratio for mass detection. Pre-trained models for high accuracy should be employed to solve the challenges associated with early screening for breast cancer abnormalities.

CHAPTER THREE

RESEARCH AND METHODOLOGY

3.0 Introduction

In this chapter we introduce a paradigm shift system for diagnosing breast cancer and we will be doing that using a web app powered by a machine learning model. The entire process outlined in this chapter can be seen in **Fig 6**.

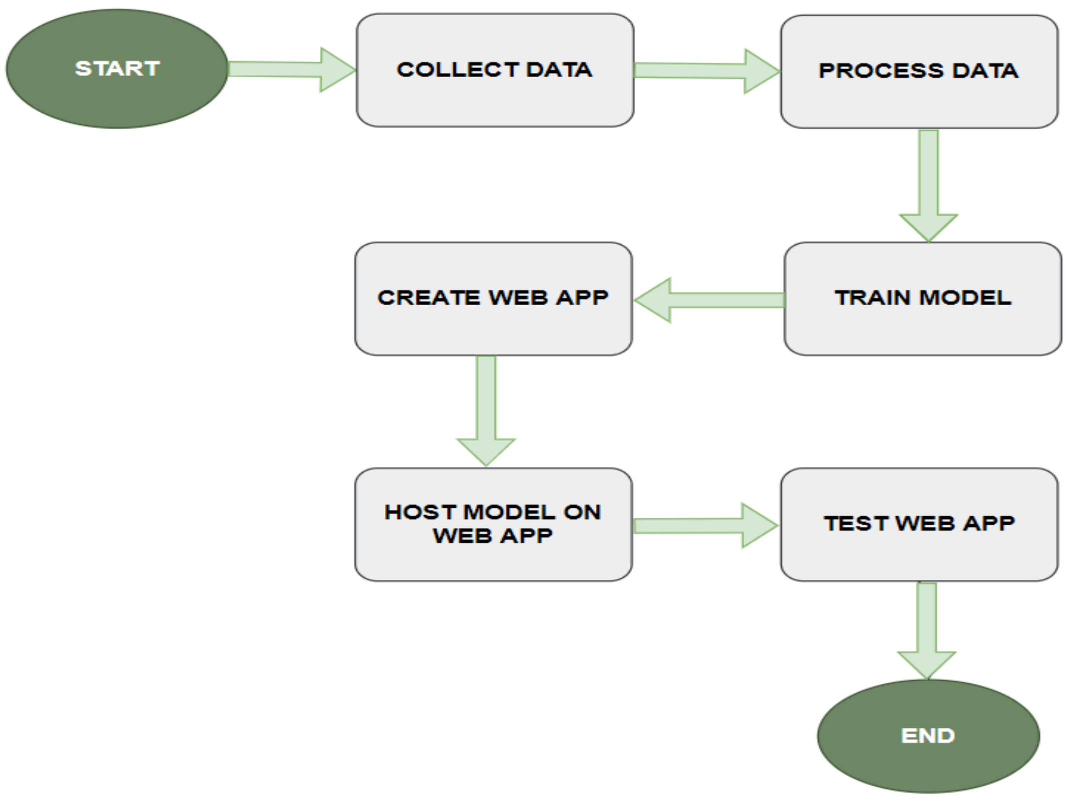


Figure 3.1: Shows the methodology of the research project.

3.1 COLLECTION OF DATA

With the dataset used to train our model, we will use the Digital Database for Screening Mammography (DDSM) - a resource for use by the mammographic image analysis research community.

The dataset is made up of 5970 benign samples and 7185 malignant samples and out of this, we will use 3000 benign samples and 3540 malignant for the training dataset, 2360 benign and malignant 2360 for the validation dataset and 140 samples for the test dataset.

3.2 PREPROCESSING THE DATA

With the data pre-processing, we will implement techniques such as rescaling, rotation range, width shift range, height shift range, shear range, zoom range and horizontal flop when generating the data for training. These processes will help condition our data well for training our model.

3.3 TRAINING OF THE MODEL

With the model we will use for the training, we will leverage on the VGG19 network architecture below.

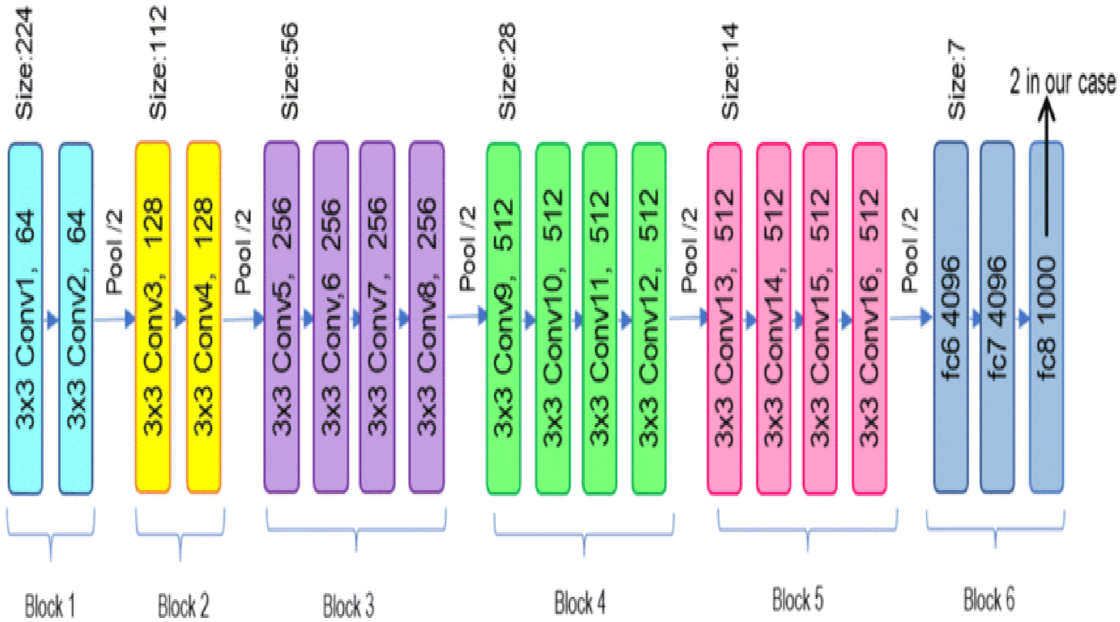


Figure 3.2: Shows the architecture of VGG19

The reasons for we going for this architecture are as follows:

a) **Deep Architecture:** VGG19 is a deep network with 19 layers, including 16 convolutional layers and 3 fully connected layers. This depth allows the network to learn more complex features and patterns from the images, which can be particularly beneficial for distinguishing between malignant and benign breast tissue.

b) **Pre-trained on Large Datasets:** VGG19 is often pre-trained on large datasets like ImageNet, which contains millions of images across thousands of classes. This pre-training helps the network learn general image features, which can then be fine-tuned on smaller datasets specific to breast cancer. Transfer learning using pre-trained models like VGG19 can improve performance, especially when labelled medical data is limited.

c) **Good Feature Extraction:** The design of VGG19, with small receptive fields (3x3 convolution filters), allows it to capture fine details in the images. This is crucial in medical imaging, where small differences in texture, shape, and structure can be indicative of different types of breast cancer.

Leveraging on these capabilities, we will freeze the last layer of the VGG19 network and add 2 extra fully connected layers. We will also implement some regulation techniques such as dropout and L2 regularization in order to cater for issues of bias and variance after the training of the model. Fig 8 and Fig 9 is a representation of our model structure.

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0

Figure 3.3: Representation of our model structure

block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 512)	12,845,568
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 1024)	525,312
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 2)	2,050

Total params: 33,397,314 (127.40 MB)
Trainable params: 13,372,930 (51.01 MB)
Non-trainable params: 20,024,384 (76.39 MB)

Figure 3.4: Representation of our model structure continuation

3.4 CREATION OF WEB APP AND HOSTING OF THE MODEL

3.4.1 FRONT-END

We will use React JS and Tailwind CSS to build the front-end part of our web app. We will have included features such as a sidebar containing patient data, an upload box for uploading the mammograms and a text field for inputting patient full name.

3.4.2 BACK-END

For the back-end of our web app, we will use Fast API to serve the prediction to the front-end with the help of the model. That is, when a mammogram is uploaded, the image is sent to the back-end through an address. This image received by the back-end is processed and sent to the trained model for the model to make a prediction. This prediction by the model will then be sent to the front-end to be displayed to the user.

3.4.3 HOW THE WEBAPP WORKS

The figure below shows how our web app will work.

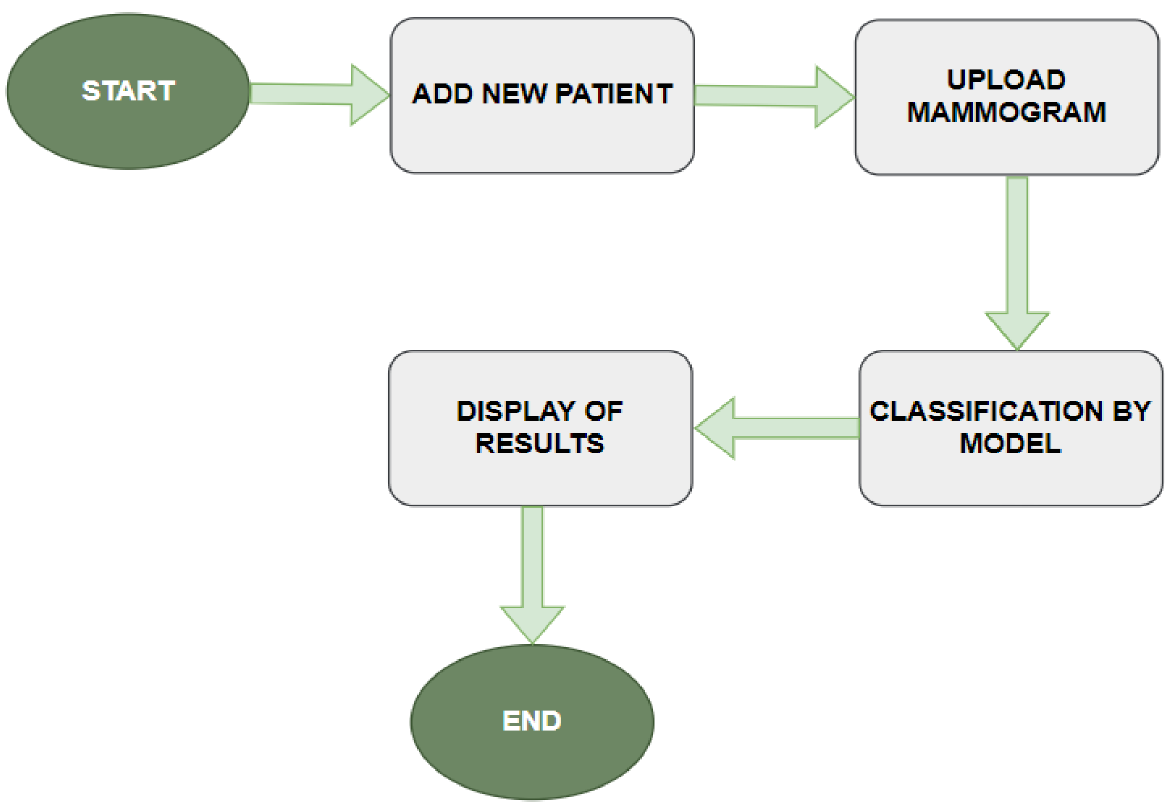


Figure 3.5: How the Web App operates

1. ADD NEW PATIENT: Here the radiologist will input the name of the patient whose mammogram is going to be analysed.
2. UPLOAD MAMMOGRAM: Here the radiologist will upload the mammogram on to the web app to be processed at the back end.
3. CLASSIFICATION BY MODEL: The processed image will then be sent to the model for it to classify whether the image is benign or malignant.
4. DISPLAY OF RESULT: After the model is done with the classification of the image, the result will be sent to the front end of the web app for it to be displayed.

3.5 TESTING THE WEB APP

We will test our web app with several mammograms to gauge how well the model performs.

CHAPTER FOUR

RESULTS AND DISCUSSION

Within this chapter, an evaluation will be conducted to assess the results of the research, taking into consideration the methodology adopted in chapter three. The evaluation encompasses the examination of the performance of the model that was trained and the operation of the web app.

4.0 MODEL RESULTS

We trained our model for 40 epochs and tested it on some unseen data and these were the results;

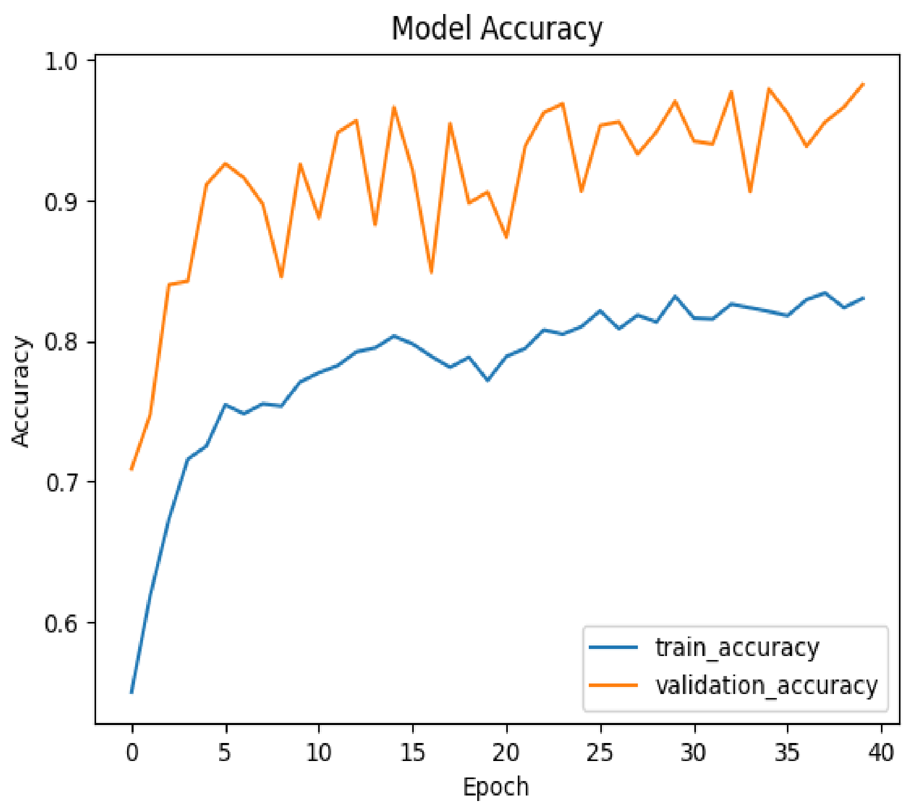


Figure 3: Shows the training and validation accuracy of the model

Table 0.1 shows the model's performance on unseen data

	ACCURACY	RECALL	F1 SCORE
OUR MODEL	97.92	97.92	97.92

Comparing this performance with that of the VGG19 model used in our benchmark paper – Shen, L., Margolies, L.R., Rothstein, J.H. *et al* – this is how our model performed;

Table 0.2 compares the performance of the model with our benchmark

	ACCURACY (%)
Shen, L., Margolies, L.R., Rothstein, J.H. *et al.*	89
Our Model	97.92

4.1 SIGNIFICANCE OF RESULTS

Our model having an accuracy of 97.92% indicates that it is able to correctly classify 97.92 out of 100 mammograms that it is fed with and this is an encouraging step towards curbing the issue of misdiagnosis caused by radiologists thereby saving patient life.

4.2 THE WEB APP

For our web app, we successfully built the front-end part using React Js and Tailwind CSS and the backend using Fast API. The image below shows how the web page looks.

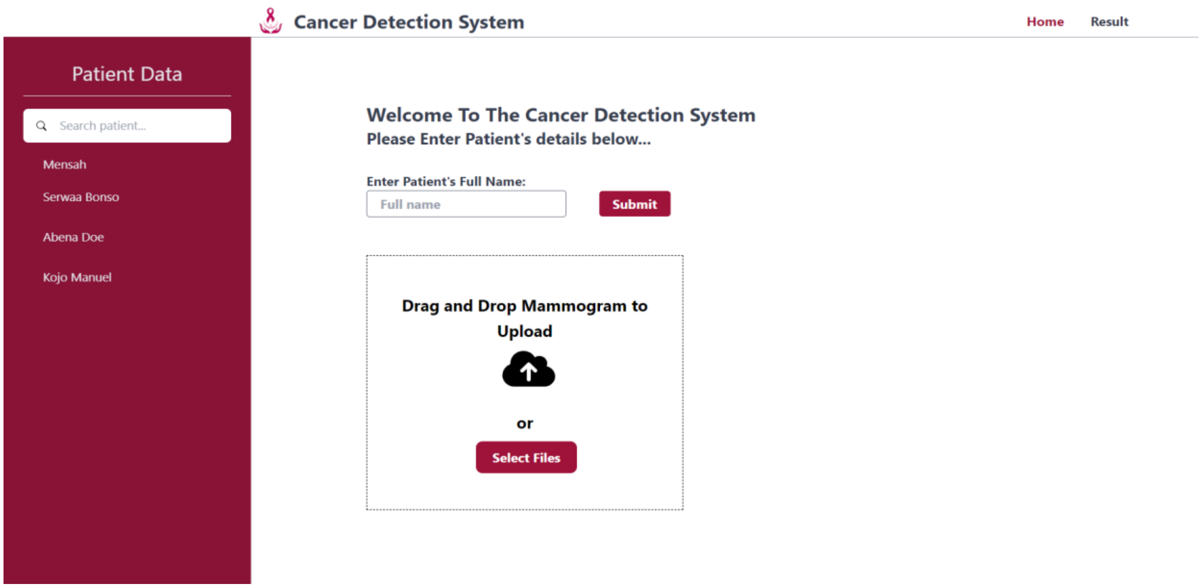
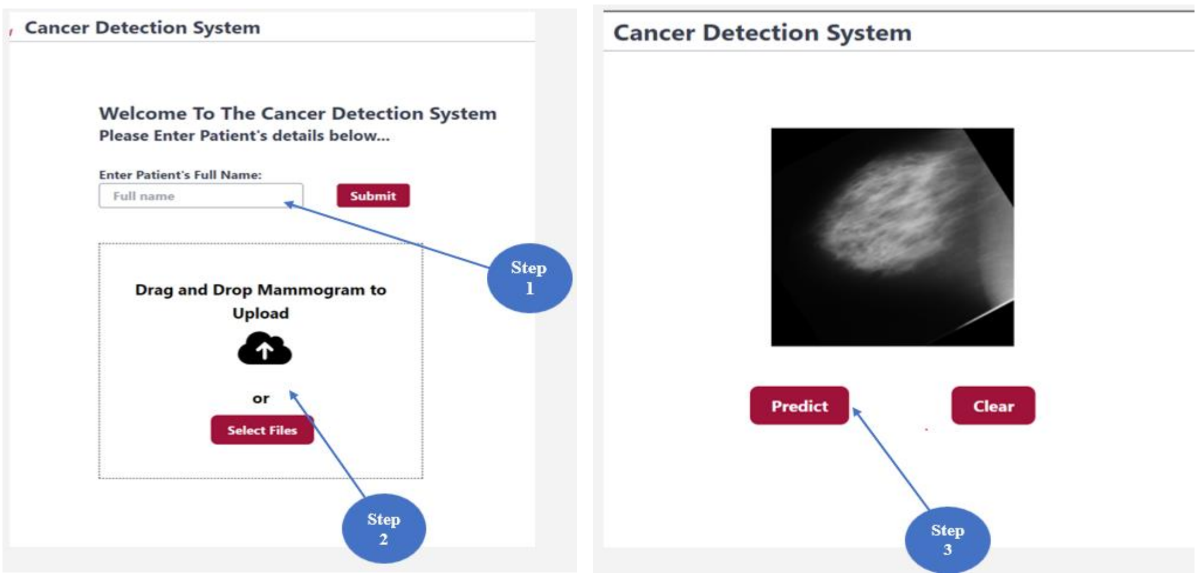


Figure 4: Shows the homepage of the web app.

The subsequent images show the steps involved in getting the model to make a classification.



CHAPTER FIVE

CONCLUSION AND FUTURE RECOMMENDATION

5.0 Conclusion

This research project focused on the early detection of breast cancer using an AI powered web app. We were able to create a tool that will aid radiologist in diagnosing and detecting breast cancer, thereby boosting radiologist efficiency, reduce the number of misdiagnosed cancer cases and enable early diagnose of cancer.

5.1 Future Recommendation

Going forward we recommend that more stakeholders come on board and make datasets available to be used to train the model.

Also, the model can be trained on other cancers in order to increase its capabilities and use by radiologists. Lastly, we recommend that federated techniques are implemented in the training of the model. This will solve the issue of privacy when implemented successfully and will make more stakeholders come on board.

5.2 REFERENCES

1. Desai, S. D., Giraddi, S., Verma, N., Gupta, P., & Ramya, S. (2020, September). Breast cancer detection using gan for limited labeled dataset. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 34-39). IEEE.
2. Rakhlin, A., Shvets, A., Iglovikov, V., & Kalinin, A. A. (2018). Deep convolutional neural networks for breast cancer histology image analysis. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15* (pp. 737-744). Springer International Publishing.
3. Mahmood, T., Li, J., Pei, Y., Akhtar, F., Jia, Y., & Khand, Z. H. (2021, July). Breast mass detection and classification using deep convolutional neural networks for radiologist diagnosis assistance. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 1918-1923). IEEE.
4. Sami, H., Sagheer, M., Riaz, K., Mehmood, M. Q., & Zubair, M. (2021, December). Machine learning-based approaches for breast cancer detection in microwave imaging. In *2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)* (pp. 72-73). IEEE.
5. Shen, L. (2017). End-to-end training for whole image breast cancer diagnosis using an all convolutional design. *arXiv preprint arXiv:1711.05775*.
6. Zhang, B., Shi, H., & Wang, H. (2023). Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. *Journal of multidisciplinary healthcare*, 1779-1791.
7. Guo, K., Chen, J., Qiu, T., Guo, S., Luo, T., Chen, T., & Ren, S. (2023). MedGAN: An adaptive GAN approach for medical image generation. *Computers in Biology and Medicine*, *163*, 107119.
8. Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. *Journal of infection and public health*, *13*(9), 1274-1289.
9. Seely, J. M., Eby, P. R., Gordon, P. B., Appavoo, S., & Yaffe, M. J. (2022). Errors in conduct of the CNBSS trials of breast cancer screening observed by research personnel. *Journal of Breast Imaging*, *4*(2), 135-143.
10. Mahapatra, D., Bozorgtabar, B., & Garnavi, R. (2019). Image super-resolution using progressive generative adversarial networks for medical image analysis. *Computerized Medical Imaging and Graphics*, *71*, 30-39.
11. Shen, T., Hao, K., Gou, C., & Wang, F. Y. (2021). Mass image synthesis in mammogram with contextual information based on GANs. *Computer Methods and Programs in Biomedicine*, *202*, 106019.
12. Cai, J., Zhang, Z., Cui, L., Zheng, Y., & Yang, L. (2019). Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network. *Medical image analysis*, *52*, 174-184.
13. Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., & Wang, Z. (2019). Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology. *Engineering*, *5*(1), 156-163.
14. Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. *Journal of infection and public health*, *13*(9), 1274-1289.