Convolutional Neural Networks based Breast Cancer Detection Using Feature Fusion

Doğu Manalı

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical and Electronic Engineering

Eastern Mediterranean University
August 2023
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

Assoc. Prof. Dr. Rasime Uyguroğlu
Chair, Department of Electrical and Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

Prof. Dr. Hasan Demirel
Supervisor

Examining Committee

1. Prof. Dr. Hasan Demirel
2. Prof. Dr. Önsen Toygar
3. Assoc. Prof. Dr. Sertan Serte

ABSTRACT

Currently, one of the most significant health issues affecting women is breast cancer. Breast cancer, which has the highest mortality and morbidity rates among diseases that affect women, poses a severe threat to their lives and health. It is crucial to diagnose breast cancer early. Recently, the technological and theoretical developments in innovative techniques such as machine learning made it possible to achieve early diagnoses of breast cancer. Using Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) for breast cancer detection, we developed a Computer-Aided Diagnosis (CAD) system to identify breast tumors from mammogram images. The proposed system is composed of four stages. Firstly, CNN is used to classify mammogram images. Secondly, CNN-based features are extracted and used with a standard classifier, which is SVM, to identify potential tumors. Thirdly, SVM is used to distinguish between different types of tumors based on the extracted features. Finally, a data fusion method is employed to combine the results obtained from the first, second, and third methods to improve the overall performance of the system. The individual techniques, accuracy is increased, and it reaches around 99% by using fusion methods.

Keywords: Breast Cancer, Convolutional Neural Network, Computer-Aided Diagnosis, Support Vector Machine, Fusion Methods

ÖZ

Günümüzde kadınları etkileyen en önemli sağlık sorunlarından biri meme kanseridir. Kadınları etkileyen hastalıklar arasında mortalite ve morbidite oranları en yüksek olan meme kanseri, kadınların yaşamları ve sağlıkları için ciddi bir tehdit oluşturmaktadır. Bu hastalığı erken teşhis etmek çok önemlidir. Son zamanlarda makine öğrenimi gibi yenilikçi tekniklerdeki teknolojik ve teorik gelişmeler meme kanserinin erken teşhisini mümkün kılmıştır. Bu çalışmada, meme kanseri tespiti için evrişimli sinir ağları ve destek vektör makinesi kullanarak mamogram görüntülerinden meme tümörlerini tespit etmek için bilgisayar destekli bir teşhis sistemi geliştirmeye odaklandık. Önerilen sistemimiz dört adımdan oluşmaktadır. İlk olarak, mamogram görüntülerini sınıflandırmak için Evrişimli Sinir Ağı kullanılır. İkinci olarak, potansiyel tümörleri belirlemek için Evrişimli Sinir Ağı tabanlı öznitelikler çıkarılır ve standart bir sınıflandırıcı olan Destek Vektör Makinesi ile kullanılır. Üçüncüsü, Destek Vektör Makinesi ile çıkarılan özniteliklere dayalı olarak farklı tümör tipleri arasında ayrım yapılır.Son olarak, sistemin genel performansını iyileştirmek için birinci, ikinci ve üçüncü yöntemlerden elde edilen sonuçları birleştirmek için bir veri füzyon yöntemi kullanılmıştır.Füzyon teknikleri sayesinde bireysel tekniklerin doğruluğu artmış ve yaklaşık %99'a ulaşmıştır.

Anahtar Kelimeler: Göğüs Kanseri, Evrişimli Sinir Ağı, Bilgisayar Destekli Tanı, Destek Vektör Makinesi, Füzyon Teknikleri

ACKNOWLEDGMENT

I am incredibly grateful to God Almighty for allowing me to complete my postgraduate thesis successfully. My sincere gratitude goes out to Prof. Dr. Hasan Demirel, who served as my supervisor for the duration of my M.Sc. thesis. In addition, I want to thank the department for its important contributions to education, which were very helpful to me in finishing my research. My family's unwavering moral and financial support of my pursuit of research objectives is something else for which I am grateful. Finally, I want to thank everyone who helped me finish this thesis, whether they were directly involved in it or not.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accuracy
AE	Auto encoders
ANN	Artificial Neural Network
C4.5	Decision tree
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Networks
DDSM	Digital Database for Screening Mammography
DL	Deep Learning
ENVI	Environment for Visualizing Images
FN	False Negative
F1	F-score
FP	False Positive
FPR	False Positive Rate
GANs	Generative adversarial networks
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
K-NN	K-nearest neighbors
LBP	Local Binary Patterns
MACs	Multiply-accumulates
mResnet	Multiview ensemble deep ResNet
MRI	Magnetic Resonance Imaging
NA	Not applicable
NB	Naive Bayes
PRE	Precision
RELU	Rectifier Linear Unit
RF	Random Forest

ROC Receiver Operating Characteristic

ROI Region of Interest

S-DPN Stacks Deep Polynomial Networks

SEN Sensitivity

SGD Stochastic Gradient Descent

SPE Specificity

SVM Support Vector Machine

TN True Negative

TP True Positive

TPR True Positive Rate

WEKA Waikato Environment for Knowledge Analysis

Chapter 1

INTRODUCTION

This chapter gives background information on breast cancer, discusses the problem statement, focuses on the thesis's contribution, lists the thesis's goals, and provides a summary of the thesis's organizational elements.

1.1 Introduction

Breast tissue may experience excessive cell and cell group proliferation form of breast cancer, and these cells may later develop into malignant structures. The result of this uncontrolled proliferation, which frequently affects the mammary ducts or glands, is the formation of masses and other similar structures in the breast. Cancer cells settle and grow in a specific place in the tissues around them, where they quickly continue to grow. Contrary to malignancies in other analogous organs in the abdominal cavity, this tumor is evident on peritoneal biopsy and offers a more favorable environment for breast cancer growth and dissemination [1]. The most prevalent cancer type is breast cancer in women, and it influences all ages of females. According to World Health Organization records, Breast cancer affected over 2 million women globally in 2020, and in total amount of this cancer rise to 25 % in females. This approach that 1 out of each 4 women is diagnosed with cancer and mostly recognized from breast cancer. Women can also develop other cancers, like ovarian and cervical cancer. But none of them are fatal like breast cancer. Breast cancer continues to be a significant global cause of cancer-related deaths in women, so it makes a critical public health issue [2].

Routine screening is recommended for women above a certain age, as early analysis and remedy of breast cancer extensively will increase a female possibility of survival. In order to ensure prompt detection, it is crucial to be fully aware of the common symptoms and to get regular mammograms. We can increase our chances of finding cancer in its early stages by putting a high priority on disease prevention and being vigilant [1]. Modern methods have advanced, allowing, for instance, the use of machine learning for the early detection of breast cancer. Deep Learning, a branch of machine learning that employs multi-layer neural networks to extract and analyze functions from data, has been a major driving force behind this improvement.

Deep learning focuses specifically on the application of multilayer artificial neural networks, also known as deep neural networks, while machine learning refers to a larger set of methods that allow machines to learn from data [3]. SVM is a well-known and popular machine learning algorithm. It is used for classification and regression analysis. To distinguish between different classes of data points in a feature space, which can have many dimensions, SVM determines the best hyperplane.

In this study, the effectiveness of CNN alone and in combination with SVM for the task of image classification is evaluated. To assess the SVM model's performance when used solely, without any pre-processing or feature extraction from a CNN model. It is crucial to understand, though, that each model may have certain limitations on how accurately it can classify data. As a result, we also look into the possibility of enhancing overall performance by combining the CNN and SVM model outputs. The goal of this fusion approach is to take advantage of the distinct advantages of each model: SVMs offer reliable classification boundaries, while

CNNs excel at learning hierarchical features from images. By using these fusion strategies, we can improve the precision and robustness of our image classification system by combining the strengths of CNNs and SVMs.

The results of this study will advance computer vision by shedding light on how CNNs and SVMs perform differently on different image classification tasks. Additionally, investigating model fusion techniques will provide helpful direction for creating more efficient and precise image classification systems across various domains and applications.

1.2 Problem Definition

Breast cancer represents a substantial public health issue that persists as a prominent contributor to cancer-induced mortality in the female population worldwide. Breast cancer constitutes a form of malignancy that is frequently insidious and not easily discernible, thereby posing challenges to timely detection. Women categorized as having a heightened susceptibility to breast cancer, specifically those with a familial background of the ailment, may necessitate more customized and frequent screening protocols in order to achieve optimal identification and supervision of prospective malignancies. The importance of routine mammography screening in the timely detection of breast cancer, a period during which medical intervention is most effective, is widely recognized in academic discourse. However, interpreting mammography images can be difficult, especially for radiologists with little training. In addition, many areas, especially rural and underserved areas, lack radiologists, which further complicates early detection [4].

For instance, according to the Ministry of Health [5], Ethiopia boasts a considerable number of hospitals, health centers, and health posts within its healthcare infrastructure. However, a considerable number of these establishments are currently without medical practitioners. The proportion of individuals to medical practitioners is significantly inadequate, with an average of one physician serving a populace of 57,876 individuals, as conveyed by reference [6]. In the southwestern and west-central regions of Ethiopia, the ratio of doctors to population is notably deficient, with a solitary physician available for every 200,000 to 300,000 individuals[7].

In order to overcome these hurdles, the current study strives to develop a computational aid for diagnosis that accurately detects and categorizes instances of breast cancer captured through mammography imaging. The contemporary diagnostic approach embraces advanced machine learning techniques, which integrate SVMs and CNNs. The aim of this investigation is to elevate the degree of precision and effectiveness in identifying and determining breast cancer, focusing specifically on areas where the accessibility of radiologists is limited. The primary objective of CAD system advancement is to address the difficulty of discriminating between noncancerous and cancerous tumors while simultaneously refining the accuracy of breast cancer identification in areas of dense breast tissue. A comprehensive analysis shall be undertaken to ascertain the efficacy of the system via methodically conducted testing and validation procedures. Sensitivity, specificity, accuracy, and an examination of the characteristics of the receiver operating characteristic (ROC) curve are the variables being evaluated.

1.3 Thesis Contributions

Mammography is a crucial imaging technique for detecting and diagnosing breast cancer, but it has some drawbacks. One of the key concerns related to mammography pertains to the inadequacy of radiologists who possess the necessary expertise in the interpretation of mammography images. This shortage of radiologists has significantly impeded the prompt dissemination of mammogram results, leading to a delay in diagnostic procedures and the initiation of treatment for patients. Furthermore, the complex and intricate nature of mammography images poses a significant challenge for radiologists to accurately and correctly interpret them. The complexity and nuanced characteristics of mammography images pose a significant challenge in the timely detection of early symptoms of breast cancer. Moreover, the interpretation of these images is subject to several variables, including breast density, image quality, and the level of expertise and education of the radiologist, which can potentially affect the efficacy of mammography evaluations [8]. The employment of CAD systems that utilize machine learning has emerged as a significant development within the domain of medical image analysis, particularly in the interpretation and reading of mammography images. CNNs are a prominent and esteemed deep learning model that is extensively utilized for image classification tasks because of their exceptional precision and effectiveness.

SVMs are a different kind of machine learning algorithm in addition to CNNs that can be applied in CAD systems to label mammography images. Radiologists can review and validate the results more quickly and effectively by using CAD software to help automate the interpretation of mammography images.

Healthcare industry problems may be solved by computer vision technologies, including machine learning.supporting doctors or other medical professionals in particular when the availability of such experts is limited. The ability to diagnose cancer more quickly and increase survival chances depends on early detection and timely prediction. In medical settings, the study's findings can be applied to assess women's breast cancer risk while they see their mammograms. Finding and diagnosing breast cancer early can save lives because women are so important to their families. Since early breast cancer can be treated more successfully and more easily, early detection is crucial. As a result, it is recommended that women start screening for breast cancer at the age of 40 and continue to do so annually [9]. The contributions of thesis are listed below:

- Integration of Fusion Methods: To combine the findings from SVM and CNN models, the thesis introduces and assesses three distinct fusion methods. With this integration, image classification tasks can now be completed with greater accuracy.
- Accuracy Improvement through Fusion: The thesis shows that using fusion techniques significantly improves image detection accuracy. This result emphasizes how well SVM and CNN outputs can be combined to enhance classification performance.
- Determination of the Best Fusion Method: The thesis determines which of the three suggested techniques is the Best Fusion Method through comparative analysis and evaluation. This information advances our understanding of the best fusion methods for fusing SVM and CNN models.
- Development in Image Classification: By introducing a fresh strategy that combines SVM and CNN models, the thesis advances the field of image classification. The improved accuracy that has been seen and the discovery of the best fusion methods have improved image classification methodologies.

- Real-World Applications: The research presented in this thesis has applications in autonomous driving, medical diagnosis, and surveillance systems. The increased accuracy made possible by fusion techniques improves the dependability and efficiency of these systems across a variety of domains.

1.4 Thesis of Objective

There are two objectives of the study: general and specific objectives. The objectives, details are given below.

1.4.1 General Objective

The main goal of this study is to improve breast cancer detection accuracy and efficiency by creating a model that combines three different approaches, including CNNs and SVMs, and using data fusion to enhance system performance.

1.4.2 Specific Objectives

The study aims to achieve the following specific objectives in order to reach its main objective;

- To perform an extensive review of literature to identify the techniques, algorithms, and methods used in this research.
- The mammogram images are labeled by domain experts during the data collection process.
- After collecting mammogram images, to produce a dataset for training and testing the breast cancer detection model.
- To indicate appropriate machine learning models for supervised learning.
- To propose a highly effective model for detecting and classifying breast cancer.
- To evaluate and validate the effectiveness of the developed model in detecting and classifying breast cancer.

1.5 Thesis Outline

The present thesis is structured as follows in the subsequent sections. The following chapter shall place emphasis on the execution of a comprehensive literature review, followed by the consolidation of pertinent literature. This study seeks to present a succinct analysis of pertinent literature concerning, computer vision, breast cancer disease, machine learning, and deep learning principles. Within this section, an overview of pertinent literature shall be presented, emphasizing the contributions of various scholars whose work is deemed of utmost significance and bears a strong correlation to the identification of breast and other forms of cancer through the implementation of deep learning methodologies. Subsequently, Chapter 3 shall expound upon the research methodologies utilized in the present investigation. Chapter 4 will provide a comprehensive account of the experiment conducted. Moreover, Chapter 5 will expound upon the findings and present an in-depth analysis of the conducted investigation. Subsequently, the sixth chapter shall culminate the investigation by consolidating its discoveries, merits, and presenting suggestions.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

The chapter surveys breast cancer detection methods and works, providing an introduction to abnormalities and medical screening techniques.

2.1 Introduction

The focus of this chapter is a thorough analysis of the literature that is relevant to the dissertation's main topic. Using various sets of images and datasets, we give a succinct and instructive overview of breast cancer classification in this review. Machine learning techniques with a variety of architectures have been used to publish and improve numerous methodologies for the detection and classification of breast cancer. The classification of mammography images based on cell structure poses a formidable challenge in the field of medical image analysis and classification. Contemporary techniques have been employed to address the arduous task of image categorization, encompassing deep learning architectures, trained deep neural networks, and machine learning methodologies. The paper also covers a variety of machine learning techniques used for breast cancer classification. These strategies extract significant features from mammography images and categorize them using a variety of algorithms, including ensemble techniques and SVMs.

Overall, the literature analysis offers a thorough grasp of the ongoing investigations into the classification of breast cancer. It clarifies the developments achieved in the area, particularly the use of trained deep neural networks, deep learning architectures, and machine learning techniques. The review establishes the groundwork for the dissertation's chapters to follow, which suggest novel approaches and make contributions to the field of breast cancer classification utilizing image analysis and machine learning methods.

2.2 Breast Cancer Overview

Previously, the illness was generally regarded a societal taboo and source of embarrassment, resulting in rare instances of detection and diagnosis. The incidence of breast cancer references in literary works beyond the domain of medical publications and books was unusual. The active engagement of women and the promotion of openness around sickness have arisen as relatively new phenomena, having gained ground during the past few decades. In the 1990s, the pink ribbon symbolizing breast cancer triggered a revolution in the fight against this disease[10]. Uncontrollably growing and spreading abnormal cells can give rise to tumors [11]. Tumors that have spread outside of their original location and are currently invading nearby healthy tissue are referred to as invasive tumors. Breast cancer in women is the most common cancer type. Doctors have been aware of breast cancer since the turn of the century. Because of the visible symptoms, particularly in later stages, nearly all of recorded history that is still available mentions it [10].

The human body's exocrine glands, which also include other glands that produce breast milk, include the breast. The majority of its structure is composed of numerous tiny pieces called lobes and fat. Each lobe contains a large number of alveoli, or cavities. These cavities are surrounded by my epithelial cells, which link them to the epithelial cells that secrete milk. Milk is drawn from the lobes into the tiny, duct-like tubes created by my epithelial cells. A network of pipes is connected to one another. The nipple marks the end of the larger pipe that is formed when these pipes are joined [12]. The density, mass, or calcification of a breast on a mammogram is indicated based on the area of the breast anatomy that is affected.

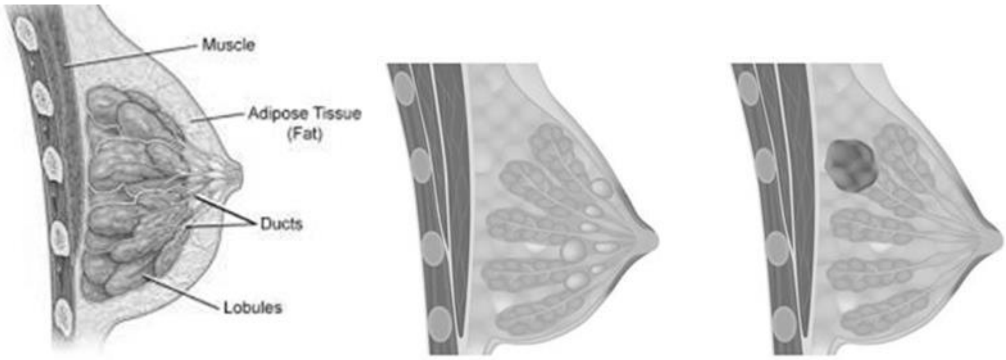


Figure 1: From left to right,the anatomy of the breasts in cases of typicality, mass, and calcification [13].

2.3 Breast Abnormality

Breast tissues may change structurally or morphologically in a number of ways, including physical distortions or the presence of strange lumps. The following categories idealize these lumps mass and calcification.

2.3.1 Calcification

Since calcium is unrelated to an individual's diet, calcium deposits in dense tissue lead to calcification, which may be a sign of an underlying tissue process. As people get older or when milk glands release milk into the ducts, this happens more frequently as shown in Figure 1.

Calcification is not always thought to be cancerous. There are two possibilities:

- Benign/not cancerous
- Malignant/cancerous.

In general, calcification manifests itself in the breast tissue as a few minuscule, white, clustered spots. When calcification is present, it can take on a variety of shapes, such as spheres, popcorn-like shapes, or lines that range in size from 0.1 to 1.0 mm. The larger the lines, the less likely it is that they are malignant[14-16].

2.3.2 Mass

A mass or lump is a region of dense tissue in the breast with edges and distinct shapes that is distinct from the surrounding healthy tissues, as previously seen in Figure 1. When we compare malignant and benign cases, malignant cases have less masses of calcification than benign cases.

2.4 Imaging Techniques for Breasts

Breast cancer patient survival and life expectancy have been shown to be significantly improved and extended by early detection [1]. Imaging modalities are required to support radiologists in the detection of potential breast abnormalities. In the medical field, a number of screening techniques are available for that purpose, including:

- MRI
- X-ray
- Mammography.

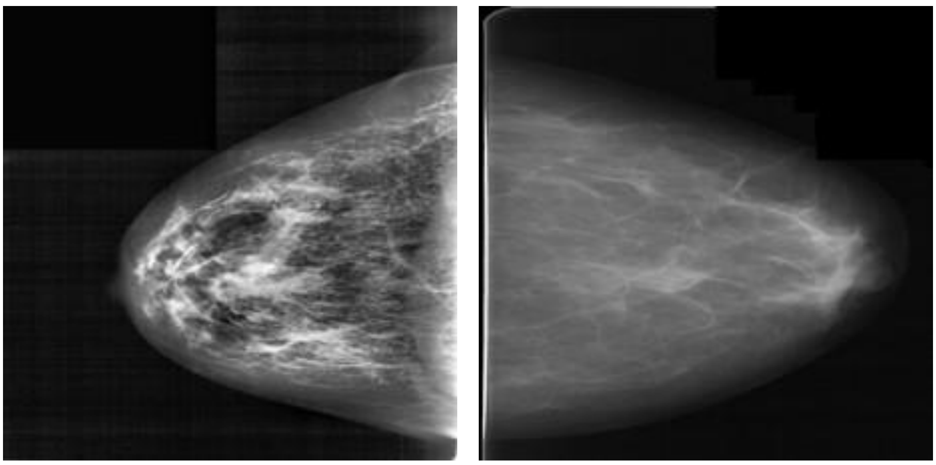


Figure 2: The left mammogram is an example of calcification, the other an example of a mass[13].

The most effective method for examining human breasts for any abnormalities is to use low-dose x-rays to view the internal structure of the breast tissues during mammography [8,9]. Figure 2 shows the mammogram samples of mass and calcification.

2.4.1 Analysis of Breast Images

To efficiently use the amount of data, explore it, and display it for specific medical tasks, a medical image is required. The development of machine learning algorithms has improved this field's accuracy and garnered a lot of interest. In radiology, types and intensities of anomalies in X-ray images are identified, and whether an anomaly is normal or abnormal is determined using the well-known medical image analysis. Even though they frequently yield false positive results, CAD systems were developed and put into use in the healthcare system in the 1980s [17]. Medical image processing, in contrast to conventional image processing, only seeks to improve the readability of the content generated, with no other primary goals (such as enhancing image aesthetics or creative art). The image itself may need to be improved in order to improve the perception of specific features and automatically extract information [17]. One of the radiologists' most important tasks is to make a precise differential diagnosis based on the medical images of each patient. Determining the type of cancer or whether a disease is present could be included in this [18]. For doctors, examining the medical image is a difficult and time-consuming task. In order to streamline the process and free up the doctor's time for other duties, the CAD system has been used to evaluate patients since 1980 [17]. However, the CAD system in use at the time produced more false positives than doctors who extended examination times and insisted on pointless examinations after a biopsy [18]. Methods for using medical images to enhance image capture, disease detection, therapy, and prediction have changed and improved as a result of advancements in computer vision.

2.5 Machine Learning

For systems to learn on their own, algorithms are developed in the area of artificial intelligence known as machine learning. Machine learning's core idea is the creation of algorithms that can analyze input data to find patterns and features, forecast an output, and update outputs as new data becomes available. The three different types of algorithms that make up machine learning are supervised learning, unsupervised learning, and reinforcement learning [19].

In supervised learning, the algorithm is trained on sets of data with exact labels and training examples to get it ready to decide what to do when fresh data is presented. Unsupervised learning, in contrast, does not tag the data. The program must identify patterns and relationships in unlabeled and uncategorized data without any prior training. In reinforcement learning, the environment in which the system continually practiced making decisions by making mistakes is exposed to it. It makes the most accurate predictions it is able to using historical knowledge [20,21]. Many typical activities, including prediction, medical diagnosis, picture and speech recognition, and others, are performed using machine learning.

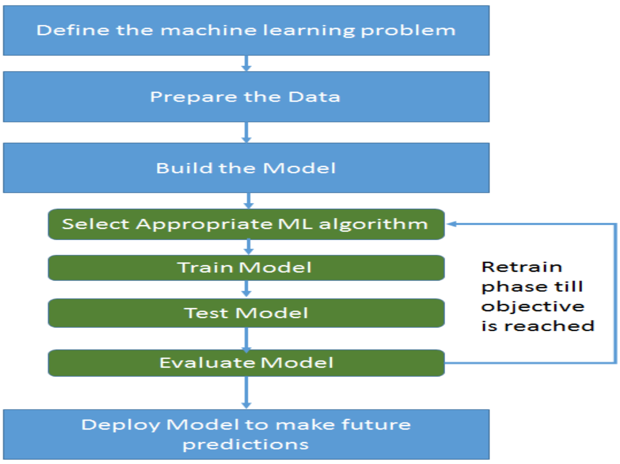


Figure 3: A description of the steps in machine learning [22].

2.5.1 Support Vector Machine

A supervised classification technique called SVM typically produces accurate classification results from challenging and noisy data. Its foundation is statistical learning theory. It uses a decision surface to divide the classes that maximizes their margin. The hyperplane is often referred to as the best hyperplane since support vectors are the data points that are closest to it. Support vectors are the most crucial subset of the training set. By incorporating nonlinear kernels into SVM, a nonlinear classifier can be created. There are four different kinds of kernels. Basis functions that are linear, polynomial, sigmoid, and radial basis function. SVM is a binary classifier in its simplest form, but it can be combined with other binary classifiers to create a multiclass classifier (a binary classifier is created for every potential pair of classes) [23]. Environment for Visualizing Images (ENVI) SVM implementation uses pairwise classification as its multiclass classification technique.

Table 1: Representation of each kernel in mathematics.

Linear	$K(x_i,x_j) = x_i^T x_j$
Polynomial	$K(x_i,x_j) = (gx_i^T x_j + r)^d$, $g > 0$
Sigmoid	$K(x_i,x_j) = \exp(-g\|\|x_i - x_j\|\|^2)$, $g > 0$
Radial basis function	$K(x_i,x_j) = \tanh(gx_i^T x_j + r)$

In Table 1, X_j and X_i^T represent two input vectors. G stands for the gamma term in the kernel function for all kernel types other than linear. The polynomial degree term is represented by d in the kernel function for the polynomial kernel. R is the bias term in the kernel function for the polynomial and sigmoid kernels.

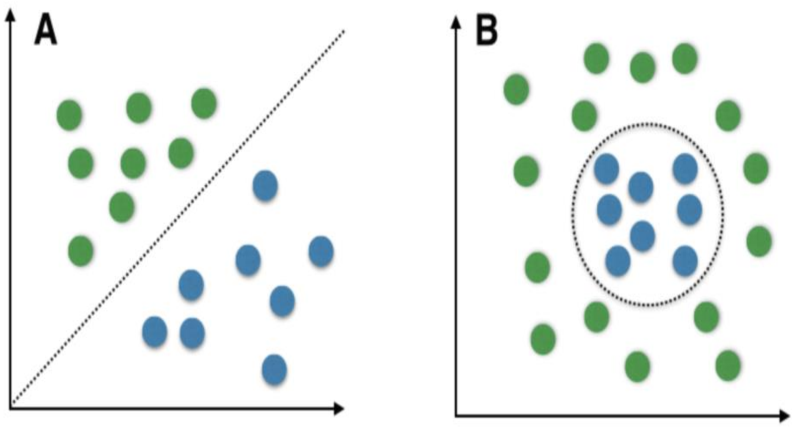


Figure 4: The image A is an example of linearly separable data, and the image B is an example of non-linearly separable data [24].

2.6 Deep Learning

In conventional machine learning methods, which call for a good representation of the input data, it is crucial to have a good set of characteristics that will provide a feature vector that can be used by algorithms as a learning bias. By learning the answer directly from the data, deep learning's main objective is to solve this problem [25]. The neural network gradually learns more about the images as it advances through each layer. Curves, lines, and edges are basic features that early layers detect; later layers combine these features to produce a more meaningful representation.

For simpler problems, they might not be able to generalize new data because they require a large amount of data to function well. Deep learning algorithms are still better at handling images even though they can automatically identify features in images.Convolutional neural networks are a key component of almost all deep learning techniques. Other approaches, like generative adversarial networks (GANs) and autoencoders (AE) are also excellent at addressing various machine learning problems.

2.6.1 Convolutional Neural Networks

Convolutional neural networks have recently been successfully applied to and widely adopted in image-focused tasks due to their capacity to encode image-specific features in the architecture while requiring less setup data for a model. It is challenging to use precise mathematical descriptors to describe the feature technique because medical images differ from other types of images in their local structures, such as texture, edges, curves, and corners. As a result, over the past ten years, CNN applications have significantly increased in the medical sector. CNNs are composed of neurons that are capable of optimizing their own operations, much like traditional neural networks. By ingesting inputs (like a scalar value followed by a nonlinear function), the neurons process information. The whole network expresses a single evaluation function (weights), from the vector image to the final output of the class. Convolutional neural networks' top layer includes class-specific loss functions as well as all the construction instructions for a typical neural network.

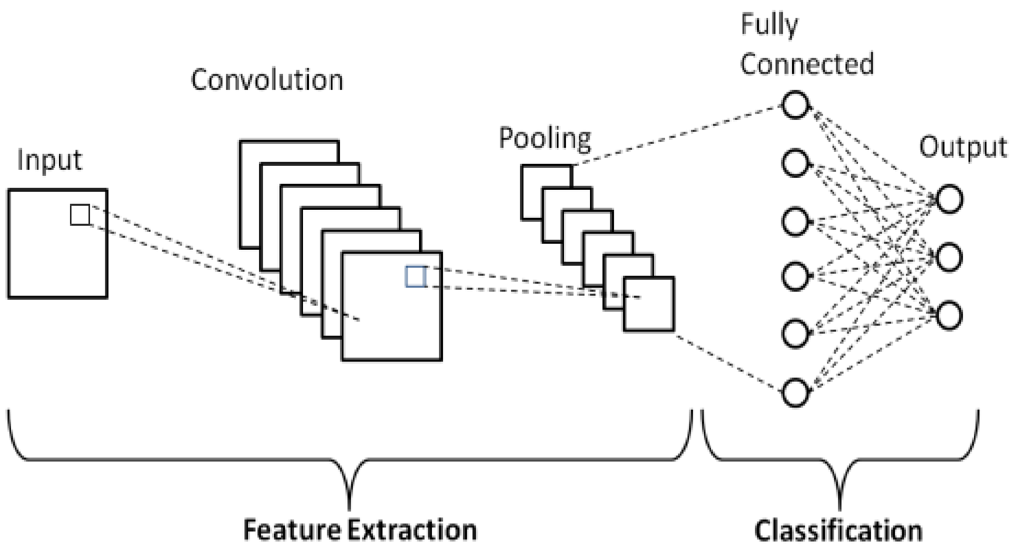


Figure 5: Convolutional Neural Network concept [26].

Convolutional layers, pooling (downsampling) layers, activation feature layers, and fully connected layers are the fundamental CNN building blocks [25].One, two, or even three dimensions may be present in the input data from these networks. One, two, or even three dimensions may be present in the input data from these networks. Convolution is the process of applying m x m filters to the entire input picture matrix in the convolution layer. The term "convolution" in this context describes a linear operation that multiplies the image matrix and filter weights in a specific manner; each weight is a model parameter that must be learned.

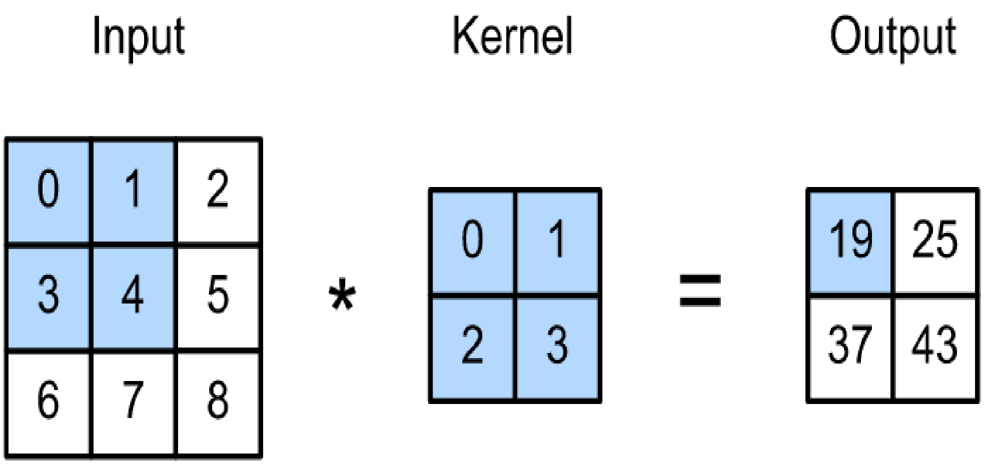


Figure 6: Example of convolution process [27].

After convolution, each filter generates a feature map, which is an affine transformation of the input used as the input for the layer as follows. Artificial neural networks' activation features are incredibly helpful. Based on these characteristics, a neuron is either activated or not. Whether the information the neuron receives should be ignored or is pertinent to the information provided.

The nonlinear transformation we perform on the input pixels is referred to as the activation function. The layer of neurons below receives this altered output as input. The sigmoid function, rectified linear unit (ReLu) function, and softmax function are frequently used in convolutional neural networks as activation functions. Usually, the convolutional layer is followed by the pooling layer. By down sampling each of the feature maps created by the previous convolutional layer to create a new set of the same number of pooled feature maps, the layer aims to reduce the spatial dimensionality of the data. Average pooling and maximum pooling are the two main functions used in pooling operations [28].

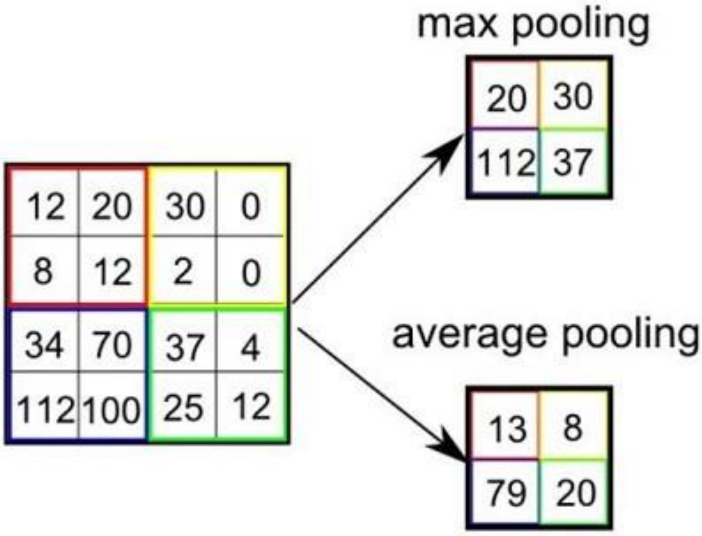


Figure 7: Max and average pooling examples [29].

The network becomes slightly more translation invariant as the spatial dimensionality is decreased, which also reduces the amount of training data required.

After a large number of convolutional and pooling layers, fully connected layers are used. The high density of neurons in these layers flattens the feature maps from the earlier layers into a single vector.. This flattened vector can be used with fully connected layers for classification.

$$Z = activationfunction(\sum_i(wi * xi + bias)) \quad (1)$$

From the above formula is used in fully connected layer and we can summarize the process as follows: to determine the output of the neuron, the formula computes the weighted total of the inputs, including the bias term, and then passes it through an activation function. Each neuron in the network goes through this process once, allowing the network to learn and develop predictions based on the inputs and learned weights.

2.6.1.1 Convolutional Neural Networks Parameters

Without knowledge of some of the hyper-parameters used to modify training and learning to address the specific challenge, understanding the general architecture of CNN will not be sufficient to produce a successful model.

A statistical model is said to be overfitted when it becomes trained on data and starts to learn unreliable and noisy dataset entries. Overfitting is a significant challenge in deep learning and machine learning because an overfitted model is unable to generalize new data during model testing [30]. Therefore, a test set is necessary in order to assess a deep learning model's performance. All of the strategies on the following list have been proposed to lessen overfitting.

➢ More training data: By making more training data available, overfitting can be reduced. With more diverse examples, the model is less likely to memorize specific examples or noise in the data and more likely to find patterns that can be applied to novel situations.

- Data augmentation: In order to generate new samples, data augmentation involves applying various transformations or perturbations to the training data already present. Changing an image's orientation, size, or noise, for instance. The training dataset's size and diversity are artificially increased, which can expose the model to more variations and prevent overfitting.
- Regularization: Regularization techniques like L1 and L2 regularization introduce a penalty term to the loss function during training. This penalty discourages the model from learning complex patterns or relying too much on specific features, which promotes simpler and more general models.
- Batch normalization: By normalizing them with respect to the mean and standard deviation of the most recent mini-batch, the batch normalization technique normalizes the activations of each layer in a neural network. This can lessen overfitting and speed up training by helping to regularize and stabilize the learning process.
- Decrease architecture complexity: Simplifying the model architecture can reduce overfitting. Limiting the number of layers or nodes in a neural network or the maximum depth of a decision tree are two examples. A simpler model can concentrate on the most significant and generalizable patterns while being less likely to memorize noise in the data.

Determining the learning process' optimization algorithm is a crucial additional parameter. One of the favored algorithms is stochastic gradient descent, which merely updates the parameter for each training run and descends until it finds the best option parameter.

The loss function, which aims to reduce the discrepancy between the model's output and the desired outcome, should then be established.

2.6.2 Convolutional Neural Networks Architectures

Over the past ten years, various CNN architectures have been introduced [31,32]. Model architecture is essential for enhancing the performance of numerous applications. The structural organization of CNN has undergone significant change since 1989. The redesign of the processing units and the addition of new blocks, however, is something that must be emphasized because it significantly improved CNN's performance. Such modifications include, for instance, regularization, parameter optimization, and structural reformulation. The most inventive developments in CNN architectures have been in the use of network depth. In this section, the residual network that triumphed in the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is examined[32].

ResNet employs the concept of a bypass pathway in a novel way.In contrast to earlier networks, they wished to construct an ultra-deep network free of vanishing gradient problem. Layers are recast in order to learn residual functions with refence to the layer inputs rather than employing the conventional method. Depending on the number of layers (34 to 1202), different versions of ResNet were created. ResNet50, which has 49 convolutional layers and one FC layer, was the most prevalent type.

While there were 3.9 M total Multiply-Accumulates (MACs), there were 25.5 M network weights overall.

2.7 Ensemble Learning Technique

Neural network modeling is one nonlinear method. They are flexible and can detect intricate nonlinear relationships in data. This adaptability has a disadvantage because of their sensitivity to the initial conditions, such as the initial random weights and the statistical noise in the training dataset. Each time a network is trained, it learns a (slightly) updated version of the functions that convert input to network output. This will lead to variations in predicted performance even with the same dataset. Several models are trained on a single problem using ensemble learning, and the outcomes of their forecasts are combined to create a single, precise forecast [29].There are numerous configurations for ensemble networks, including the bagging method, voting method, boosting method, and stacking method. According to [28], the bagging method combines independent, parallel learning through an averaging process.

The methods mentioned above operate in accordance with a set of combination rules that are already built into them, such as the mean rule, sum rule, product rule, weighted majority voting, etc. One of the many available rules can be used to combine ensemble classifiers based on various models, each trained with a different subset of training data.

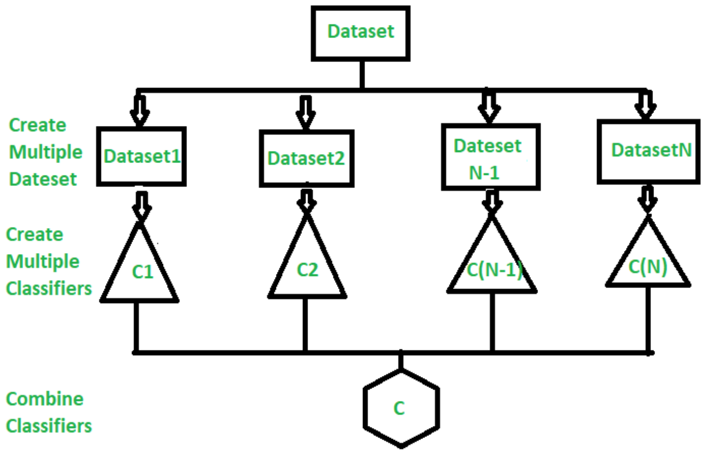


Figure 8: Example for ensemble learning [34]

2.8 Literature Review

This chapter gives a summary of the pertinent studies that have been done and the methods that are employed at each stage of breast cancer detection.

2.8.1 Classical methods

One of the most fundamental and significant tasks in machine learning is prediction [35]. In particular, for the prediction of breast cancer, machine learning algorithms have been extensively used in research on a variety of medical datasets. Almost all machine learning (ML) techniques showed respectable prediction accuracy [36].

SVM, Random Forest (RF), Naive Bayes (NB), and K-Nearest Neighbors (K-NN), four machine learning algorithms, were used by Khourdifi [37] to the Wisconsin breast cancer data is available in the machine learning library at UC Irvine. To simulate the method, the authors used the Waikato Environment for Knowledge Analysis (Weka) tool. They discovered that SVM performed the best overall in terms of effectiveness and efficiency.

An artificial neural network (ANN) and SVM were used by Boeri et al. [38] in 2020 to forecast the likelihood that a patient will pass away within 32 months after having breast cancer surgery. SVM had the best performance, with an accuracy of 96.46%.

Another research illustrate which is Asri et al. [39] compared SVM, decision tree (C4.5), NB, and K-NN among other machine learning approaches in the Wisconsin Breast Cancer (original) datasets for breast cancer risk prediction and diagnosis. The experimental SVM fared best in terms of accuracy and error rate.

A different study was conducted by Zheng et al[40].The proposed method achieved 97.2% accuracy and improved feature representation by stacking Deep Polynomial Networks (S-DPN) algorithms. It also developed a two-stage multi-modality S-DPN technique in conjunction with an SVM classifier to produce more discriminate, robust features.

2.8.2 Deep Learning Methods

Breast cancer early detection relies heavily on mammographic density. The radiologists' scores differ greatly, and the qualitative evaluation is very arbitrary [41]. Additionally, according to recent studies, commercial software used to calculate breast scores frequently produces results that are inconsistent with radiologists' assessments [42] by either overreporting or underreporting. Deep learning (DL) algorithms can significantly lessen the amount of time radiologists must spend manually scoring data and enhance their ability to conduct risk assessments [43].

In a related study, Xu et al [44] used residual CNN to categorize the breast density estimation method. Both single-view and multiple-view images responded favorably to the method. They divided the BI-RADS densities into 4 groups in their study using the residual CNN. Seven residual learning blocks were distributed over 70 layers in the residual CNN. To compare the performance, two additional networks with 36 and 48 weighted layers but fewer residual blocks were also trained. Cross-entropy loss could be decreased while maintaining classification accuracy with ResNets. Their research revealed that classification accuracy rose as the residual layer was raised. The price of computation, however, went up.

Dhungel et al. introduced a multiview ensemble deep ResNet (mResNet) for identifying malignant and benign tumors in a different work [45]. They used a deep

ResNet component in their ensemble network that could handle six input images with various viewpoints, including cranio caudal and mediolateral oblique views. The lesions' binary maps can be automatically created by the mResNet. The combined output of the mResNet can be used to build a fully linked layer that can categorize the lesions into benign or malignant groups.

Deep convolutional neural networks were employed by Geras et al. [46] to forecast breast densities in multiview data. The technique identified three different breast density types: BI-RADS0, BI-RADS1, and BI-RADS2. Additionally, it used the region of interest (ROIs) extracted from these images to distinguish between benign and malignant anomalies. The study also looked at how training size and image size affected the accuracy of predictions. More training samples were found to increase the testing phase's prediction accuracy. Furthermore, changing the image size had little to no impact on the method's ability to generate accurate predictions. The outcomes are generally in line with manual scores generated by seasoned radiologists.

Jiao et al.'s study [47] creates an eight-layered network with five convolutional layers and three full-connection layers. The model is pretrained on ImageNet to get around the issue of the lack of medical examples. SVM then executes the classifier, and a decision mechanism is presented. Then, in order to categorize lesions, the mammographic breast cancer diagnosis model applies 2048 high-level features and 256 mid-level features.

Chapter 3

METHODOLOGY

The chapter covers the methods used for decision fusion techniques and feature extraction from images. Details about the study's toolkit and platform are also covered. Moreover, the specifics of the database used for research.

3.1 Introduction

By creating index measures that enable precise quantification, it is possible to determine whether it is feasible to measure a system's performance subjectively or objectively [48]. Even for straightforward tasks like binary classification, a variety of indices must be used to accurately assess classification performance. The instances of a set are categorized using binary classification based on whether they belong to a class or not. The classification system will produce four results: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) [49]. False positives will be correctly categorized as false negatives, true positives as true positives, and false negatives as true negatives for negative classes. If the prediction set is unbalanced, with more occurrences of one class than the other, or balanced, with an equal number of positive and negative predictions, there may be a discrepancy between absolute and relative forecasts. Due to all of these considerations, it is challenging to develop a distinctive metric for assessing prediction skill. This explains why binary classification employs a variety of indexes. The list that follows contains a summary of some of the most well-known indexes for binary classification.

3.2 Feature Extraction

A feature extractor is a part or algorithm that turns a set of pertinent and representative features from raw data. These traits are used to identify significant patterns or traits that are pertinent to a particular task or analysis. They are a clearer, more streamlined, and more informative version of the original data. The pretrained CNN and Local Binary Pattern (LBP) features are utilized in the technique in this thesis proposed.

3.2.1 Local Binary Patterns (LBP)

Local Binary Patterns (LBP), a technique for extracting texture features, is one of the straightforward and effective methods [51]. Each pixel was used by LBP as a threshold before its 3 x 3 neighborhood was converted into an 8-bit binary code.

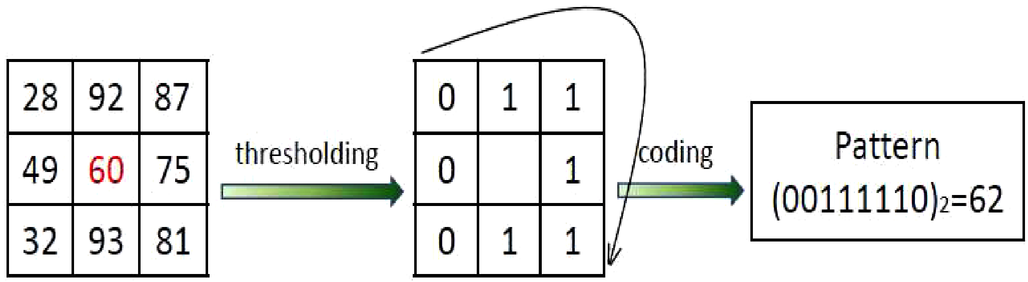


Figure 9: LBP Operator [52].

The texture direction information surrounding pixels is reserved by the fixed order of this binary code. There are 2P different ways to do this. When there are only two instances of either 0 to 1 or 1 to 0, the binary pattern is known as uniform LBP. P2 - P + 2 is the number of uniform patterns in a sampling density P.

3.2.2 Pretrained CNN Extractor

In the real world, machine learning applications frequently use small amounts of training data. Training a deep network from scratch can be difficult and time-consuming. Due to the complexity and sizable number of model parameters, this is the case. However, a model from the previous section that was developed using a sizable dataset, such as ResNet, can be applied once more to issues unrelated to the original categorization issue [25]. A pre-trained CNN such as AlexNet and ResNet-50 can be used as feature extractor. The last layer is omitted, which is the classification layer of the pre trained CNN model, and the remaining layers are used as a feature extractor. Then these features were passed to a new classifier, such as SVM, to classify the images.

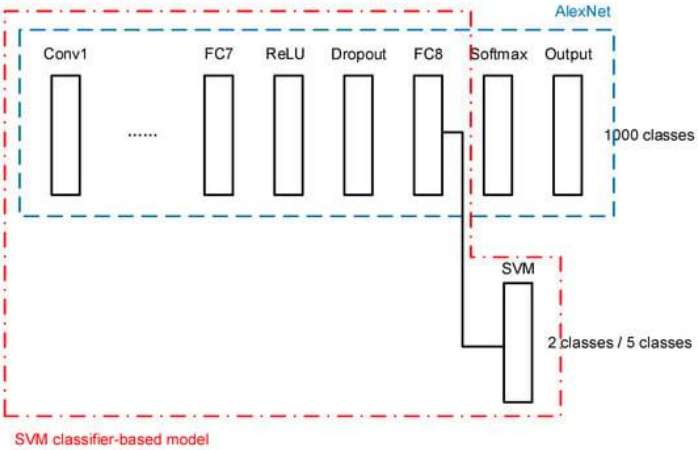


Figure 10: Example for pretrained CNN model for feature extraction [33].

3.3 Decision Fusion

When using decision fusion, the decision fusion block receives output probabilities from various methodologies, and at the conclusion of the process, a diagnosis of breast cancer is made. The sum rule, product rule and majority voting methods are used in our thesis' decision-making process to determine whether a breast cancer is obvious or not. Equation 2 demonstrates the sum rule, and equation 3 provides an example of the product rule's formula. Finally, equation 4 illustrates the majority voting method.

Sum Rule $$\hat{y} = mean\{c1(x), c2(x), \dots \dots \dots . cm(x)\} \quad (2)$$

Product Rule $\hat{y} = \{c1(x) * c2(x), \dots \dots \dots * cm(x)\}$ (3)

Majority Vote Rule $\hat{y} = mode\{c1(x), c2(x), \dots \dots \dots . cm(x)\}$ (4)

Where;

$\hat{y}$= Final prediction

c1 (x), c2 (x), … … . . cm (x) = predictions from classifier

3.4 Software and Hardware Platform

In order to choose the best software tool for breast cancer detection from mammography images using the CNN algorithm, all currently offered software products and their libraries are evaluated. The analysis showed that while SVM is specific and can only be used with one of the algorithms, some tools, like Matlab, are general and can be used with both machine learning and deep learning algorithms. Before selecting the tools, we took into consideration the following factors because they help us find the appropriate software tools and companion libraries. The algorithm's primary determining factor is the computer language used. The second step is to select tools with adequate learning resources, such as free video tutorials and prior knowledge. The use of the tools on computers with limited resources is the third prerequisite. The programming language used was Matlab 22. Windows 10 is used to run Matlab software. A number of toolboxes are uploaded to run the proposed system there, including deep learning and machine learning toolboxes. The hardware specifications are like this: 8 GB of NVIDIA GTX 1660i, 16 GB of random access memory, and a 250GB solid state drive are used.

3.5 Database of Breast Cancer Images

The data needed to train the model must be collected if machine learning algorithms are to be used in this study. Data from mammography imaging for breast cancer is the main source of input for the model in this thesis. For study and research, there are a number of well-known mammography datasets that are openly accessible. The digital database for screening mammography (DDSM) is one of the most commonly used datasets in machine learning. The DDSM dataset is well recognized and extensively used in the field of breast cancer research. It includes a collection of mammograms that have been digitally archived together with the pertinent metadata, notes, and ground truth labels.The DDSM dataset was established by the Department of Radiology at the University of South Florida to be used in the creation and assessment of CAD algorithms for breast cancer. Included are mammograms taken during screening and diagnostic procedures that show various breast abnormalities. The training and test portions of the dataset are initially separated. After the model has been trained using the training split, it is tested using the test split, which is not used during model training. The validation split is used to evaluate the performance of the model created during training and to change model parameters in order to choose the model that performs the best overall. To prevent the problem of overfitting, an equal number of images from each class must be used for both validation and training.

Chapter 4

BREAST CANCER DETECTION BASED ON MULTI-MODAL FRAMEWORK

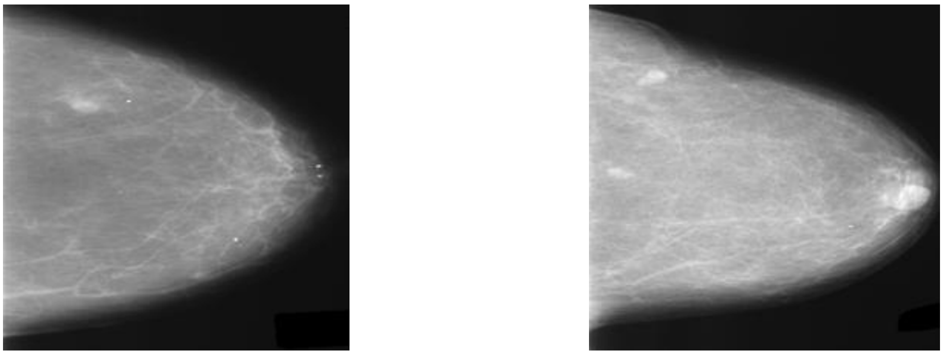
Chapter discusses multi-modal framework implementation, focusing on system overview, model training, decision fusion, and preprocessing pipeline.

4.1 Introduction

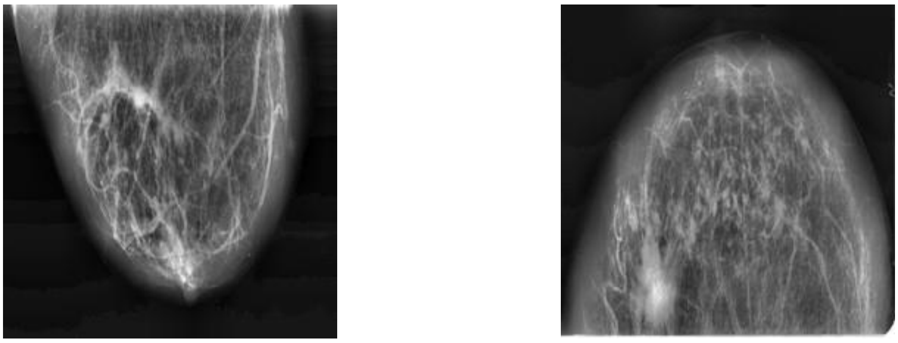
This chapter covers the specifics of the suggested multi-modal framework and how it was put into practice. It is divided into five sections, the first of which gives an overview of the entire system. Next, model training, decision fusion, and the preprocessing pipeline are introduced.

4.2 Dataset Description

There are many free databases with top-notch hand annotations available. Based on the following factors, we selected the DDSM dataset for this study from [53]: The database was accessible to the general public, the images were rated as suitable for medical analysis (gradable) or not (ungradable), and expert medical professionals handled the database annotation. To facilitate research on computer-aided breast cancer diagnostics, this database was developed. The database contains 434 grayscale pictures altogether. The dataset is divided into two categories: benign and malignant. There are 217 images per category.Figure 11 shows images for each class from dataset.



(a)

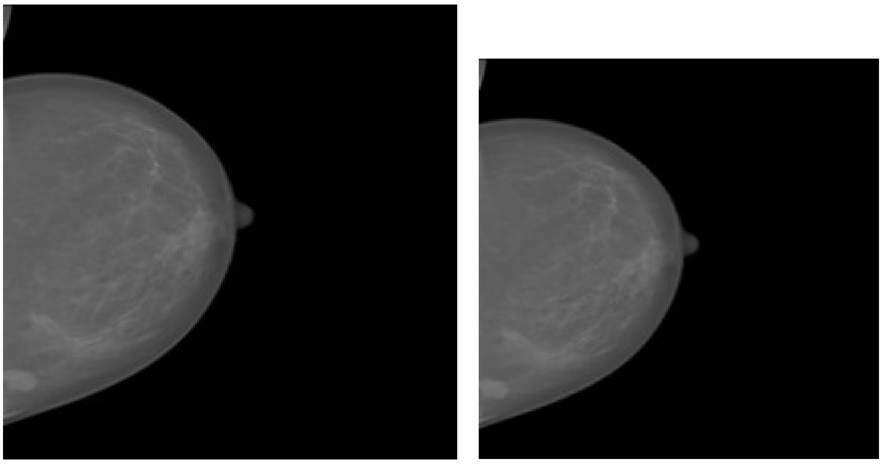


(b)

Figure 11: (a) Benign images from dataset (b) Malignant images from dataset

4.3 Image Preprocessing

An image preprocessing step is necessary for image consistency, noise reduction, and improved image features. The images have been scaled down to 224 by 224 because of CNN(Resnet)'s architecture configuration, which we are using in our research. Additionally, the unnecessary portions of the images were cropped out using a cropping technique. Figure 12 depicts the outcome.



(a) (b)

Figure 12: (a) Original Image (b) Preprocessed Image

4.4 Decision Fusion

In the decision fusion process, each of the three models is individually applied to the input data. From each model, probabilities are obtained, representing the likelihood of different outcomes. These probabilities are then combined using either the sum rule product rule or the majority vote, which are fusion techniques described in detail in section 3.3. The sum rule involves adding the probabilities from each model, while the product rule involves multiplying them and majority vote works the predicted labels from each classifier are compiled, and the class label with the highest frequency or count is selected as the final predicted label.By employing these fusion techniques, an improved single prediction is generated, taking into account the collective information from all three models.

4.5 Training of Model

Before training the models, the dataset should be balanced, meaning that each category has the same number of images. This helps prevent overfitting. The dataset is then divided into two parts: the training set and the testing set. Firstly, the CNN architecture is trained in the proposed model. The architecture includes layers such as convolutional layers, pooling layers, batch normalization layers, activation layers, and fully connected layers with a softmax output for classification. These layers are defined using MATLAB's Deep Learning Toolbox's Layer Graph API. Once the network layers are defined, training options are set. These options involve selecting an optimizer and specifying parameters like the mini-batch size, number of epochs, initial learning rate, shuffling of data, validation data, and plotting options. These options help configure the training process for the network, allowing for effective model optimization.

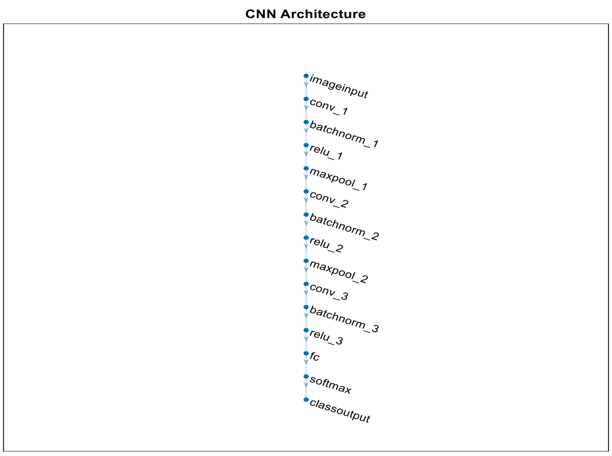


Figure 13: Proposed CNN model generated in matlab.

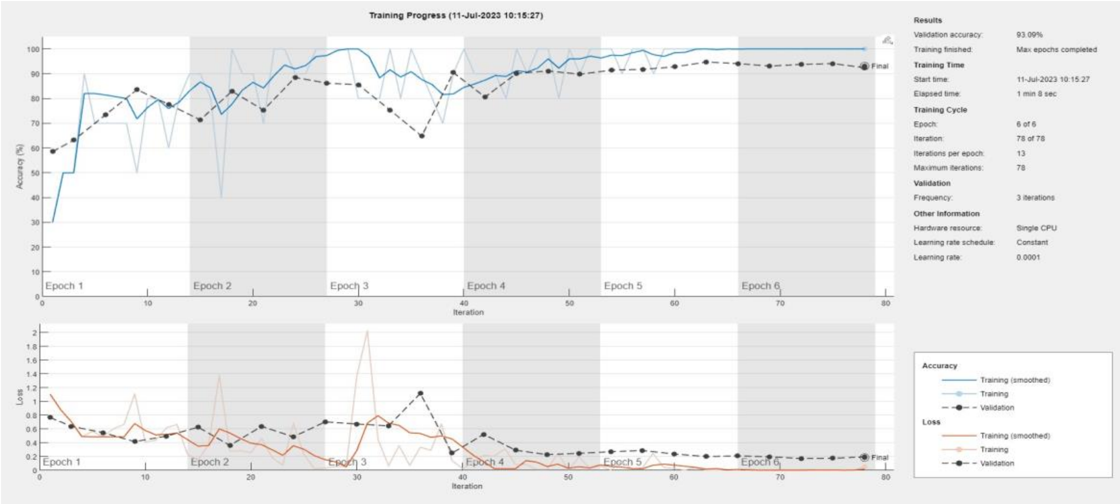


Figure 14: Proposed CNN model training process.

Table 2: Experimentally decided hyper-parameters settings for CNN model

Hyper-parameter	
Epochs	6
Optimizer	Stochastic gradient descent with momentum
Batch size	10
Activation function	Relu
Stride	2
Shuffle	Every epoch
Validation frequency	3
Initial learn rate	$1 * 10^{-4}$

The table 2 shows hyper-parameters settings for CNN model to classify the image. The hyperparameters were chosen through experimentation. The two cases of underfitting and overfitting are considered when choosing an epoch number. Monitoring is done of the training and validation losses. If the training loss increases while the validation loss decreases, overfitting is present. Or if you have poor performance, it means underfitting. The optimizer SGD with momentum is chosen. The weights of the model are updated in traditional SGD after the gradient of the loss function with respect to the weights for each distinct training is calculated. A moving average of previous gradients is used in SGD with Momentum to smooth out updates and build momentum during the optimization process. Not only does the pass gradient affect velocity, but also velocity. Compared to traditional SGD, it is faster. The batch size controls the training's speed. But memory is required. Model nonlinearity is introduced using the relu activation function. The convolutional filter moves around the input image in steps of a certain size called strides during the convolution process. When a convolutional layer has a stride of 2, it means that when moving across the input image to perform the convolution, the filter skips 2 pixels (in both the height and width directions). The epoch shuffle process is where the training data is randomly shuffled before the start of the subsequent epoch following the conclusion of each epoch. It aids in avoiding overfitting. The learning rate and validation frequency have finally been decided. Validation frequency affects how often the system is assessed during the training process. The time of convergence is influenced by the learning rate.

In the second technique, training is carried out on a pre-trained ResNet-50 convolutional neural network. The network has been trained up to layer FC100 with a batch size of 32. In order to train an SVM classifier, image features are extracted from that layer. To expedite the training process when dealing with high-dimensional CNN feature vectors, a fast Stochastic Gradient Descent (SGD) solver is employed by setting the 'Learners' parameter of the fitcecoc function to 'Linear'. This approach helps accelerate training when working with high-dimensional CNN feature vectors. For the SVM classifier, the input vector dimension is extracted from ResNet50 512, and the output dimension is two for benign and malignant. Figure 15 demonstrates ResNet50 model architecture.

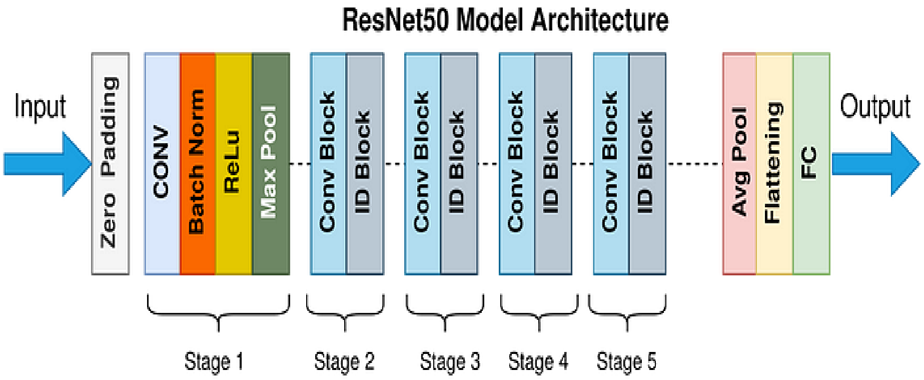


Figure 15: ResNet50 model architecture [54].

In the final approach, the SVM classifier is trained using features extracted using the LBP method. To reduce the training time, an efficient optimization algorithm called SGD solver is employed.The SGD solver updates the SVM model parameters incrementally by considering a subset of training samples at each iteration. This incremental update strategy allows for faster convergence compared to traditional solvers that consider the entire training set in each iteration.

4.6 System Overview

First of all, a traditional CNN with three hidden layers and max pooling is used to classify images. Then, a pre-trained CNN called ResNet-50 is used. The last layer is omitted, which is the classification layer of ResNet-50, and the remaining layers are used as a feature extractor.

The features extracted from the FC100 layer,that is the last layer of ResNet 50 before the classification layer. Then features passed to an SVM classifier for classification. This approach is commonly known as feature extraction process. Thirdly, the SVM classifier is used for image classification using LBP features. Finally, decision fusion is implemented to combine probabilities from independent multi models to classify class.

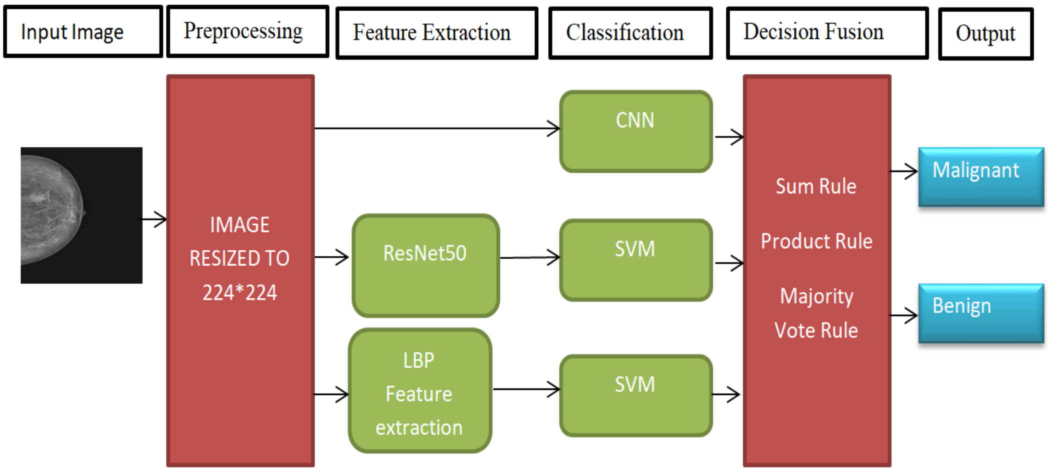


Figure 16: Detection of Breast Cancer According to the Proposed Approach

Chapter 5

RESULTS AND DISCUSSION

The system architecture, its hyper-parameters, and how to select the ideal classifier for the purpose were covered in the prior chapter. This chapter analyzes the outcomes of various experiments while mentioning the impacts of particular parameters on the model's capability.

5.1 Evaluation Metrics

Evaluation is essential to any project in order to determine how effective a learning model or algorithm is. The effectiveness of a predictive model is typically measured using a combination of various individual evaluation matrices.

5.1.1 Confusion Matrix

A confusion matrix, also known as a contingency table, can be used to depict the four possible outcomes such as TP, TN, FP and FN. Figure 17's table serves as a representation of information. Confusion matrices summarize all the details of a situation. The data in the table is utilized to calculate the other indices. By simply adding rows and columns for the number of projected classes, this confusion matrix can be applied to situations involving the categorization of more than one class.

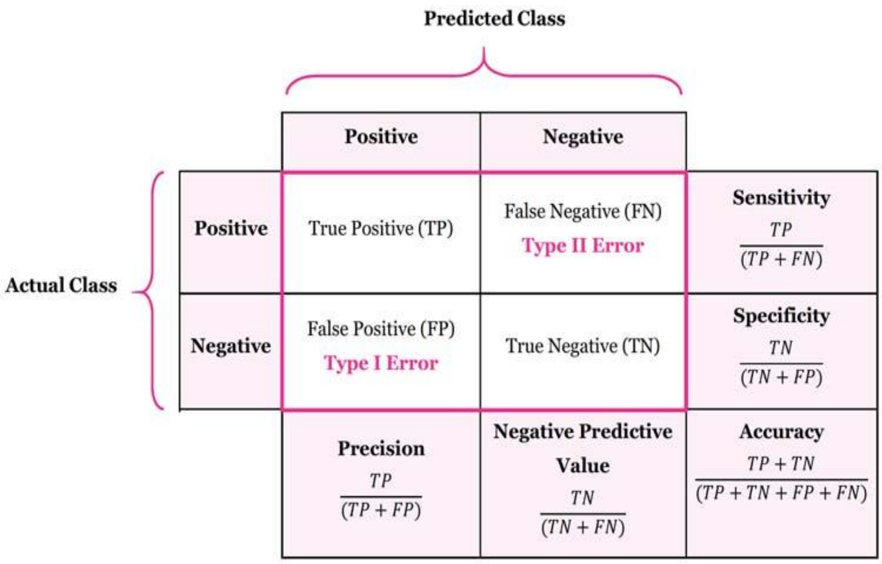


Figure 17: Confusion Matrix for Binary Classification [50].

5.1.2 Sensitivity and Specificity

The most crucial metrics for confusion matrices are Sensitivity and Specificity. Recall (eq. 5), also known as sensitivity, is a measure of how frequently true positive outcomes are correctly identified. To assess whether true negatives are correctly identified as such as a result of actual testing, the specificity (eq. 6) is used. This can be mathematically stated as:

$$SEN = \frac{TP}{(TP+FN)} \times 100 \tag{5}$$

$$SPE = \frac{TN}{(TN+FP)} \times 100 \tag{6}$$

In medical diagnosis, sensitivity refers to the model's capacity to identify patients with the condition properly, while specificity refers to its capacity to identify patients without the disease.

5.1.3 False Positive and Negative Rates

The risk that the model will report incorrect values is gauged by the incorrect false positive rate (eq 7) and negative rate (eq 8) indices. In statistical hypothesis testing, these indices are referred to as type I and type II mistakes. They have a significant role in deciding screening and testing procedures among medical diagnoses.

$$FPR = \frac{FP}{(FP+TN)} \times 100 \quad (7)$$

$$FNR = \frac{FN}{(FN+TP)} \times 100 \quad (8)$$

5.1.4 Accuracy

The effectiveness of a diagnostic test is assessed based on a number of variables, including its "sensitivity" and "specificity," among others. Two distinct kinds of accuracy are sensitivity and specificity. These metrics, along with the related indices TPR and FPR, are superior to "accuracy," but they do not accurately describe diagnostic performance because they rely on an arbitrary decision threshold.Equation 9 states the accuracy.

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP+TN+FP+FN})} \quad (9)$$

5.1.5 Precision

The percentage of real positives compared to all true positives and false positives is displayed here. The quantity of false positives included in the sample is examined by precision. The accuracy of a model is 100% if there are no FPs. The precision will appear to be deteriorating as more FPs are added to the mixture. The precision is given by equation 10 below.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \quad (10)$$

5.1.6 F1-Score

A model's ability to fit a dataset is gauged by the F-score, also known as the F1-score. It is used to assess categorization systems that categorize examples as either "positive" or "negative." The F-score, also referred to as a model's harmonic mean, combines a model's recall and precision. The answer is given in Equation 11.

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

5.1.7 ROC Analysis

An examination of the ROC curve offers a graphic representation of the trade-offs between sensitivity and specificity for various test result values. It can also be used to choose the ideal threshold for a particular application. The y-axis and x-axis of a two-dimensional unit square plot, on which the ROC curve can be displayed, can represent the TPR (or sensitivity) and FPR (1-specificity), which are both close to (0, 1), respectively. Figure 18 shows the binary classification ROC analysis performed for this thesis. More information is provided in the chapter that follows.

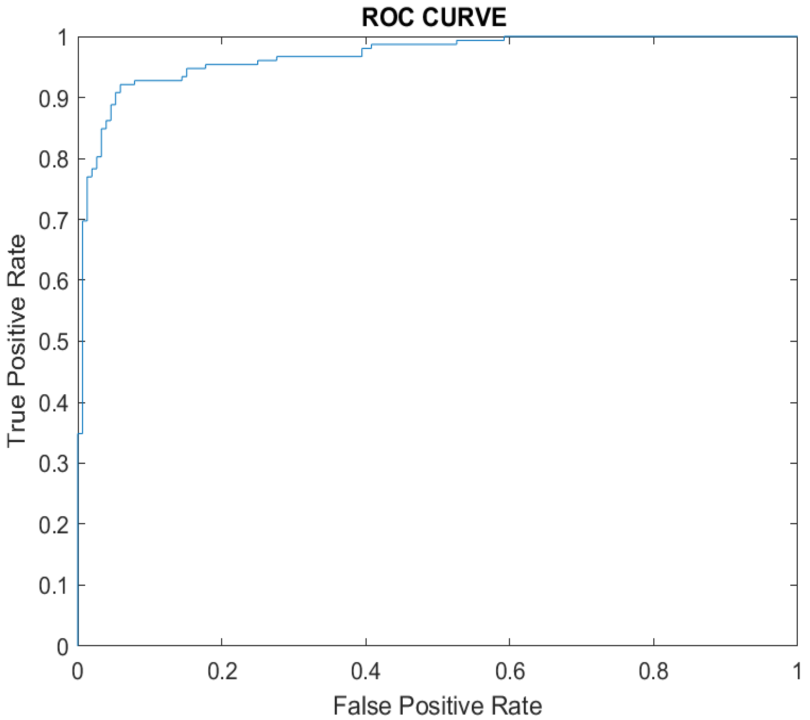


Figure 18: ROC curve of Binary Classification for Breast cancer.

5.2 Experimental Results and Discussions

The proposed framework is evaluated based on its predictions on test data following training for each fold as stated in 5.1. This is done in order to thoroughly assess the framework's effectiveness and obtain a reliable evaluation of how well it will perform when used to diagnose patients who have never received a diagnosis.The testing accuracies obtained after training and testing on all folds are averaged to find each model's overall accuracy. A confusion matrix is calculated for each fold and then added together to create the overall confusion matrix for the model.

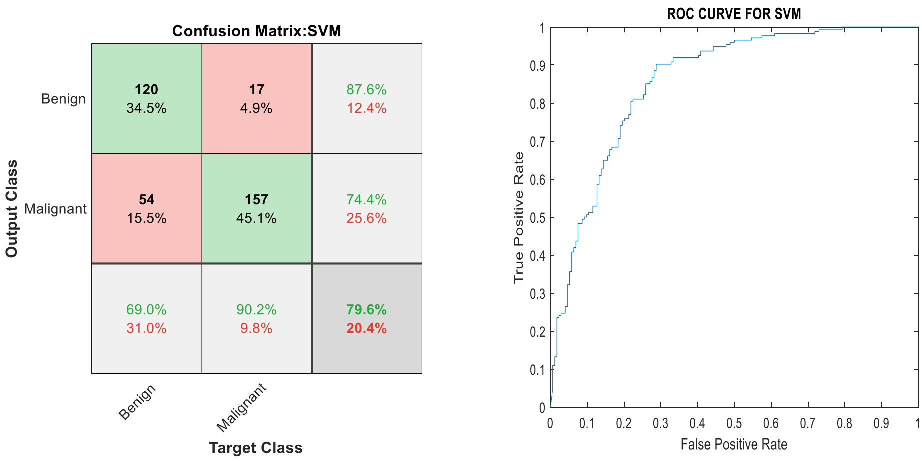
The predictions from each pertinent model using the pertinent data point are then combined to create a single, more precise prediction as mentioned in section 3.10. In addition, all predictions from the three models are applied, accounting for all testing images, using the sum rule product rule and majority vote methods.

Table 3: Accuarcy results after each training

TRAINING /TESTING RATIO	20 / 80 %		30 / 70%		40 / 60%		50 / 50%		70 / 30%	
	SVM	79.8	SVM	81.1	SVM	80.3	SVM	80.1	SVM	83.7
TRAINING 1 ACCURACIES (%)	SVM+Resnet50	83.2	SVM+Resnet50	87.1	SVM+ResNet50	90.3	SVM+Resnet50	93.1	SVM+Resnet50	88
	CNN	84.9	CNN	88.7	CNN	96.2	CNN	95.9	CNN	98.4
	SVM	80.4	SVM	80.9	SVM	81.2	SVM	79.3	SVM	84.7
TRAINING 2 ACCURACIES (%)	SVM+Resnet50	80.6	SVM+Resnet50	87.9	SVM+Resnet50	92	SVM+Resnet50	89.5	SVM+Resnet50	94.3
	CNN	86.1	CNN	91.9	CNN	96.1	CNN	97.1	CNN	97.1
	SVM	78.9	SVM	80.9	SVM	81.1	SVM	83.1	SVM	82.8
TRAINING 3 ACCURACIES (%)	SVM+Resnet50	80.8	SVM+Resnet50	84.7	SVM+Resnet50	91.4	SVM+Resnet50	94.2	SVM+Resnet50	96.6
	CNN	83.2	CNN	90.9	CNN	97	CNN	96.8	CNN	96.9
	SVM	79.1	SVM	80.8	SVM	79.8	SVM	82.8	SVM	84.1
TRAINING 4 ACCURACIES (%)	SVM+Resnet50	80.4	SVM+Resnet50	84.5	SVM+Resnet50	91.2	SVM+Resnet50	94.4	SVM+Resnet50	96.4
	CNN	83.1	CNN	89.9	CNN	97.1	CNN	96.9	CNN	96.8
	SVM	80.1	SVM	80.7	SVM	80.2	SVM	80.1	SVM	80.9
TRAINING 5 ACCURACIES (%)	SVM+Resnet50	83.9	SVM+Resnet50	87.5	SVM+Resnet50	91.1	SVM+Resnet50	92.1	SVM+Resnet50	95.1
	CNN	84.1	CNN	91.6	CNN	96.5	CNN	97.1	CNN	96.9
	SVM	79.6	SVM	80.9	SVM	80.4	SVM	81.0	SVM	83.1
AVERAGE ACCURACY RESULTS FOR EACH METHOD	SVM+Resnet50	81.9	SVM+Resnet50	86.2	SVM+Resnet50	90.0	SVM+Resnet50	92.6	SVM+Resnet50	93.1
	CNN	84.2	CNN	90.8	CNN	95.8	CNN	96.8	CNN	97.7

Table 3 shows results for each technique's accuracy and training/testing ratio. Each method was trained five times for each training and testing ratio. After that, the average is taken for each method to improve stability and generalization. From the table,it is clearly seen that when the training ratio is increased,the accuracy of each method is soar.When the dataset is divided by 70% for training and 30% for testing, the SVM method reaches 83.1%, Resnet50+SVM 93.1%, and CNN 97.7% accuracy.

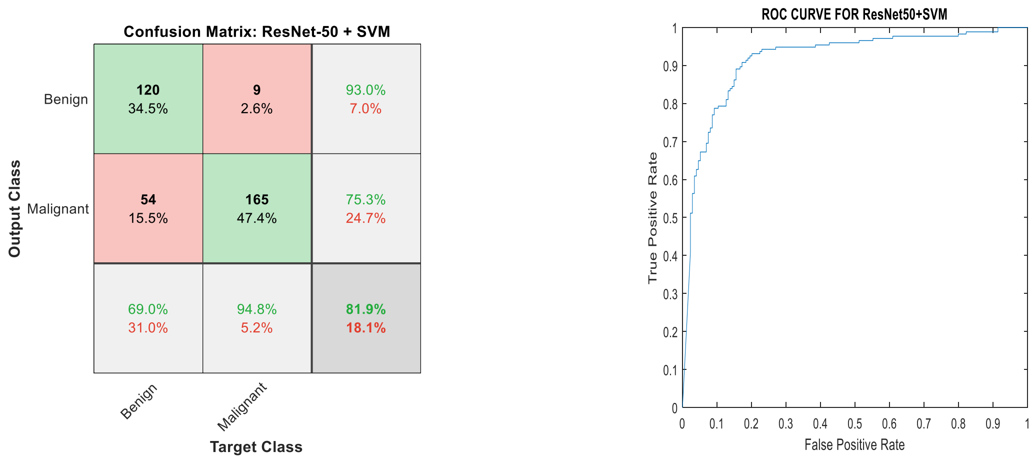
When the dataset is divided into 20% for training and 80% for testing, the average confusion matrices and ROC curves are as follows for each method:



(a) (b)

Figure 19: (a) Confusion matrix for SVM. (b) ROC curve for (a) SVM.

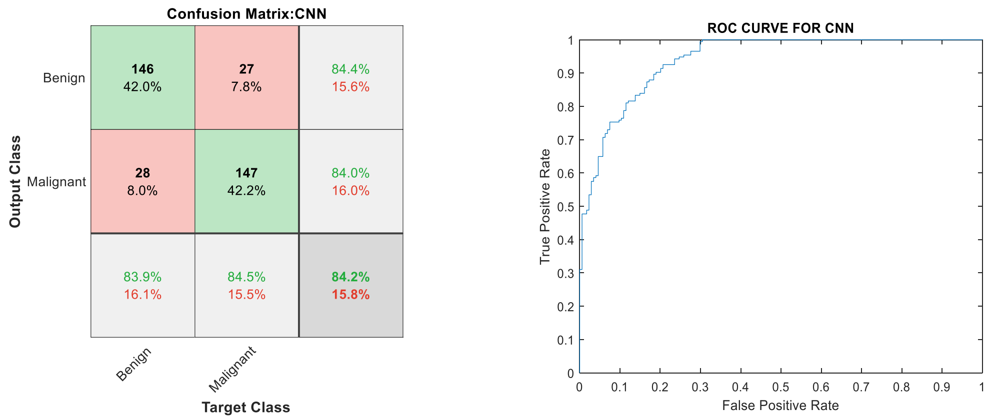
In Figure 19, the confusion matrix shows that 277 images out of 348 are correctly identified.



(a) (b)

Figure 20: (a) Confusion matrix for ResNet50+SVM. (b) ROC curve for ResNet50+SVM.

The confusion matrix shown in Figure 20 indicates that 285 of the 348 images were correctly identified.



(a) (b)

Figure 21: (a) Confusion matrix for CNN. (b) ROC curve for CNN.

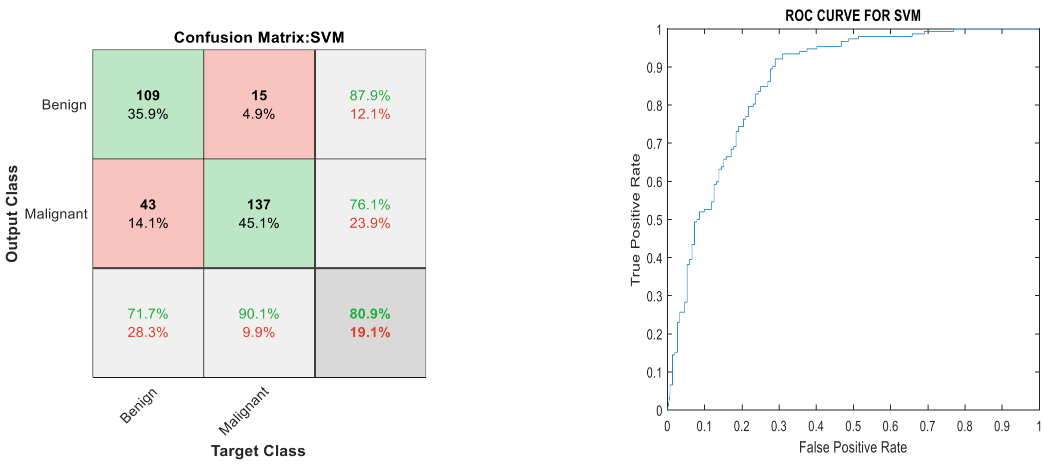
Figure 21 shows the confusion matrix of CNN, showing that 293 images out of 348 are correctly identified.

Table 4: Average scores (%) when the dataset is 20% for training and 80% for testing

MODEL	SEN(%)	SPE(%)	PRE(%)	F1-SCORE(%)	ACC(%)
SVM	87.6	74.4	69.0	77.2	79.6
SVM+ResNet50	93.0	75.3	69.0	79.2	81.9
CNN	84.4	84.0	83.9	84.1	84.2

According to Table 4, the CNN showed better performance in all scores except for its sensivity score. Also in the ROC curve, CNN has the highest area under the curve value.

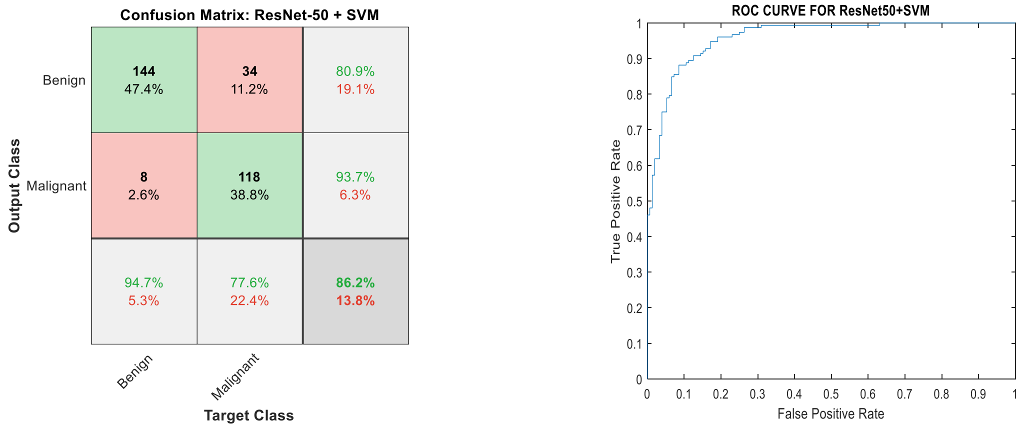
The average confusion matrices and ROC curves for each method are as follows when the dataset is split into 30% for training and 70% for testing:



(a) (b)

Figure 22: (a) Confusion matrix for SVM. (b) ROC curve for SVM.

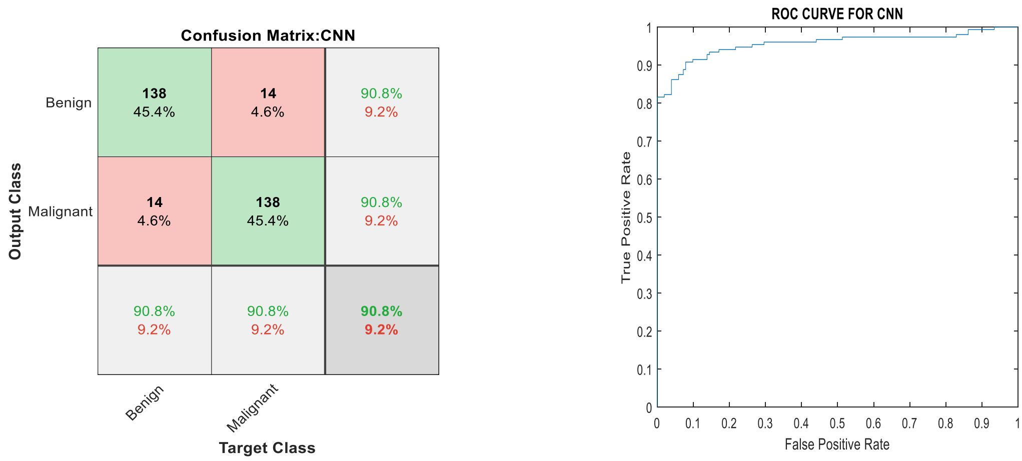
Figure 22, the confusion matrix illustrates that 58 images out of 304 are falsely identified, and it achieves 80.9% accuracy.



(a) (b)

Figure 23: (a) Confusion matrix for ResNet50+SVM. (b) ROC curve for ResNet50+SVM.

The confusion matrix shows that 262 out of 304 images were correctly identified, yielding an accuracy rate of 86.2% in Figure 23.



(a) (b)

Figure 24: (a) Confusion matrix for CNN. (b) ROC curve for CNN.

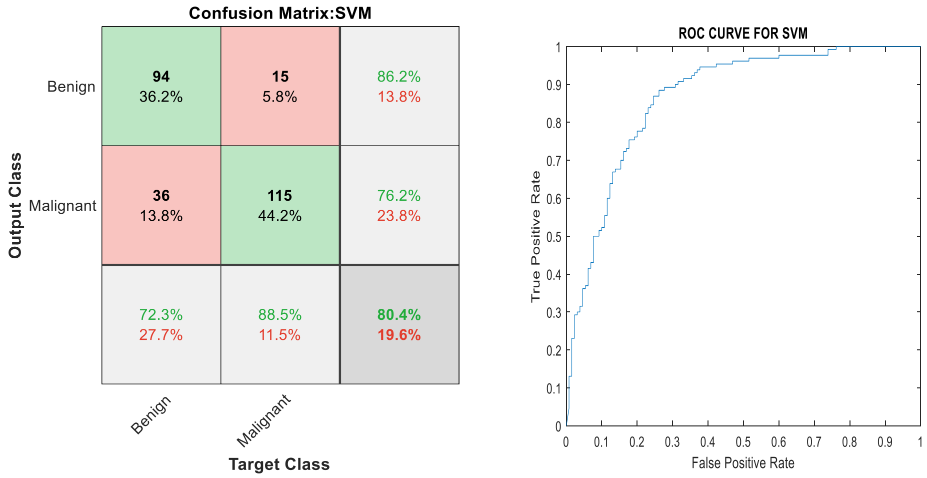
The confusion matrix, shown in Figure 24, indicates that 276 out of 304 images were correctly identified, resulting in an accuracy rate of 90.8.

Table 5: Average scores (%) when the dataset is 30% for training and 70% for testing

MODEL	SEN(%)	SPE(%)	PRE(%)	F1-SCORE(%)	ACC(%)
SVM	87.9	76.1	71.7	79.0	80.9
SVM+ResNet50	80.9	93.7	94.7	87.3	86.2
CNN	90.8	90.8	90.8	90.8	90.8

From table 5, in every category, SVM performed poorly except in sensitivity. CNN has the highest area under the curve value in the ROC curve.

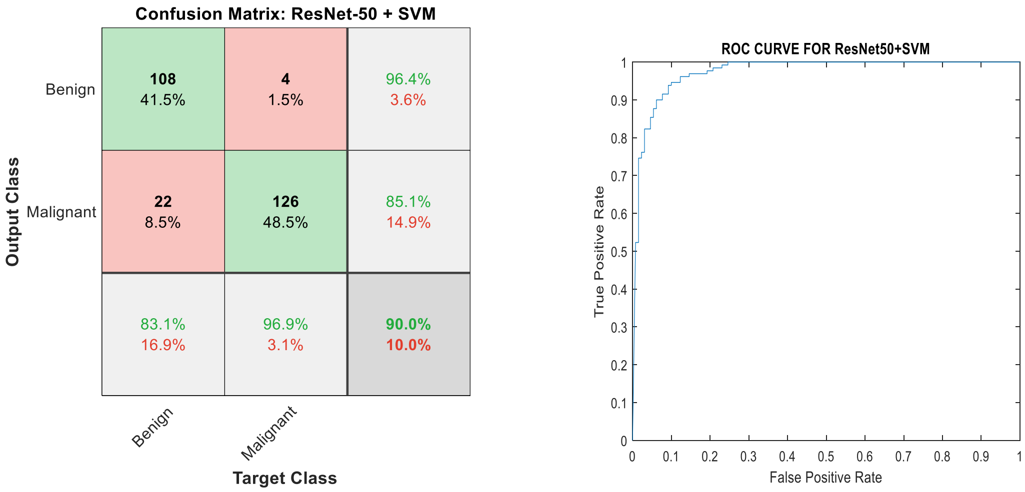
Following are the average confusion matrices and ROC curves for each technique when the dataset is split into 40% for training and 60% for testing:



(a) (b)

Figure 25: (a) Confusion matrix for SVM. (b) ROC curve for SVM.

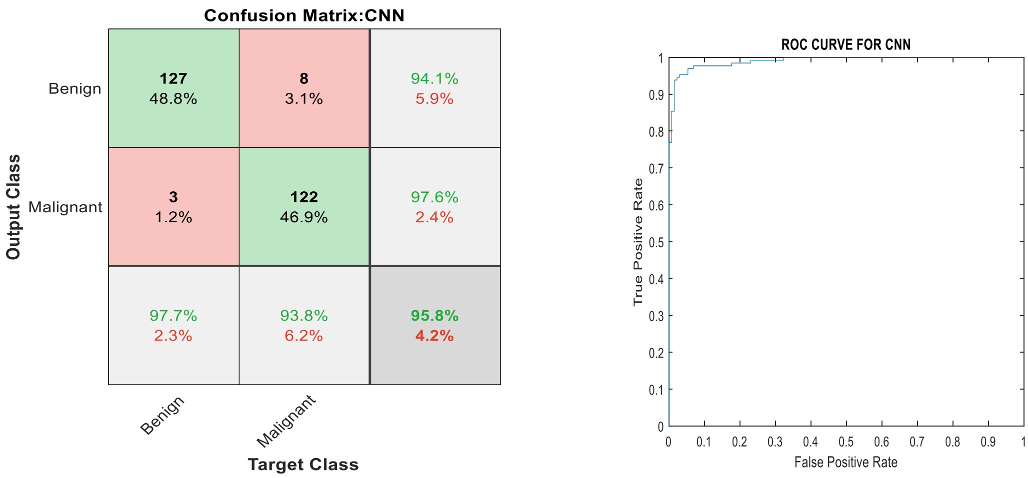
Figure 25 shows that 209 out of 260 images were correctly identified, according to the confusion matrix, for an accuracy rate of 80.4%.



(a) (b)

Figure 26: (a) Confusion matrix for ResNet50+SVM. (b) ROC curve for ResNet50+SVM.

The confusion matrix indicated that 234 out of 260 images had a 90.0% accuracy rate in terms of identification in Figure 26.



(a) (b)

Figure 27: (a) Confusion matrix for CNN. (b) ROC curve for CNN.

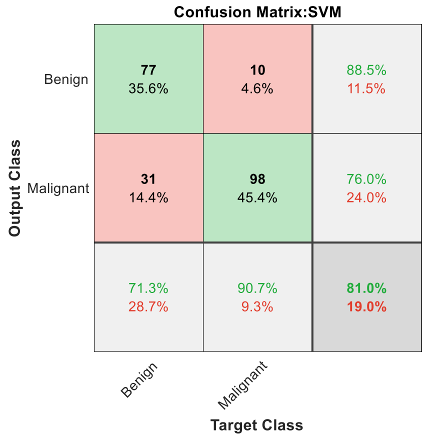
Figure 27 illustrates the confusion matrix of CNN, which indicated that 249 out of 260 images had a 95.8% accuracy rate in terms of identification.

Table 6: Average scores (%) when the dataset is 40% for training and 60% for testing

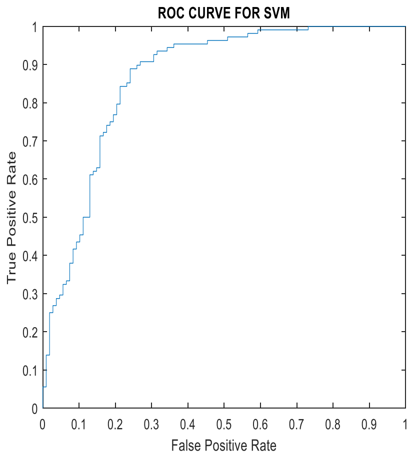
MODEL	SEN(%)	SPE(%)	PRE(%)	F1-SCORE(%)	ACC(%)
SVM	86.2	76.2	72.3	78.6	80.4
SVM+ResNet50	96.4	85.1	83.1	89.3	90.0
CNN	94.1	97.6	97.7	95.9	95.8

With the exception of its sensivity score, CNN outperformed rivals in every category, as shown in Table 6. In the ROC curve, CNN also has the highest area under the curve value.

The average confusion matrices and ROC curves for each technique are as follows when the dataset is split into 50% training and 50% testing:



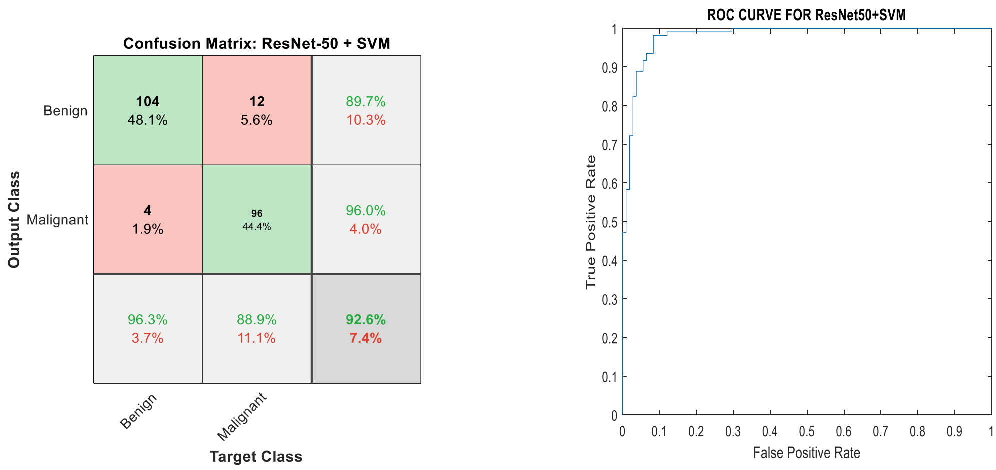
(a)



(b)

Figure 28: (a) Confusion matrix for SVM. (b) ROC curve for SVM.

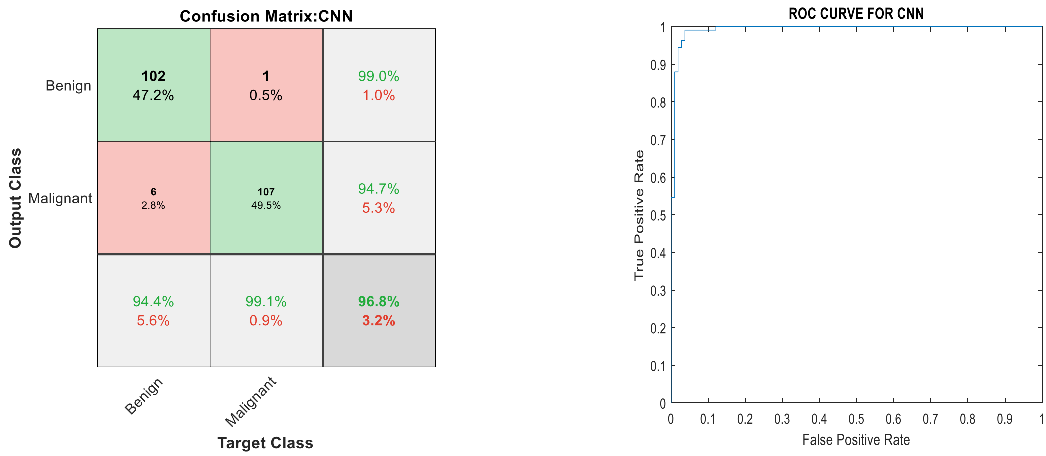
Figure 28, the confusion matrix, shows that 81.0% accuracy was achieved and that 41 out of 216 images were incorrectly identified.



(a) (b)

Figure 29: (a) Confusion matrix for ResNet50+SVM. (b) ROC curve for ResNet50+SVM.

In Figure 29, according to the confusion matrix, 200 out of 216 images were correctly identified with a 92.6% accuracy rate.



(a) (b)

Figure 30: (a) Confusion matrix for CNN. (b) ROC curve for CNN.

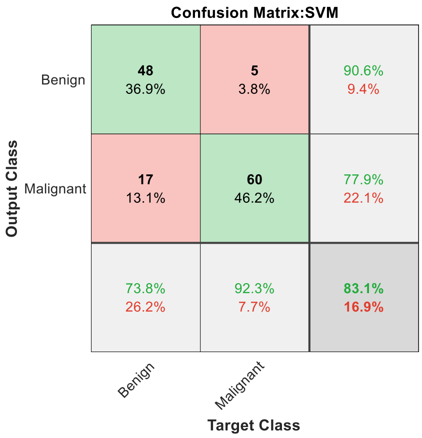
The confusion matrix shows that 209 out of 216 images were correctly identified with a 96.8% accuracy rate in Figure 30.

Table 7: Average scores (%) when the dataset is 50% for training and 50% for testing

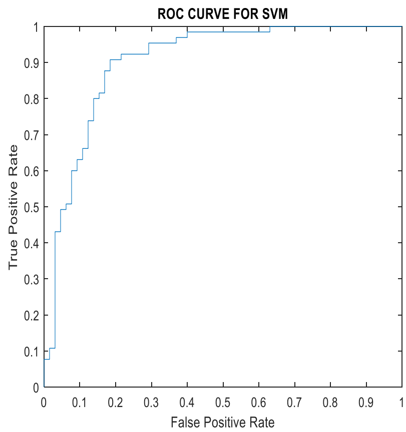
MODEL	SEN(%)	SPE(%)	PRE(%)	F1-SCORE(%)	ACC(%)
SVM	88.5	76.0	71.3	79.0	81.0
SVM+ResNet50	89.7	96.0	96.3	92.9	92.6
CNN	99.0	94.7	94.4	96.6	96.8

Table 7 illustrates that in terms of sensitivity, F1-score, and accuracy, CNN performed best. In the other two scores, ResNet50+SVM passed other techniques. When we analyze the ROC curve, it is clearly seen that CNN has more area under the graph, which means CNN's classification model performs better than other methods.

When the dataset is divided into 70% training and 30% testing, the average confusion matrices and ROC curves for each technique are as follows:



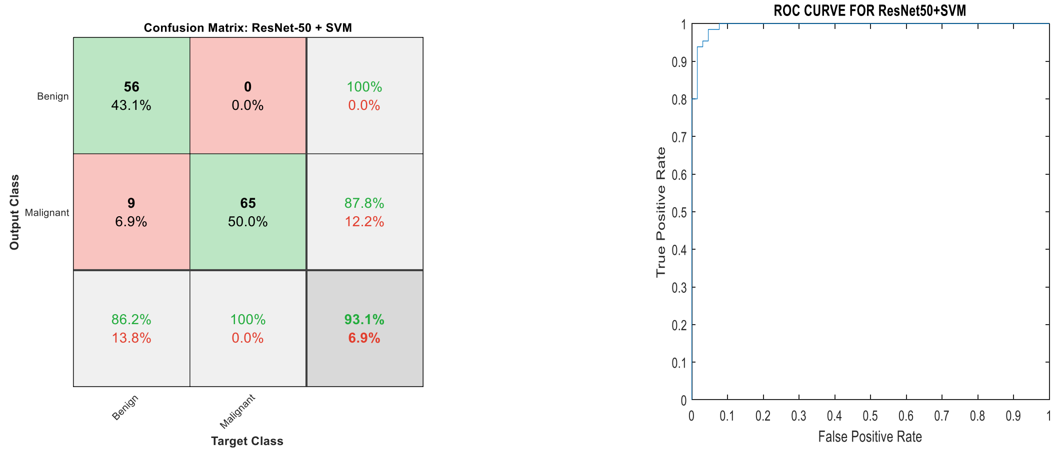
(a)



(b)

Figure 31: (a) Confusion matrix for SVM. (b) ROC curve for SVM.

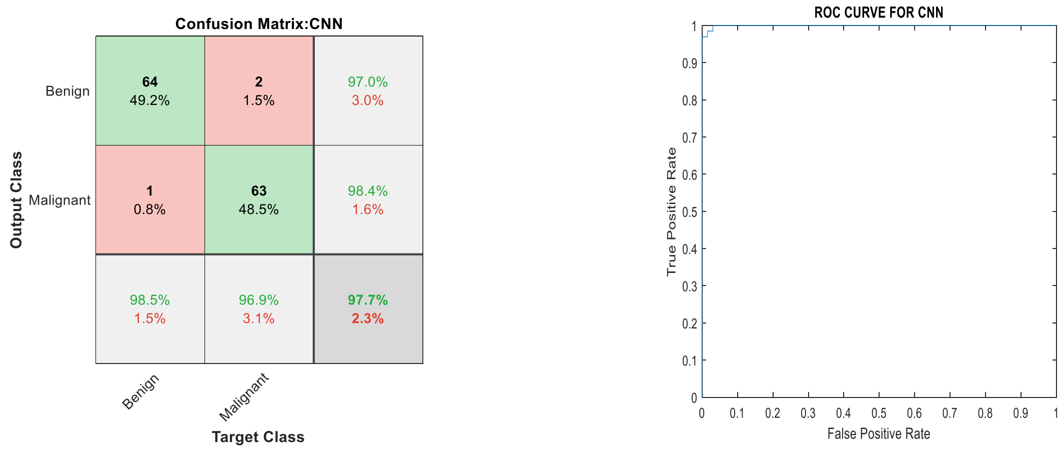
In Figure 31, the confusion matrix of SVM images illustrates that 108 out of 130 images were correctly identified, with an 83.1% accuracy rate.



(a) (b)

Figure 32: (a) Confusion matrix for ResNet50+SVM. (b) ROC curve for ResNet50+SVM.

The confusion matrix in Figure 32 indicated that 121 out of 130 images had a 93.1% accuracy rate in terms of identification.



(a) (b)

Figure 33: (a) Confusion matrix for CNN. (b) ROC curve for CNN.

In Figure 33, according to the confusion matrix, 127 out of 130 images were correctly identified with a 97.7% accuracy rate.

Table 8: Average scores (%) when the dataset is 70% for training and 30% for testing

MODEL	SEN(%)	SPE(%)	PRE(%)	F1-SCORE(%)	ACC(%)
SVM	90.6	77.9	73.8	81.3	83.1
SVM+ResNet50	100	87.8	86.2	92.6	93.1
CNN	97	98.4	98.5	97.7	97.7

According to Table 8, in every category but sensivity, CNN outperformed the competition. In the ROC curve, CNN also has the highest area under the curve value.

Individual classification results were obtained, and various fusion rules were then applied to improve accuracy. These included the product rule, which multiplied probabilities, and the sum rule, which added confidence scores. The majority vote rule also made use of numerous predictions. To balance contributions, an average fusion was finally computed. These fusion techniques are designed to increase robustness and take into account different classifier strengths. The combined prediction that results offers a more thorough and trustworthy classification outcome.

Table 9 demonstrates the average accuracies for each training and testing ratio after each fusion rule is applied.

Table 9: Average accuracies (%) after fusion rules

TRAINING DATASET RATIO	20%	30%	40%	50%	70%
AVERAGE ACCURACY RESULTS AFTER SUM RULE FUSION	93.8	95.8	96.5	97.9	98.7
AVERAGE ACCURACY RESULTS AFTER PRODUCT RULE FUSION	95.4	95.6	96.9	97.2	99.1
AVERAGE ACCURACY RESULTS AFTER MAJORITY VOTE RULE FUSION	93.4	95.4	96.8	97.4	98.9

5.3 Comparison with other Methods

This section discusses an accuracy-based comparison of the results to those of other recent methodologies. Structural MRI images were used for the implementation of all the techniques because they are highly comparable across datasets, especially when preprocessing has been applied and the breast has already been recorded and segmented in the published datasets. Even though some of the methods may not have used the same datasets and/or experimental setups as in this work, the results can still be compared. We contrast our findings using the deep learning and machine learning strategies covered in section 2.8.

Table 10: The system classifier and earlier classifiers are compared in the comparison results

Researcher	Classification	Training/Test ratio(%)	Dataset	Model	Acc(%)
Khourdifi & Bahaj [37]	Benign,Malignant	0.75/0.25	Wisconsin Diagnostic dataset	K-NN,RF,NB	NA
				SVM	0.971
Xu et al [44]	Benign,Malignant	0.85/0.15	Public, INbreast dataset	CNN	0.924
Dhungel et al. [45]	Benign,Malignant	0.60/0.40	DDSM	ResNet50	0.907
Jiao et al. [47]	Benign,Malignant	NA	DDSM	CNN+SVM	0.964
Proposed system	Benign,Malignant	0.60/0.40	DDSM	ResNet50	0.915
Proposed system	Benign,Malignnt	0.70/0.30	DDSM	SVM	0.808
				ResNet50	0.921
				ResNet50+SVM	0.931
				CNN	0.977
Proposd system	Benign,Malignant	0.70/0.30	DDSM	Sum rule fusion	0.987
Proposd system	Benign,Malignant	0.70/0.30	DDSM	Product rule fusion	0.991
Proposd system	Benign,Malignant	0.70/0.30	DDSM	Majority vote fusion	0.989

The proposed method, which uses data fusion methods with mammogram images, outperforms all other methods in terms of accuracy, as shown in the table above. Each method has particular advantages and disadvantages, but using fusion techniques, it is possible to combine them and build a robust detection model, as shown in table 10.

5.4 Multi-modal Performance

The offered Multi-modal method outperforms every method in Table 5.2.7, achieving an accuracy of 99.1 with product rule, 98.7% with sum rule and 98.9% with majority vote. The outcome of fusion is generally regarded as precise and robust, which is a crucial factor in diagnosis.

Chapter 6

CONCLUSION AND FUTURE WORK

This chapter expresses in brief the results obtained as well as the contribution of this thesis and recommendationsfor future work.

6.1 Conclusion

The issue of breast abnormality classifications in mammography images is the main topic of the thesis. Radiologists can diagnose cases of breast cancer with the least amount of time and effort by using a classification system that has been suggested. The outcome revealed increased accuracy compared to correctness of a person. When comparing the results with earlier research efforts in this area, the robustness of the methods adopted was clearly seen in terms of thorough classification and marginally improved accuracy.Four steps make up the system that is proposed in this thesis suggest. In the beginning, mammogram images are classified using CNN. In order to detect potential tumors, standard classifier SVM is used in conjunction with features extracted from CNN-based images. Thirdly, using the features that are extracted from the tumors, SVM is used to differentiate between various tumor types. In order to improve the system's overall performance, data fusion methods such as sum rule,product rule and majority voting are used to combine the findings from the first, second, and third methods. In this study, 434 grayscale images were used to build the proposed model. The efficacy of the proposed method is tested using metrics that are frequently used to evaluate the performance of models when working with medical images, such as Accuracy, ROC, recall, and F1 score.

The results of the proposed strategy are compared to those reported in the literature. The outcomes demonstrate that suggested method yields results that are comparable to those described in the literature.

6.2 Future Work

The performance of predictive models could be significantly improved by including a trainable decision fusion layer in future research. With the help of this trainable layer, decisions drawn from various component models can be blended adaptively to produce better results. The system as a whole becomes trainable, enabling the optimization of decision combination strategies, by allowing the training weights of the fusion layer to be changed. The addition of a trainable fusion layer is anticipated to produce better outcomes than the current method. The most efficient method for combining the discrete outputs produced by each model can be dynamically determined by the trainable fusion layer, improving prediction performance. The fusion layer can learn from and adapt to the characteristics and strengths of each component model through an iterative training process, which improves the decision fusion process.This development creates opportunities for a variety of applications, including ensemble learning, which combines multiple models to create a more reliable and accurate predictive system. Furthermore, the trainable fusion layer's adaptability enables the seamless integration of new models or changes to existing ones without the need for significant system-wide adjustments.

REFERENCES

[1] Breast cancer. (2023, July 12). Breast Cancer. https://www.who.int/news room/fact-sheets/detail/breast-cancer

[2] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal For Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

[3] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. *(2021, March 31)*. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions - *Journal of Big Data*. *SpringerOpen*.

[4] Wilkinson, L., & Gathani, T. (2022). Understanding breast cancer as a global health concern. *The British Journal of Radiology*, *95*(1130), 20211033.

[5] Wilkinson, L., & Gathani, T. (2022). Understanding breast cancer as a global health concern. *The British journal of radiology*, *95*(1130), 20211033. https://doi.org/10.1259/bjr.20211033

[6] Allaouzi, I., & Ben Ahmed, M. (2019). *A Novel Approach for Multi-Label Chest X-Ray Classification of Common Thorax Diseases.* IEEE Access, 7, 64279–64288. https://doi.org/10.1109/access.2019.2916849

[7] *Debelee, T. G., Kebede, S. R., Schwenker, F., & Shewarega, Z. M. (2020, November 10). Deep Learning in Selected Cancers' Image Analysis—A Survey. Journal of Imaging, 6(11), 121. https://doi.org/10.3390/jimaging6110121*

[8] Assefa, T., Haile Mariam, D., Mekonnen, W., Derbew, M., & Enbiale, W. (2016). Physician distribution and attrition in the public health sector of Ethiopia. *Risk management and healthcare policy*, *9*, 285–295. https://doi.org/10.2147/RMHP.S117943

[9] Narula, M. K., Chaudhary, V., Baruah, D., Kathuria, M., & Anand, R. (2010, January). Pictorial essay: *Orbital tuberculosis. Indian Journal of Radiology and Imaging*, *20*(01), 6–10. https://doi.org/10.4103/0971-3026.59744

[10] American College of Obstetricians and Gynecologists. (2017). Breast cancer risk assessment and screening in average-risk women. *Practice bulletin*, (179), 2019-1.

[11] Lakhtakia R. (2014). A Brief History of Breast Cancer: Part I: Surgical domination reinvented. Sultan Qaboos University medical journal, 14(2), e166–e169.

[12] Hossain, A., Islam, M. T., Islam, M. T., Chowdhury, M. E., Rmili, H., & Samsuzzaman, M. (2020). A planar ultrawideband patch antenna array for microwave breast tumor detection. *Materials*, *13*(21), 4918.

[13] Kolla, V. R. K. (2015). Heart Disease Diagnosis Using Machine Learning Techniques In Python: A Comparative Study of Classification Algorithms For Predictive Modeling. *International Journal of Electronics and Communication Engineering & Technology*.

[14] Pediconi, F., Catalano, C., Roselli, A., Dominelli, V., Cagioli, S., Karatasiou, A., ... & Passariello, R. (2009). The challenge of imaging dense breast parenchyma: is magnetic resonance mammography the technique of choice? A comparative study with x-ray mammography and whole-breast ultrasound. *Investigative radiology*, *44*(7), 412-421.

[15] Wei, M., Du, Y., Wu, X., Su, Q., Zhu, J., Zheng, L., ... & Zhuang, J. (2020). A benign and malignant breast tumor classification method via efficiently combining texture and morphological features on ultrasound images. *Computational and Mathematical Methods in Medicine*, *2020*.

[16] Schreer, I. (2009). Dense breast tissue as an important risk factor for breast cancer and implications for early detection. *Breast Care*, *4*(2), 89-92.

[17] Lanyi, M. (2012). *Diagnosis and differential diagnosis of breast calcifications*. Springer Science & Business Media.

[18] Latif, J., Xiao, C., Imran, A., & Tu, S. (2019, January). Medical imaging using machine learning and deep learning algorithms: a review. In *2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1-5). IEEE.

[19] Ismail, N. S., & Sovuthy, C. (2019, August). Breast cancer detection based on deep learning technique. In *2019 International UNIMAS STEM 12th engineering conference (EnCon)* (pp. 89-92). IEEE.

[20] Tellaeche, A., & Arana, R. (2011). Three-dimensional machine vision and machine-learning algorithms applied to quality control of percussion caps. *IET computer vision*, *5*(2), 117-124.

[21] Liu, X., Beheshti, I., Zheng, W., Li, Y., Li, S., Zhao, Z., ... & Hu, B. (2022). Brain age estimation using multi-feature-based networks. *Computers in Biology and Medicine*, *143*, 105285.

[22] Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, *55*(3), 856-867.

[23] Asadi-Pooya, A. A., Kashkooli, M., Asadi-Pooya, A., Malekpour, M., & Jafari, A. (2022). Machine learning applications to differentiate comorbid functional seizures and epilepsy from pure functional seizures. *Journal of Psychosomatic Research*, *153*, 110703.

[24] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 1-27.

[25] Alahmadi, A. A., Aljabri, M., Alhaidari, F., Alharthi, D. J., Rayani, G. E., Marghalani, L. A., ... & Bajandouh, S. A. (2023). DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions. *Electronics*, *12*(14), 3103.

[26] Komin, M., & Vileykis, A. (2016). 'Good'and'Bad'Investments: Everything You Always Wanted to Know about Ukrainian Commanders but Were Afraid to Ask. *Connections*, *15*(1), 57-71.

[27] Phung, V. H., & Rhee, E. J. (2019). A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. *Applied Sciences*, *9*(21), 4500.

[28] Kang, B., Tripathi, S., & Nguyen, T. Q. (2020). Generating images in compressed domain using generative adversarial networks. *IEEE Access*, *8*, 180977-180991.

[29] Taye, M. M. (2023). Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. *Computation*, *11*(3), 52.

[30] Jadon, S. (n.d.). *Data Science Interview Preparation*. Data Science Interview Preparation.

[31] Beluffi, G. (2013). Radiology of the post surgical abdomen: J. Brittenden· DJM Tolan (Eds.) Springer London, Heidelberg, New York, Dordrecht, 2012 ISBN: 978-1-4471-2774-1 e-ISBN: 978-1-4471-2775-8 DOI: 10.1007/978-1-4471-2775-8.

[32] Ahmed, F., Kang, I. S., Kim, K. H., Asif, A., Rahim, C. S. A., Samantasinghar, A., ... & Choi, K. H. (2023). Drug repurposing for viral cancers: A paradigm of machine learning, deep learning, and virtual screening-based approaches. *Journal of Medical Virology*, *95*(4), e28693.

[33] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, *53*, 5455-5516.

[34] Fan, J., Lee, J., & Lee, Y. (2021). A transfer learning architecture based on a support vector machine for histopathology image classification. *Applied Sciences*, *11*(14), 6380.

[35] Czarnowski, I., & Jędrzejowicz, P. (2014). Ensemble classifier for mining data streams. *Procedia Computer Science*, *35*, 397-406.

[36] Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access*, *7*, 53040-53065.

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

[38] Khourdifi, Y., & Bahaj, M. (2018, December). Applying best machine learning algorithms for breast cancer prediction and classification. In *2018 International conference on electronics, control, optimization and computer science (ICECOCS)* (pp. 1-5). IEEE.

[39] Boeri, C., Chiappa, C., Galli, F., De Berardinis, V., Bardelli, L., Carcano, G., & Rovera, F. (2020). Machine Learning techniques in breast cancer prognosis prediction: A primary evaluation. *Cancer medicine*, *9*(9), 3234-3243.

[40] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, 1064-1069.

[41] Zheng, X., Shi, J., Li, Y., Liu, X., & Zhang, Q. (2016, April). Multi-modality stacked deep polynomial network based feature learning for Alzheimer's disease diagnosis. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)* (pp. 851-854). IEEE.

[42] Sprague, B. L., Conant, E. F., Onega, T., Garcia, M. P., Beaber, E. F., Herschorn, S. D., ... & PROSPR Consortium*. (2016). Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. *Annals of internal medicine*, *165*(7), 457-464.

[43] Gemici, A. A., Arıbal, E., Özaydın, A. N., Gürdal, S. Ö., Özçınar, B., Cabioğlu, N., & Özmen, V. (2020). Comparison of qualitative and volumetric assessments of breast density and analyses of breast compression parameters and breast volume of women in Bahcesehir mammography screening project. *European Journal of Breast Health*, *16*(2), 110.

[44] Lehman, C. D., Yala, A., Schuster, T., Dontchos, B., Bahl, M., Swanson, K., & Barzilay, R. (2019). Mammographic breast density assessment using deep learning: clinical implementation. *Radiology*, *290*(1), 52-58.

[45] Mohamed, A. A., Berg, W. A., Peng, H., Luo, Y., Jankowitz, R. C., & Wu, S. (2018). A deep learning method for classifying mammographic breast density categories. *Medical physics*, *45*(1), 314-321.

[46] Dhungel, N., Carneiro, G., & Bradley, A. P. (2017, April). Fully automated classification of mammograms using deep residual neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 310-314). IEEE.

[47] Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Computer methods and programs in biomedicine*, *157*, 19-30.

[48] Jiao, Z., Gao, X., Wang, Y., & Li, J. (2016). A deep feature based framework for breast masses classification. *Neurocomputing*, *197*, 221-231.

[49] De la Torre Gallart, J. (2019). Diabetic retinopathy classification and interpretation using deep learning techniques.

[50] Mushtaq, Z., Yaqub, A., Hassan, A., & Su, S. F. (2019, February). Performance analysis of supervised classifiers using PCA based techniques on breast cancer. In *2019 international conference on engineering and emerging technologies (ICEET)* (pp. 1-6). IEEE.

[51] Shajihan, N. (2020). Classification of stages of Diabetic Retinopathy using Deep Learning. *Bournemouth University United Kingdom*.

[52] Li, K., & Wang, P. (2013). A New Extension of LBP for Texture Classification. In *Intelligence Computation and Evolutionary Computation: Results of 2012 International Conference of Intelligence Computation and Evolutionary Computation ICEC 2012 Held July 7, 2012 in Wuhan, China* (pp. 7-12). Springer Berlin Heidelberg.

[53] Li, Z. M., Huang, Z. H., & Zhang, T. (2016). Gabor-scale binary pattern for face recognition. *International Journal of Wavelets, Multiresolution and Information Processing*, *14*(05), 1650035.

[54] CBIS-DDSM: Breast Cancer Image Dataset. (n.d.). CBIS-DDSM: Breast Cancer Image Dataset | Kaggle. https:///datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset

[55] Mukherjee, S. (2022, August 18). *The Annotated ResNet-50*. Medium.