DEEP LEARNING AND INFORMATION FUSION BASED CLASSIFICATION OF BREAST CANCER IN MAMMOGRAPHY IMAGES

By

DINA AHMED RAGAB

Ph.D. Thesis Submission

UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

Supervisors

Prof. Stephen Marshall

Director of Institute for Sensors, Signals, and Communications (InstSCC),
University of Strathclyde, Glasgow,
United Kingdom

Prof. Jinchang Ren

National Subsea Centre,
Robert Gordon University, Aberdeen,
United Kingdom

Prof. Maha Sharkas

Professor, and Head of Electronics and Communications Dept.
Arab Academy for Science, Technology and Maritime Transport (AASTMT), Alexandria,
Egypt

Acknowledgments

First, praises and thanks to **ALLAH**, the Almighty, for His showers of blessings throughout my research work to complete it successfully.

I would like to express my deepest thanks and sincere gratitude to **Prof. Stephen Marshall, Prof. Jinchang Ren,** and **Prof. Maha Sharkas** for their supervision, patient guidance, encouragement, advice, motivation, unlimited effort, and support to conduct this work and illustrate it in quite a vivid manner. I have been extremely lucky to have supervisors who cared so much about my work.

My sincere thanks go to my friend **Dr. Omneya Attallah,** Assistant Professor of Electronics and Communications Engineering Department at the AASTMT, for her great help, support, and unlimited effort.

I would like to thank the **Armed Forces Hospital** in **Egypt** and especially **Dr. Gehan Roushdy** for providing us a detailed mammography dataset to be used in this work. Without her precious support, it would not have been possible to conduct this research.

I am extremely grateful to **Dr. Saleh Eisa**, Head of Electronics and Communication Engineering Department, AASTMT Smart Village campus for his endless support.

I wish to express my deepest gratitude to all my colleagues and friends at AASTMT.

Special thanks to my mother, my brother, and my family for their love, prayers, caring, and continuous backup.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

Abstract

Breast cancer is one of the most dangerous diseases that can afflict especially women. Computer-aided diagnosis (CADx/CAD) systems may help radiologists in fast and accurate decision making to detect early stage of breast cancer and reduce mortality. By fusion of different techniques in image processing, machine learning (ML), and deep learning (DL), several techniques are proposed in this thesis for analysing breast cancer images, aiming to classify normal/ abnormal lesions in mammography images. The classification method is generally implemented using ML classifiers and DL methods based on the deep convolutional neural networks (DCNN).

This thesis presents new methods, assembled on recently developed DL models based on different fusion techniques to develop three different frameworks. The first framework presents an approach for segmenting the region of interest (ROI) followed by classifying it using DCNN. In the second framework, a fusion-based novel approach is presented to classify the mammogram lesions using several DCNNs. Additionally, it employs some ML techniques to improve the classification accuracy. However, the third framework introduces the decision-level fusion rather than feature fusion forming a second stage classification to improve the classification accuracy as well.

Overall, the methods proposed in this thesis achieved promising classification accuracy results that improve the performance of the state-of-the-art approaches and may help to improve the diagnosis of breast cancer.

The methodologies presented in this work are evaluated on several publicly available datasets, including the digital database for screening mammography (DDSM), the curated breast imaging subset of DDSM (CBIS-DDSM), and the mammographic image analysis society digital mammogram dataset (MIAS). Considering their limitations, a new mammogram dataset, “DAR-Breast,” is collected from the Armed Forces Hospital in Egypt, the first such in Egypt, to benefit the advancement in this area.

Contents

List of Figures

Chapter 1: Introduction

Chapter 2: Introduction to Breast Cancer

Chapter 3: Computer-Aided Diagnosis: Datasets and State-of-the-Art

Chapter 4: AlexNet-SVM: A Framework for Breast Cancer Segmentation and Classification

Chapter 5: Deep Features Fusion Framework for Breast Cancer Classification

Chapter 6: A Breast Cancer Classification Framework Based on the Decision-Levels Fusion using Multiple DCNNs

Appendix A: State-Of-The-Art DCNN Architectures

List of Tables

Chapter 2: Introduction to Breast Cancer

Chapter 3: Computer-Aided Diagnosis: Datasets and State-of-the-Art

Chapter 4: AlexNet-SVM: A Framework for Breast Cancer Segmentation and Classification

Chapter 5: Deep Features Fusion Framework for Breast Cancer Classification

Chapter 6: A Breast Cancer Classification Framework Based on the Decision-Levels Fusion using Multiple DCNNs

Appendix A: State-Of-The-Art DCNN Architectures

List of Abbreviations

ACR	American College of Radiology
AdaBoosting	Adaptive boosting
AHE	Adaptive Histogram Equalization
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANOVA	One-way analysis of variance
ASR	Age Standardized Incidence Rate
AUC	Area Under the ROC Curve
BCDR-F03	Breast Cancer Digital Repository Film Mammography dataset #3
BI-RADS	Breast Imaging-Reporting and Data System
BPNN	Back-Propagation multilayer Neural Network
CADe	Computer-Aided Detection
CADx/CAD	Computer-Aided Diagnosis
CBIS-DDSM	Curated Breast Imaging Subset of DDSM
CC	Craniocaudal view
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNNI- BCC	Convolutional Neural Network Improvement for Breast Cancer Classification
CNN	Convolutional Neural Network
CNN-CT	Convolutional Neural Network Curvelet Transform
CNN-DW	Convolutional Neural Network Discrete Wavelet
CT	Computed Tomography
DBT	Digital Breast Tomosynthesis
DCA	Discriminant Correlation Analysis
DCIS	Ductal Carcinoma In Situ
DCNN	Deep Convolutional Neural Networks
DDSM	Digital Database for Screening Mammography
DICOM	Digital Imaging and Communication in Medicine
DL	Deep Learning
DT	Decision Tree
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machine
ERGS	Effective Range-based Gen Selection
Faster-RCNN	Faster Region-based Convolutional Neural Network
FC	Fully Connected layer
FCRN	Fully Complex-valued Relaxation Network Classifier
FFDM	Full-Field Digital Mammogram
FN	False Negative
FP	False Positive
FNR	False Negative Rate
FPR	False Positive Rate
GANs	Generative Adversarial Networks
GB	Gradient Boosting
GLCM	Grey Level Co-occurrence Matrix
GNB	Gaussian Naïve Bayes

GPUs	Graphic Processing Units
GRNN	General Regression Neural Network
GT	Ground Truth
HOG	Histogram of Oriented Gradient
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IRB	Institutional Review Board
IRMA	Image Retrieval in Medical Applications
JPEG	Joint Photographic Experts Group
K-NN	K-Nearest Neighbour
LBP	Local Binary Pattern
LCIS	Lobular Carcinoma In Situ
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LSTM	Long Short-Term Memory
MC	Microcalcifications
MCS	Multiple Classifier System
MIAS	Mammographic Image Analysis Society Digital Mammogram Database
ML	Machine Learning
MLO	Mediolateral-Oblique view
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MVFF	Multi-View Feature Fusion
NB	Naïve Bayes
NBSS	National Breast Screening System
NCRP	National Cancer Registry Program
OMI-DB	OPTIMAM Database
PCA	Principal Component Analysis
PEIPA	Pilot European Image Processing Archive
PET	Positron Emission Tomography
PGM	Portable Grey Map
RAM	Random Access Memory
ResNet	Residual Network
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver-Operating Curve
ROI	Region of Interest
RT	Random Tree
SEER	Surveillance, Epidemiology, and End Results program
SFS	Sequential Forward Search
SGDM	Stochastic Gradient Descent with Momentum
SIFT	Scale-Invariant Feature Transform
SMO	Sequential Minimal Optimization
SOM	Self-Organizing Map
Std	Standard Deviation

SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TNR	True Negative Rate
TPR	True Positive Rate
WBCD	Wisconsin Breast Cancer Dataset
WDBC	Wisconsin Diagnostic Breast Cancer
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting
YOLO	You-Only-Look-Once

Chapter 1

Introduction

1.1 Background and Motivation

Breast cancer is considered a crucial problem among women all over the world [1]. It is one of the most common kinds of cancer, as well as the leading cause of mortality among women [1]. However, early detection and diagnosis of breast cancer increase the chances for successful treatment and complete recovery for the patient [2]. There are many imaging modalities for detecting breast cancer among them ultrasound, computed tomography (CT), mammogram, magnetic resonance imaging (MRI), thermography, nuclear medicine, and positron emission tomography (PET) [3], [4]. However, mammography is the primary recommended imaging modality for breast cancer screening [5]. This is because breast cancer has a known asymptomatic phase that can be detected with mammography [5]. Moreover, mammography is considered a useful early detection tool before the appearance of the physical symptoms [5]. However, even though a mammogram may show suspicious regions, it cannot prove that an abnormal area is cancerous. If a mammogram presents a suspicion of cancer, a biopsy should be performed [4]. The most important abnormalities that could be found on an image modality are masses and microcalcifications (MC) [6].

Manual readings of mammograms may result in misdiagnosis due to human errors caused by visual fatigue [7]. Consequently, in the last three decades, various computer-aided diagnosis (CADx/CAD) systems have been proposed for breast cancer diagnosis using computer vision and machine learning (ML) techniques [8]. CAD systems aim at giving a second objective opinion to help the radiologist for clarification and diagnosis [7]. Moreover, they are used as applications that differentiate between different types of tumours [9]. A lesion could be classified as either normal or abnormal [9]. Additionally, abnormal lesions could be benign or malignant [9]. The advances in breast cancer diagnosis and CAD systems have led to a (30% – 50%) fall in mortality in several countries [2].

Over the years, artificial intelligence (AI) has been widely used in clinical cancer research due to its feasibility and advantages [10], [11]. The concept of AI emerged first in 1956; however, the earliest work in medical AI dates to early 1970 [12]. The main aim of AI is to build machines that can think and reason over complex tasks just like human beings and

thereby sharing the same essential cognitive characteristics [13]. Meanwhile, the field of AI has made many developments as gradually AI implementation became real in scientific research laboratories [12], [13].

The primary methodology behind the AI is machine learning (ML) [14] as shown in Figure 1.1. There are many categories of ML methods, such as unsupervised and supervised ML [12], [15]. The unsupervised ML discovers patterns from a dataset that has no labelled samples for training the model. Instead, they cluster the samples by either the similarity or distance between their features [12]. Conversely, for the supervised ML, the dataset contains labelled samples to train the model by minimising the prediction error of the training samples [12]. Most applications of ML rely on supervised ML, which is used extensively in medical applications due to its remarkable capability for extracting valuable information from medical datasets [16].

On the other hand, deep learning (DL) is the fastest-growing field of ML [17], which has emerged as the state-of-the-art ML method in many applications [18] as shown in Figure 1.1. DL is a type of representation and learning method in which a complex multilayer neural network architecture is applied to learn the representations of data automatically by transforming the input information into multiple levels of abstractions [19]. The advantage of the DL is that it uses effective unsupervised or semi-supervised feature learning and layered feature extraction instead of hand-crafted feature extraction [20]. The major aim of feature learning is to seek for better representation of data and to produce a more effective model to learn these representations from large-scale dataset. The representation is similar to the development of the real neural networks, which is based on the understanding of how information is processed and transmitted in neural-alike systems [21].

Generally, research on deep learning for medical images focuses mainly on four aspects: Structure detection, segmentation, labelling and captioning, and computer-aided detection or diagnosis [21], [22]. For pattern recognition tasks the deep convolutional neural networks (DCNN) are currently widely used for image analysis and data classification [21]. There are many DCNNs architectures proposed so far, where some of the state-of-the-art architectures are used in this thesis, including the AlexNet [23], GoogleNet [24], Inception-v3 [25], the ResNet [26], Inception ResNet-v2 [27], and MobileNet-v2 [28].

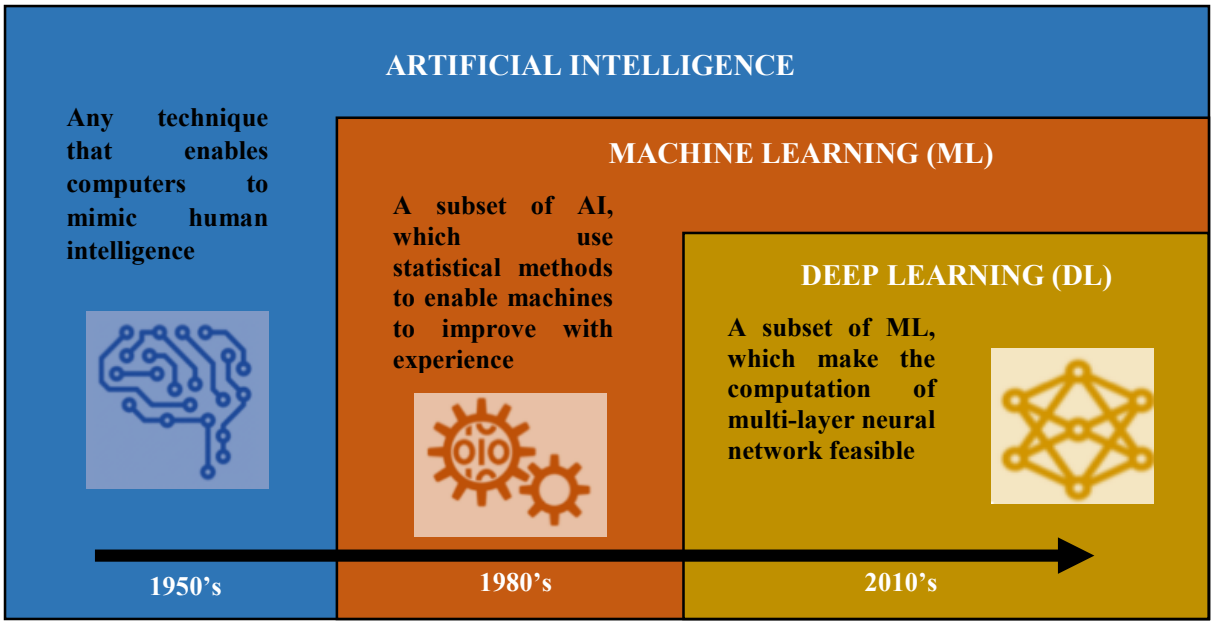


Figure 1.1: A schematic diagram showing the difference between AI, ML, and DL. Adapted from [14].

Furthermore, the transfer learning technique [29] is introduced in this thesis. This is a technique performed on DCNN networks so that it could be used in any classification problem [30]. The DCNNs networks are designed to classify 1000 classes. However, in this thesis, it is required to classify two classes. Therefore, this is accomplished by replacing the last fully connected layer in any network with a new layer for the classification of two classes [31].

1.2 Problem Statement, Aims, and Objectives

1.2.1 Problem Statement

Breast cancer is one of the most prevalent cancer types in the world. Both men and women have been diagnosed with this form of cancer, but the prevalence of women is far beyond reference [1]. Early diagnosis can prevent the disease from developing into becoming lethal [2]. In addition, it has been reported that mammography is classified as the most popular and easiest method of breast cancer detection in its early stages [32]. Currently the radiologists can predict breast cancer from mammography images with an accuracy over 90% [32]. Nevertheless, about 10% of breast cancer can be missed, and this can be further reduced by double-checking and reading of the mammography images [32].

Double-checking requires the same mammogram to be analysed by two different radiologists at different times. Although it has been shown to increase the accuracy of correct detection by up to 15% in comparison with single checking, however, this is a time consuming and costly procedure. Therefore, CAD systems are desired to cut the cost for accurate

diagnosis. This diagnosis contains several methods and techniques, including dataset collections, data analysis, image processing, and ML/ DL based decision making [9].

To tackle these issues, one proposed solution is the use of computer vision and image processing for feature extraction, followed by ML and DL algorithms to assist in the diagnosis process. Therefore, in this thesis, several DL based frameworks have been developed.

1.2.2 Thesis Aims

The work presented in this thesis aims to develop computer method for automatic segmentation and classification of different breast cancer lesions, which can act as a second opinion for assisting the radiologists to classify the breast cancer in mammography images. We aim to provide more accurate and efficient analysis to improve the classification accuracy of the state-of-the-art methods.

1.2.3 Thesis Objectives

In this thesis, new diagnostic tools will be developed, for analysing breast cancer in mammography images. These techniques are presented in three proposed frameworks respectively to tackle three major tasks in mammography analysis, i.e. detection, segmentation, and classification of cancer lesions. The objectives of the work in this thesis are summarised as follows:

1. To develop effective tools for diagnosing breast cancer lesions from mammography images.
2. To classify breast lesions into normal or abnormal in addition to benign and malignant images. To carry out this, we developed various pattern recognition techniques by proposing diagnosis frameworks based on features and decision fusion.
3. To further refine the classification accuracy of existing DL models and derive the best possible classifiers, using decision level fusion.
4. To collect and publicly share a new high-resolution dataset of mammogram images from Egypt in high resolution — labelled by radiologists.

1.3 Contributions

In this thesis, three novel DCNN based frameworks are developed and presented to differentiate between different types of breast cancer tumours in mammograms. The main contributions of this thesis are summarised as follows:

1. In the first framework, two methodologies are used to segment the region of interest (ROI), i.e., the circular contours and adaptive thresholding. The mass ROIs of breast lesions are then classified using an end-to-end fine-tuned AlexNet architecture. Additionally, to improve the classification accuracy, the deep features are extracted and classified using the SVM classifier. Detail of the hybrid approach is presented in Chapter 4.

2. In the second framework, a fusion-based novel approach is presented to classify the mammogram lesions. First, features are extracted and classified using several fine-tuned DCNN architectures including the AlexNet, GoogleNet, ResNet-18, ResNet-50, and ResNet-101. These extracted deep features are applied separately to construct SVM classifiers with different kernels. Afterward, the extracted deep features are ranked to form four feature sets that include a different combination of deep features. Finally, the principal component analysis (PCA) technique is used to reduce the feature space and the computational cost of the four feature sets. This methodology is presented in Chapter 5.

3. In the third framework, another new fusion technique is introduced for decision-level fusion rather than feature fusion. Based on the outcomes of the individual classifiers, a second stage classification is employed for improving the classification accuracy as detailed in Chapter 6.

4. Since there are few publicly available high-resolution mammogram datasets in this area, therefore, a useful dataset is collected in this thesis. Mammogram scans annotated by expert radiologists were collected from the Armed Forces Hospital located in Alexandria, Egypt, namely “DAR-Breast” dataset. To the best of our knowledge, this is the first time mammogram breast cancer scans are acquired and analysed for the purpose of breast cancer diagnosis in Egypt. The results of the third framework presented in Chapter 6 is evaluated using “DAR-Breast” dataset.

1.4 Publications Arising from this Thesis

The work presented in this thesis has generated the following research outputs:

- D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," PeerJ, vol. 7, p. e6201, 2019 doi:10.7717/peerj.6201 [33].
- D. A. Ragab, O. Attallah, M. Sharkas, J. Ren, and S. Marshall, "A Framework for Breast Cancer Classification using Multi-DCNNs," Computers in Biology and Medicine, vol. 131, p. 104245, April 2021, https://doi.org/10.1016/j.compbiomed.2021.104245 [34].

1.5 Thesis Structure

This thesis is composed of seven chapters. Chapter 1 is the overall introduction of the thesis, which covers the motivation and summarised the contributions of the work as well as the research aims and objectives. In Chapter 2, an overview of breast cancer is provided along with different imaging modalities used in breast cancer screening. In the following Chapter 3, a description of the conventional CAD systems and the major components are presented, along with an introduction to the DCNN. In addition, a description of the available mammogram datasets is also introduced, followed by a comprehensive overview of the results derived on these datasets from the state-of-the-art ML and DCNNs based CAD systems.

The three frameworks proposed in this thesis are detailed in Chapters 4-6, including the methodology, introduction, experimental setup, and the results. In Chapter 4, two methodologies are used to segment the ROI, i.e., the circular contours and adaptive thresholding. The mass ROIs of breast lesions are then classified using an end-to-end fine-tuned AlexNet architecture. Then, the deep features are extracted and classified using the SVM classifier to improve the classification accuracy. In Chapter 5, a fusion-based novel approach is presented. This is performed by extracting, fusing, and classifying the deep features using SVM classifiers with different kernels. Then the PCA technique is applied to reduce the feature space and the computational cost. In Chapter 6, another new fusion technique is introduced for decision-level fusion rather than feature fusion. Finally, in Chapter 7, the thesis is concluded with summarised contributions and directions for future investigation.

Chapter 2

Introduction to Breast Cancer

2.1 Introduction

For years, cancer has been one of the biggest threats to human life [1]; it is expected to become the leading cause of death over the next few decades. Based on statistics from the World Health Organization (WHO), deaths caused by cancer are expected to increase in the future [35]. Figure 2.1 shows the estimated number of incident cancer cases located in women of all ages until the year 2020 [36]. According to Figure 2.1 and as reported from WHO it was clear that breast cancer ranked in the first place with respect to other organs for women [36].

In the following sections, a brief overview of the incidence of breast cancer worldwide and especially in Egypt. In addition, breast anatomy is presented, and then breast cancer is explained, including how the cancer is produced, the types of breast cancer and tumours, stages of cancer, and the types of abnormalities. Furthermore, radiology-screening methods are discussed.

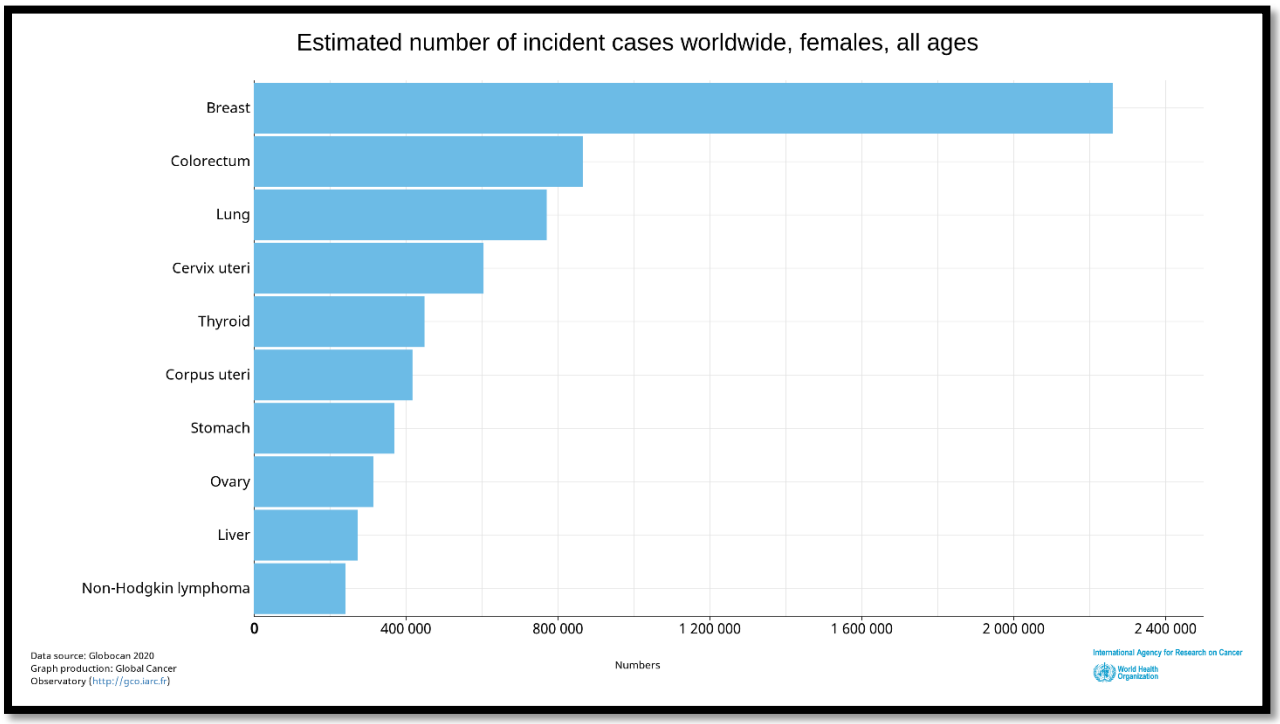


Figure 2.1: The estimated number of cancer incidence locations in females worldwide*. Adopted from [36].

** The figure is online at https://gco.iarc.fr/today/online-analysis-multi-bars*

2.2 Breast Cancer Incidence

Breast cancer is one of the most common kinds of cancer, as well as the leading cause of mortality among women, especially in developed and underdeveloped countries. Moreover, it is the second most common cancer in the world and by far the most frequent cancer among women [37] as discussed in the following sub-sections.

2.2.1 Breast Cancer Worldwide

Incidence rates vary nearly fourfold across the world regions, with rates ranging from 27 per 100,000 in Middle Africa and Eastern Asia to 96 in Western Europe [2]. In the USA about 19% of breast cancers are diagnosed in women ages 30 – 49 years, and 44% occur among women who are age 65 years or older [2]. In Japan, breast cancer in women aged younger than 35 years old comprises approximately 3% of Japanese breast cancer patients. In most African countries, breast cancer among young women comprises a high proportion of cases than among older women [2]. This is a demography-driven phenomenon rather than a true intrinsic biological significance because the African population has a low median age; generally 20 years and below [2].

Nevertheless, in North Africa, the incidence among women aged 15 – 49 is lower than in Western countries, but the very low incidence among women aged more than 50, combined with the young age pyramid of North-Africa, makes the relative proportions of young patients substantially higher (50 – 60% versus 20% in France) [2]. Such epidemiological features result mainly from peculiar risk factor profiles, which are typical of many developing countries and include notably rapid changes in reproductive behaviours [1]. In Arab women, the average age at presentation of breast cancer is a decade earlier than in the US and European women'. The median age at diagnosis in Arab populations is about 48 years, and about two-thirds of women with breast cancer are younger than 50 years [37]. Figure 2.2 shows the incidence versus the mortality of breast cancer among women of all ages for the six continents worldwide [36].

2.2.2 Breast Cancer in Egypt

Cancer is an increasing problem in Egypt. The commonest sites were liver (23.8 %), breast (15.4 %), and bladder (6.9 %) for both sexes, liver (33.6 %) and bladder (10.7 %) among men, and breast (32.0 %) and liver (13.5 %) among women [38]. The percentage of the most frequent types of cancer for women that occur in Egypt as reported from the National Cancer Registry Program (NCRP) is displayed in Table 2.1.

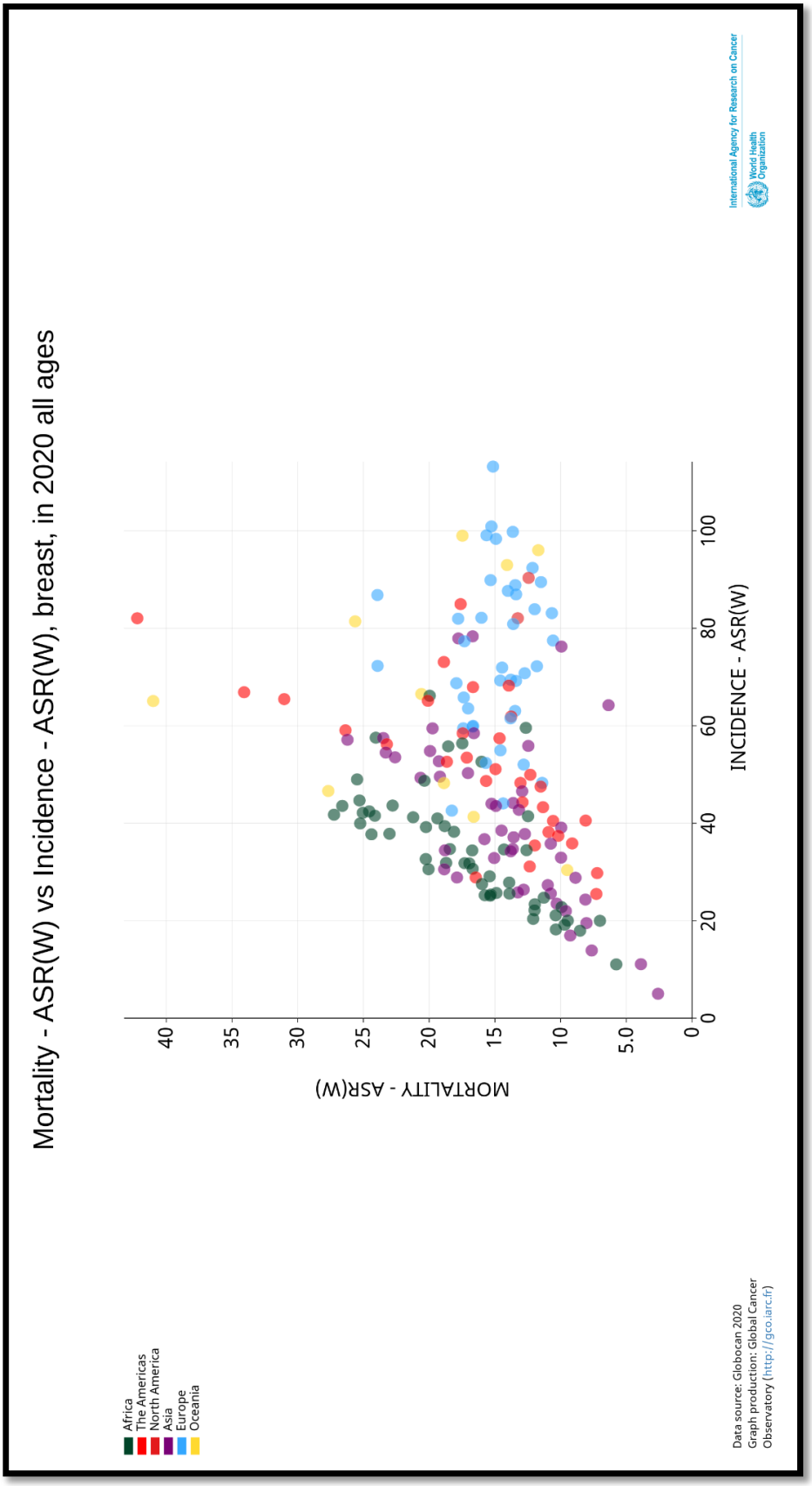


Figure 2.2: Incidence vs. Mortality for breast cancer worldwide*. Adopted from [36].

* *The figure is online at https://gco.iarc.fr/today/online-analysis-scatter-plot*

Table 2.1: The most frequent cancer sites for women in Egypt based on the results of the NCRP. Adapted from [39].

Cancer Site	Percentage (%)	Age Standardized Incidence Rate (ASR)
Breast	35.1	48.8
Liver	13.54	24.4
Brain	5.17	8.0
Ovary	4.12	6.3
Thyroid	3.28	4.3

As it is obvious from Table 2.1, the highest percentage of cancer in women is breast cancer. It is rated the highest cancers, among other sites accounting 35.1% [39]. Egypt is expected to experience the highest increase in the coming two decades [38]. Moreover, according to the study in [39], the authors estimated that the number of breast cancer cases in Egypt will increase gradually until the year 2050.

2.3 Breast Cancer

2.3.1 Breast Anatomy

The breast is composed of different tissue, ranging from very fatty tissue to very dense tissue. It is mainly composed of three types of tissues, breast fat (or adipose tissue), glandular tissue, and connective tissue [40]. The proportions of these main types of tissue may vary from one woman to another. Within this tissue is a network of lobes. Each lobe is composed of tiny tube-like structures called lobules containing milk glands. Tiny ducts connect the glands, lobules, and lobes, carrying milk from the lobes to the nipple [40]. The nipple is located in the middle of the areola, which is the darker area that surrounds the nipple. Blood and lymph vessels also run throughout the breast. Blood nourishes the cells, while the lymph system drains bodily waste products [40]. The lymph vessels connect to lymph nodes, the tiny, bean-shaped organs that help fight infection [40]. The amount of water, fat, and fibro-glandular tissue may also vary due to normal hormonal changes in different stages [41]. Breast anatomy is shown in Figure 2.3.

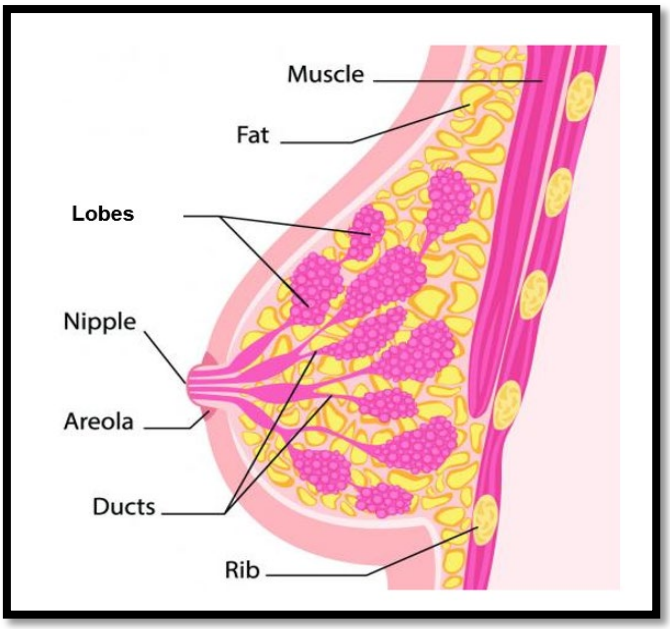


Figure 2.3: A labelled breast anatomy. Adopted from [42].

2.3.2 Types of Breast Cancer

Cancer is characterized by abnormal cells that grow and invade healthy cells in the body. Breast cancer starts in the cells of the breast as a group of cancer cells that can then invade surrounding tissues or spread to other areas of the body [43]. Figure 2.4 shows the steps of cancer cell reproduction. Cancer begins in the cells, which are the basic building blocks that form the tissue. Sometimes, the process of cell growth goes wrong and new cells form; when this occurs; a build-up of cells often forms a mass of tissue called a lump or tumour [43].

Breast cancer occurs when malignant tumours develop in the breast. These cells can spread by breaking away from the original tumour and entering blood vessels or lymph vessels, which branch into tissues throughout the body. When cancer cells travel to other parts of the body and begin damaging other tissues and organs, the process is called metastasis [43].

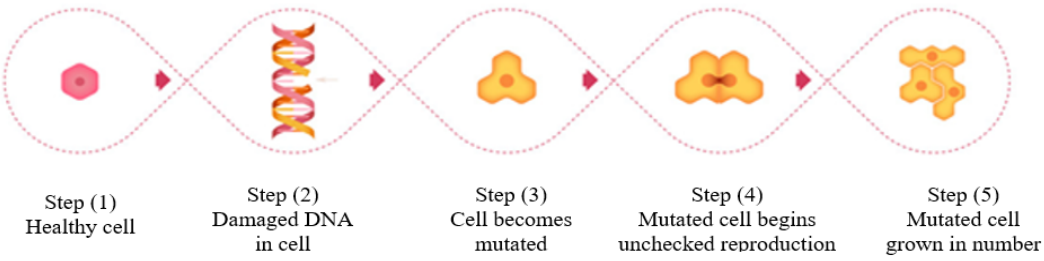


Figure 2.4: The steps of cancer cell reproduction. Adopted from [43].

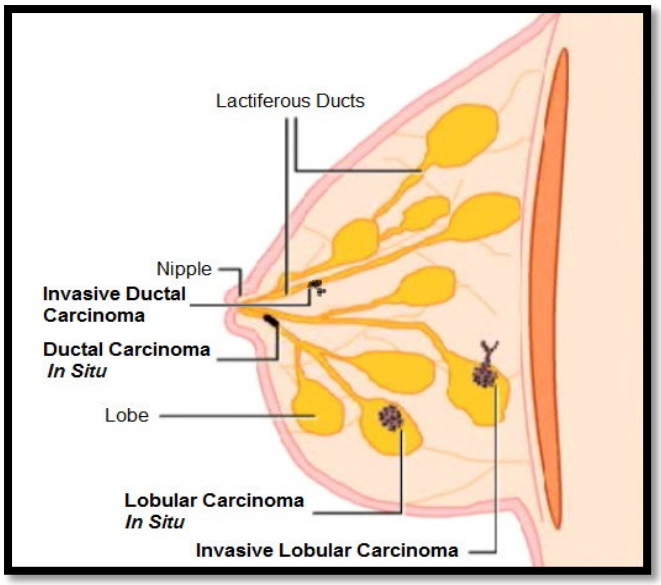


Figure 2.5: The most common types of breast cancer. Adopted from [41].

A cancer that occurs in the breast can be one of two most common types: in situ (or non-invasive) and invasive. In situ cancers are those in which cancer cells remain within the basement membrane of the lobules and the draining lactiferous duct. Therefore, the ductal carcinoma in situ (DCIS) is a type of cancer in which cancerous cells are inside some of the ducts, but have not spread to other regions of the breast or body [44], [45]. However, the lobular carcinoma in situ (LCIS) is not a type of cancer, but in the presence of this disease, there are high chances of developing cancer [46]. LCIS is characterized by changes in the cells within the breast lobes as shown in Figure 2.5.

On the other hand, invasive cancers are those, in which the cancer cells spread outside the basement membrane of the ducts and lobules into the surrounding adjacent normal tissue [41]. Thus, invasive ductal carcinoma (IDC) is the most common type of breast cancer; it represents about 70% to 80% of breast cancer cases and it occurs in the cells that line the ducts of the breast [47]. In addition, invasive lobular carcinoma (ILC) [48] represents about 10% of breast cancer cases and occurs in the cells that line the lobules of the breast as shown in Figure 2.5. Furthermore, another type of invasive breast cancer, which is inflammatory breast cancer. It is considered an uncommon type of breast cancer, and it accounts for about 1% to 5% of all breast cancers [49].

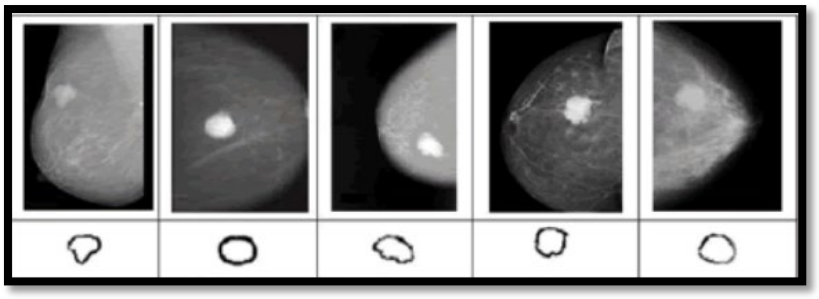


Figure 2.6: Samples for benign breast cancer tumours. Adopted from [50].

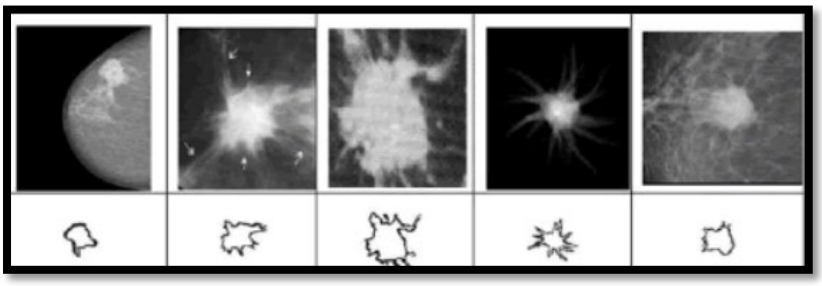


Figure 2.7: Samples for malignant breast cancer tumours. Adopted from [50].

2.3.3 Types of Tumours

A tumour is a mass of abnormal tissue, it is either benign or malignant [43].

1) Benign Tumours

The cells from benign tumours do not spread to other parts of the body. These cells are not life-threatening [43]. When a tumour is diagnosed as benign, doctors will usually leave it alone rather than remove it. Even though these tumours are not generally aggressive towards surrounding tissue, occasionally they may continue to grow, pressing on organs and causing pain or other problems. In these situations, the tumour is removed, allowing pain or complications to subside [43]. Samples of benign tumours are shown in Figure 2.6.

2) Malignant Tumours

They are cancerous and aggressive because they invade and damage the surrounding tissues and organs [43]. They can spread to other parts of the body. When a tumour is suspected to be malignant, the doctor will perform a biopsy to determine the severity or aggressiveness of the tumour [43]. Samples of malignant tumours are shown in Figure 2.7.

2.3.4 Stages of Breast Cancer

Breast cancer has been divided into five stages according to the size and location of the tumour as demonstrated in Table 2.2 [51]–[53].

Table 2.2: The stages of breast cancer. Adapted from [51].

Stages	Size of Tumour	Location
Zero	-	Cells have not spread outside the ducts or lobules
One	< 2 cm in diameter	Did not spread to the lymph nodes or outside the breast
Two	5 cm	Has not spread to the lymph nodes
	2 cm	Spread to the lymph nodes
Three	> 5 cm	Spread to lymph nodes or other tissue near the breast
Four	-	Cancer has spread to other organs or tissues, such as liver, lungs, brain, or the skeletal system

Table 2.3: BI-RADS scores description. Adapted from [54].

BI-RADS Scores	Description
0	Needs additional imaging evaluation
1	Negative – No abnormalities
2	Benign finding(s)
3	Benign finding(s) – Short-interval follow-up is suggested.
4	Suspicious anomaly – Biopsy should be considered
5	Highly suggestive of malignancy – Appropriate action should be taken
6	Biopsy proven malignancy

2.4 Breast Cancer Abnormalities

When radiologists interpret mammograms, they look for specific abnormalities. The most common findings seen on mammography are masses, calcifications, architectural distortion of breast tissue, and asymmetries [55]. Furthermore, they assign a score to the abnormalities, which is used to communicate with doctors about how they are concerned about the findings. To standardize the terminology of the mammography report, the assessment of findings, and the recommendation of action to be taken, the American College of Radiology (ACR) has developed the Breast Imaging Reporting and Data System (BI-RADS) score [56]. The BI-RADS score ranges from '0' to '6' based on the level of suspicion [54]. The scores of BI-RADS are tabulated in Table 2.3.

The masses and calcifications are considered an important early signs of the disease, although the other signs are less important, as the architectural distortions and asymmetries are similar in appearance to mass, and can often be confused by the radiologists to be masses [55]. Therefore, in the following sub-sections, the masses and calcifications will be discussed.

2.4.1 Mass

A mass is defined as a space-occupying lesion seen in more than one projection. It is usually characterized by its shape and margin [57]. Masses are quite subtle and often occurring in the dense areas of the breast tissue. They have smoother boundaries than calcifications, various shapes such as round, oval, and lobular. Moreover, they have different margins such as

circumscribed, speculated, and ill-defined [6]. A mass with a regular shape and well-circumscribed boundary has a higher probability of being benign; on the other hand, a malignant mass usually has an irregular shape and a speculated, rough, and blurry boundary [57]. Figure 2.8 shows a malignant mass sample extracted from the DAR-Breast dataset.

2.4.2 Calcifications

Oppositely, calcifications are tiny calcium deposits that have accumulated in the breast tissue. They appear as small white spots with a diameter of less than 0.5 mm. Calcifications are often difficult to detect since they are very small, have an inhomogeneous background, and have low contrast with the background [58]. Benign calcifications are usually larger and coarser with round and smooth contours. However, malignant calcifications tend to be numerous, clustered, small, varying in size and shape, angular, irregularly shaped, and branching in orientation [6]. Figure 2.9 shows an example of a malignant calcifications sample extracted from DAR-Breast dataset. There are two types of calcifications, microcalcifications (MC) they can be benign or malignant. However, macrocalcifications are often benign.

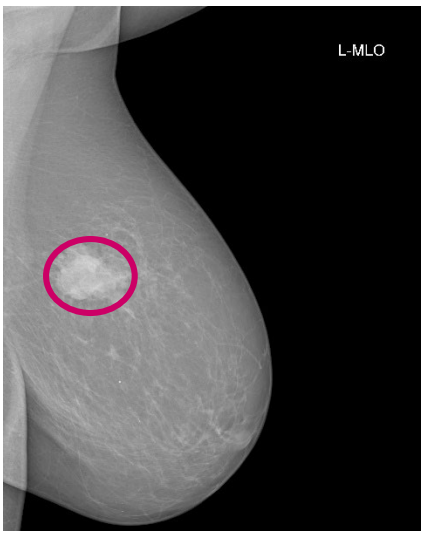


Figure 2.8: A malignant mass sample contoured with a circle extracted from the DAR-Breast dataset.

Figure 2.9: A clustered malignant calcification sample contoured with a circle extracted from the DAR-Breast dataset.

2.5 Breast Cancer Screening Methods

There are various types of screening devices for breast cancer, such as breast self-examination, breast ultrasound, computed tomography (CT), mammogram, magnetic resonance imaging (MRI), thermography, positron emission tomography (PET), and nuclear medicine imaging [3], [4].

2.5.1 Breast Self-Examination

Breast self-examination is a screening method used in an attempt to detect early breast cancer. The method involves the woman herself looking at and feeling each breast for possible lumps, distortions, or swelling [59]. Adult women of all ages are encouraged to perform breast self-exams at least once a month. The test could be performed in the shower, in front of a mirror, or lying down [60].

Steps for Breast Self-Examination:

The breast self-examination consists of five steps as summarised in Figure 2.10 [59];

1) Look at your breasts in the mirror with your shoulders straight and your arms on your hips. You have to check that breasts are in their usual size, shape, and colour, and they are evenly shaped without visible distortion or swelling.
2) Raise your arms over your head and look for the same changes.
3) With your arms still over your head, look for any signs of fluid coming out of any of both nipples.
4) Feel your breasts while lying down, using your right hand to feel your left breast and vice versa. Use a firm, smooth touch with the first few finger pads of your hand to complete an outer circle. Then move in 1 inch towards the nipple and complete another circle, and keep repeating until you reach the nipple. Follow a pattern to be sure that you cover the whole breast. You are looking for a lump.
5) Feel your breasts in the shower, as it is easier when the skin is wet and slippery. Then cover the same hand moments described in step 4.

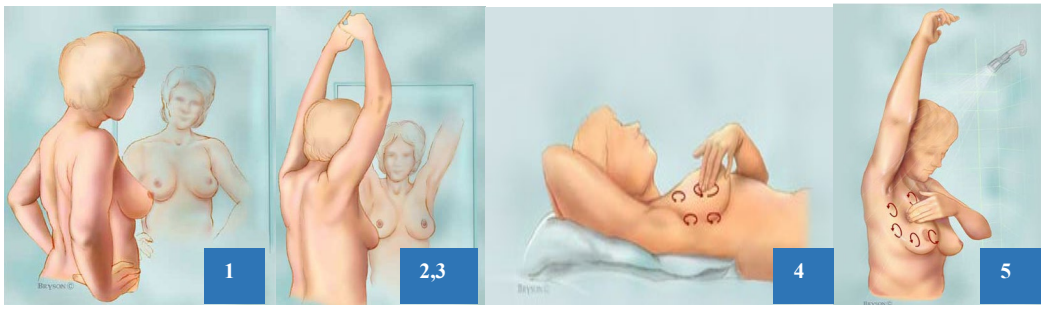


Figure 2.10: The steps of breast self-examination. Adopted from [59].

Figure 2.11: An ultrasound image of a breast mass. Adopted from [61].

2.5.2 Breast Ultrasound

Breast ultrasound uses sound waves to make a computer picture of the inside of the breast [49]. Ultrasound is useful for looking at some breast changes, such as lumps or changes in women with dense breast tissue [49]. It can differentiate between fluid-filled cysts which are very unlikely to be cancer and solid masses which might need further testing to be sure they are not cancer [49]. Additionally, ultrasound can be used as a helpful guide for the biopsy needle into an area so that cells can be taken out and tested for cancer [49]. Figure 2.11 shows an ultrasound image of a breast mass.

During screening breast ultrasound, a gel is applied to the skin of the breast, and a wand-like instrument called a transducer is moved over the skin [49]. The transducer sends out sound waves and picks up the echoes as they bounce off body tissues. The echoes are made into a picture on a computer screen [49]. The patient might feel some pressure as the transducer is moved across the breast, but it should not be painful [49]. Ultrasound has some advantages such as it is widely available, easy to have, and does not expose a person to radiation. It also costs less than a lot of other options [49].

2.5.3 Mammography

Mammography is one of the most reliable methods for early detection of breast carcinomas [62]. However, it is difficult for radiologists to provide both accurate and uniform evaluation for the enormous number of mammograms generated in widespread screening [62]. Mammography is a specific type of imaging that uses a low-dose X-ray system to examine the breast [5]. A labelled mammogram image is shown in Figure 2.12. During the test, each breast is positioned carefully on a special film cassette and then gently compressed with a paddle. This compression flattens the breast so that the maximum amount of tissue can be imaged [5]. Special energy and wavelength of the x-rays allow them to pass through the compressed breast

and create the image of the internal structures of the breast onto a film cassette positioned under the breast [62]. Differences in absorption and the corresponding varying exposure level of the film create the images, which can clearly show normal structures such as fat, fibro-glandular tissue, breast ducts, and nipples [62]. Furthermore, abnormalities such as masses and calcifications are also visible. The abnormalities appear as levels of white on a mammogram, however, fat and everything else such as glands and connective tissue appears as black regions [62].

There are two types of mammography, film mammography and digital mammography. In film mammography, the image is created directly on film, whereas digital mammography takes an electronic image of the breast and stores it directly on a computer [57]. Nowadays, in most Hospitals, film mammography is replaced with digital mammography since it has many potential advantages over film mammography [57]. In digital mammography, each breast is imaged separately in two views [6]: The craniocaudal (CC) view and the mediolateral-oblique (MLO) view as shown in Figure 2.13.

The main advantage of using mammogram screening that mammograms use low doses of radiation (x-rays) to produce a 1-D image of the breast and the calcium deposits are accurately visualised [63]. On the other hand, mammograms have some limitations such as it uses ionizing radiation, patients can feel discomfort because of breast compression, the breast must be repositioned for different views, and imaging dense breast tissue is difficult [63].

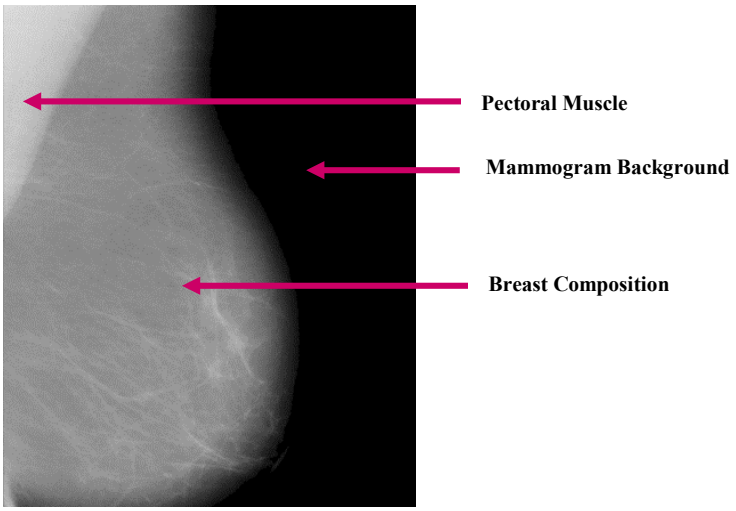
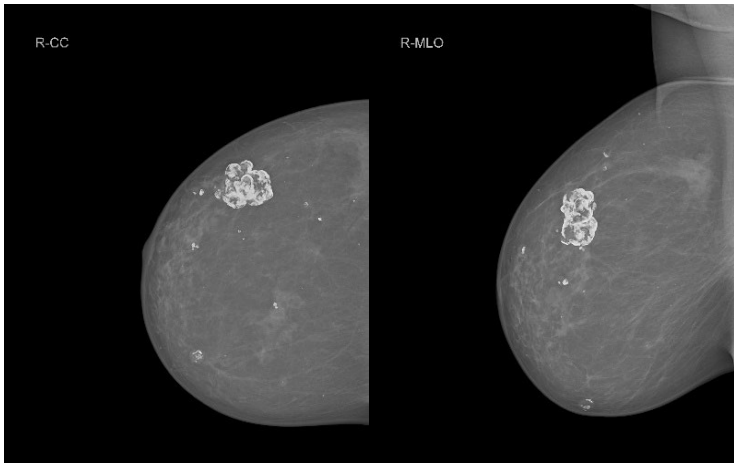


Figure 2.12: A labelled mammogram image.



(a) (b)

Figure 2.13: Two views of a mammogram image (a) CC view and (b) MLO view for the right breast extracted from the DAR-Breast dataset.

2.5.4 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) as shown in Figure 2.14 is the most attractive alternative to a mammogram [64]. MRI is sensitive for detecting some cancers, which could be missed by mammography [64]. In addition, MRI can help radiologists to determine how to treat breast cancer by identifying the stage of the disease. It is highly effective to image breasts after breast surgery or radiation therapy. To be effective; contrast-enhanced breast MRI is carried out by injecting the patient's body with a paramagnetic contrast agent [64]. This method is based on the hypothesis that after the injection of the agent, abnormalities enhance more than normal tissues due to their increased vascularity, vascular permeability, and interstitial spaces [64]. MRI forms a 3-D uncompressed image. Moreover, it can perform with all women, including those who are not suitable for mammography, such as young women with dense breasts and women with silicone-filled breast implants [64]. Since it uses magnetic fields, MRI has no harmful effects on human bodies; however, MRI takes a long time to perform and has a high cost, which is more than ten times greater than mammography [64]. The advantages of the MRI are; it is more sensitive than mammography and it uses magnets that emit radio waves to produce a 3-D view of the breast and the underlying structures and vessels [65]. Moreover, it has no ionizing radiation, thus no harmful effects. The images generated by MRI are remarkably complete, detailed, and precise than other cardiac imaging tests [65]. Furthermore, MRI may be able to visualise dense breast tissues compared with mammography [65]. On the other hand, the disadvantages of MRI scans are; it cannot visualise calcifications that typically surround DCIS lesions, and a patient could develop an allergic reaction to the contrasting agent, or that a skin infection could develop at the site of injection [65].

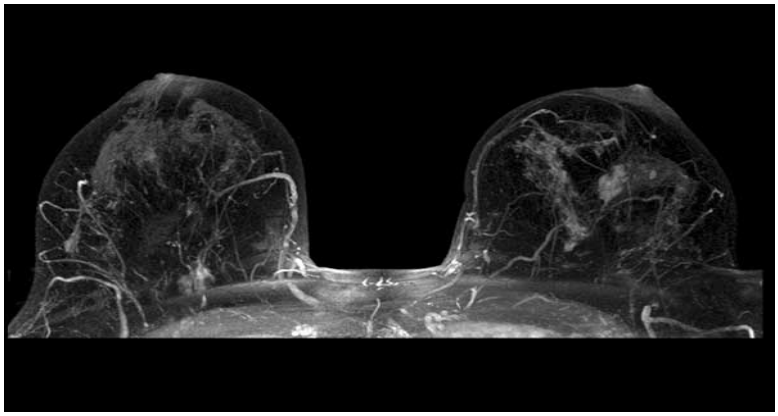


Figure 2.14: An example of a breast MRI scan. Adopted from [52].

2.5.5 Thermography

Thermography is a biological or functional exam, which is considered a high technology tool that specifically measures inflammation in the body [66]. This test is particularly good for assessing active areas of cancer cell formation. It is more effective and is significantly less invasive than mammography [66]. Thermography is a scanning device that measures the body's surface temperature and presents the information as a digitized image. This tool makes a digital map of the body and illustrates heat patterns very accurately [66]. These patterns may detect some abnormal condition such as cancer cell growth or active infection. Thermal images are usually captured from shoulder to waist [67], [68]. Figure 2.15 shows a sample of a breast thermogram scan. Thermography does not use ionizing radiation, venipuncture, or other invasive procedures. Besides that, it is a quick, painless, economic, risk-free, and patient-friendly imaging method [68]. Therefore, it is suitable for all breast sizes and densities, pregnant or nursing women, implanted breast, and even post-operative patients. Thermography detects physiological changes by measuring the infrared radiation emitted by the body [68].

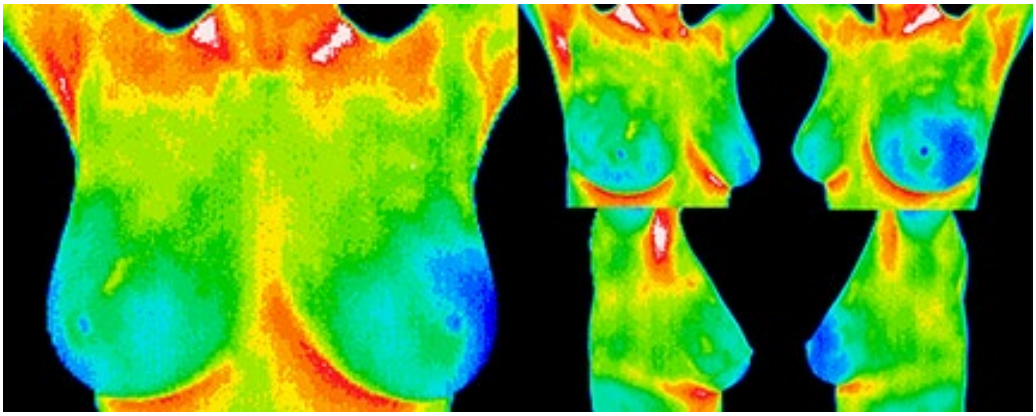


Figure 2.15: An example of a thermogram scan. Adopted from [69].

2.5.6 Positron Emission Tomography

Positron Emission Tomography (PET) is nuclear medicine and a functional imaging technique that is used to observe metabolic processes in the body. It scans all the body from the forehead to the middle of the thighs [70]. The system detects pairs of gamma rays emitted indirectly by a positron-emitting radionuclide (tracer), which is introduced into the body on a biologically active molecule [70]. Three-dimensional images of tracer concentration within the body are then constructed by computer analysis. In modern PET-CT scanners, three-dimensional imaging are often accomplished with the aid of a CT X-ray scan performed on the patient during the same session, in the same machine [70]. PET scanning is a molecular imaging procedure that allows physicians to obtain three-dimensional images of what is happening in a patient's body at the molecular and cellular level. For a PET scan, a patient is injected with a very small amount of a radiotracer such as fluorodeoxyglucose, which contains both a sugar and a radioactive element [70]. The radiotracer travels through the body and is absorbed by tumours or cancer cells. The patient then lies down on an examining table and is moved to the centre of a PET/CT scanner [70]. The PET/CT scanner contains a PET scanner and a CT scanner next to each other. The CT scan and the PET scan are obtained one after the other. The PET scanner is composed of an array of detectors that receive signals emitted by the radiotracer [70]. Using these signals, the PET scanner detects the amount of metabolic activity while a computer reassembles the signals into images [71]. Figure 2.16 shows a breast PET scan image.

The pros of PET scan are the ability to study body function through biochemical functions that can detect the disease, making it more effective in diagnosing than other imaging tests [72]. By studying the metabolic functions in patients, PET imaging can be used as an alternative to biopsy and other exploratory surgeries to determine how much the disease has spread [72]. It can distinguish between benign and malignant tumours, which gives it a benefit to be a more accurate medical tool that can reduce the number of unnecessary surgeries performed due to incorrect diagnosis and staging data [72]. Conversely, PET scan risks are caused by the radioactive component used during this procedure. Although the radioactive compounds used in PET imaging are short-lived, it may not be suitable for patients who are pregnant, suspect that they are pregnant, or are breastfeeding [72]. PET scans require cyclotrons, an expensive machine that creates the radioisotopes that are used in the radioactive tracers required for PET imaging. PET scans are not offered in the majority of medical centres in the world, consequently, it is a difficult treatment to receive [72], [73]. Figure 2.17 and

Figure 2.18 shows the PET scan device and the radioactive isotope machine, respectively. These were obtained from the West of Scotland PET Centre at Gartnavel General Hospital during my visit to the Hospital.

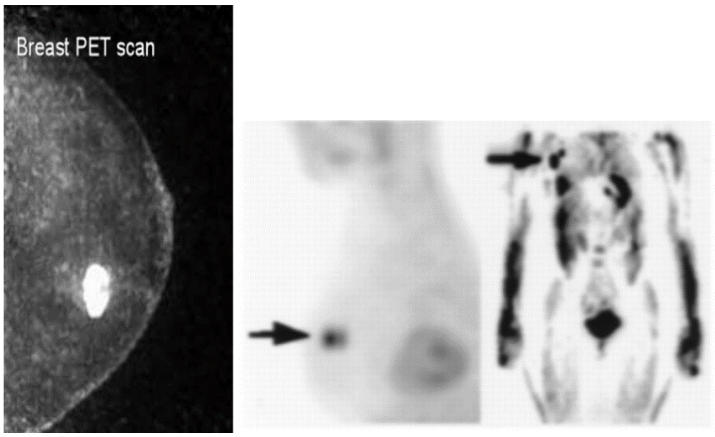


Figure 2.16: A PET scan image for breast. Adopted from [70].

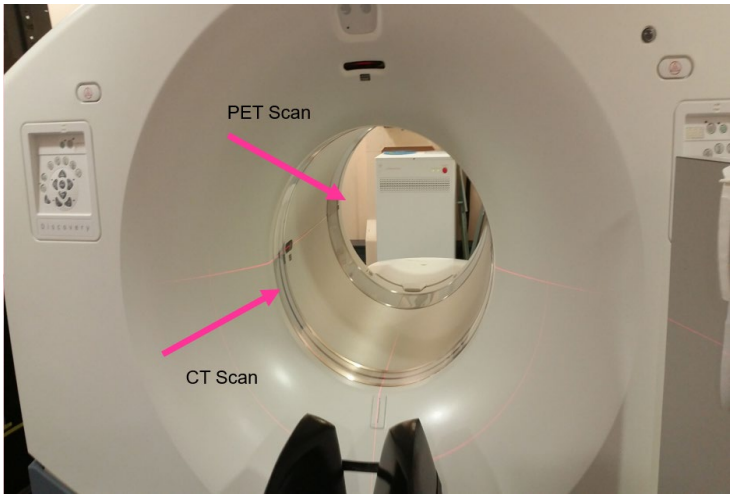


Figure 2.17: A PET scan device captured from the West of Scotland PET Centre at Gartnavel General Hospital.

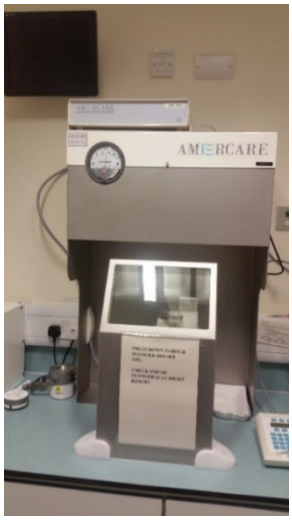


Figure 2.18: The Radioactive isotope machine captured from the West of Scotland PET Centre at Gartnavel General Hospital.

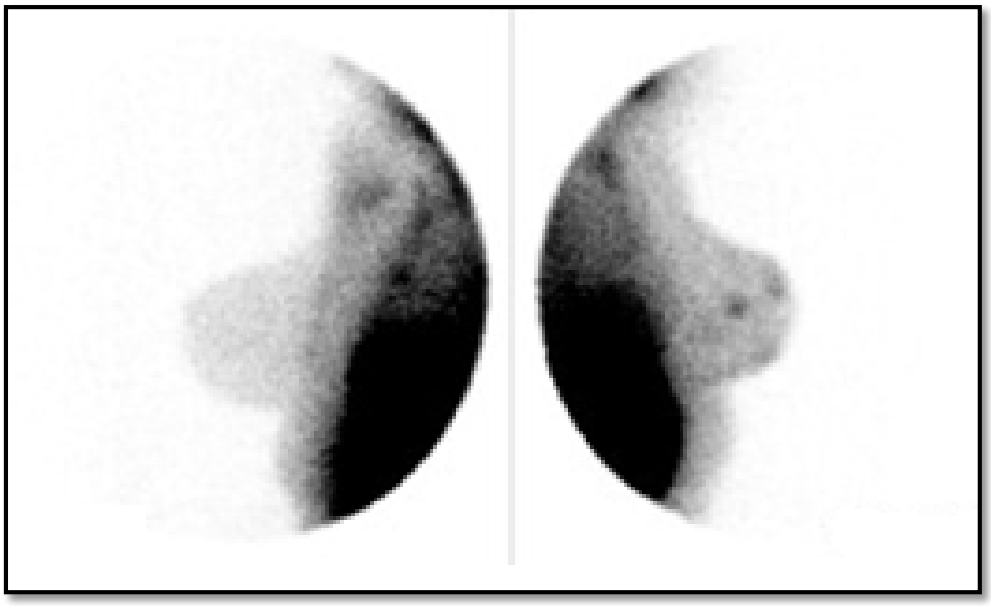


Figure 2.19: A nuclear medicine breast imaging sample. Adopted from [74].

2.5.7 Nuclear Medicine Imaging

Nuclear medicine imaging, which is called radionuclide scanning, is considered an effective diagnostic tool. This is because it shows not only the anatomy of an organ or body part, but the function of the organ as well [74]. Therefore, nuclear medicine can be valuable in the early diagnosis, treatment, and prevention of numerous medical conditions and continues to grow as a powerful medical tool [74]. On the other hand, nuclear medicine breast imaging is a supplemental breast exam that may be used in some patients to investigate a breast abnormality. A nuclear medicine test could not be considered as a primary investigative tool for breast cancer. However, it is used to investigate a breast abnormality after a mammography diagnosis [74]. Moreover, it helps radiologists to decide whether a breast abnormality requires biopsy or not. Nuclear medicine breast imaging is considered a non-invasive test. It involves injecting a radioactive tracer into the patient. Since the dye accumulates differently in cancerous and non-cancerous tissues [74], [75]. After injection, the radiotracer eventually accumulates in the breast, where it gives off energy in the form of gamma rays. A device called a gamma camera detects this energy [75]. The camera and a computer measure the amount of radiotracer absorbed by the body and produce pictures that detail organ, tissue structure, and function. Figure 2.19 shows a nuclear medicine breast-imaging sample.

Figure 2.20: The international symbol of breast cancer awareness. Adopted from [76].

2.6 Breast Cancer Awareness

2.6.1 Breast Cancer Symbol

The pink symbol of breast cancer shown in Figure 2.20, implemented in the early 1990s, represents the international symbol of breast cancer awareness [76]. In addition, it is used by various breast cancer organizations to promote breast cancer awareness and to support fundraising campaigns [76]. The first breast cancer awareness foundation was founded in 1982 in memory of Susan Komen, who died from the disease at age 36 in 1980 [76].

2.6.2 Early Detection of Breast Cancer

At present, there are no effective ways to prevent breast cancer, because its cause remains unknown. However, efficient diagnosis in its early stages can give a woman a better chance of full recovery. Therefore, early detection of breast cancer can play an important role in reducing the associated morbidity and mortality rates [57].

Manual readings of mammograms may result in misdiagnosis due to human errors caused by visual fatigue. Computer-aided diagnosis systems (CAD) serve as a second opinion for radiologists. CAD systems in the field of digital mammography are divided into two main categories; computer-aided detection systems (CADe) and computer-aided diagnosis systems (CADx/CAD) [77]. The CADe systems are capable of pinpointing suspicion regions for further analysis from an expert radiologist. However, the CAD systems are capable of making a

decision, whether the examined suspicion regions consist of abnormal or healthy tissue and distinguishing between benign and malignant lesions [78].

2.7 Summary

This chapter demonstrates a structured introduction to breast cancer. Firstly, the incidence of breast cancer worldwide, and especially in Egypt, is discussed. After a peer review it shows that breast cancer has a priority position among cancers for women. Secondly, the anatomy of the breast is illustrated, by highlighting the important parts of the breast. Then, a quick introduction about the different types and stages of breast cancers is given, as well as various types of tumours. Afterward, the mass and calcifications are presented, which are actually the most important abnormalities that occur in the breast. Due to the fact that a breast tumour examination can help to detect the tumours in the early stages, screening devices for breast cancer are introduced. To conclude, the medical imaging examination in digital mammography are found to be the most effective way of breast cancer detection. This is because breast cancer has a known asymptomatic phase that can be detected with the mammography. The CAD system played an important role in medical image classification, especially in classifying breast cancer, so the CAD system will be discussed in detail in Chapter 3.

Chapter 3

Computer-Aided Diagnosis: Datasets and State-of-the-Art

3.1 Introduction

Computer-aided diagnosis, known as CADx or CAD [79], is a fundamental tool for assisting radiologists in the image interpretation task. Since the 1980s, its popularity has grown, and it has become an important area of research. The first attempts to analyse radiographic abnormalities using computers took place in the mid of 1950s [79]. However, in the 1960s and 1970s, the first automatic detection and classification of lesions in breast images were investigated [79]. The creation of medical image databases with annotations, which are in some cases freely available for the scientific community, enables the training and testing processes, and the assessment of CAD scheme's performance as well [7]. A CAD system aims to determine the abnormalities in the breast with great accuracy and reliability. It plays as a second opinion to radiologists, which means that the radiologists can use the results of CAD in making a decision [80]. However, the radiologists must determine the final diagnostic decisions for patient treatment [80].

In this chapter, a discussion on the available mammogram datasets is provided, and then the steps of forming the CAD system are presented. Finally, this chapter provides the reader with a comprehensive survey of published works related to CAD systems for breast cancer classification. It is divided into two sections: CAD systems using ML and DL techniques.

3.2 Mammogram Breast Cancer Datasets

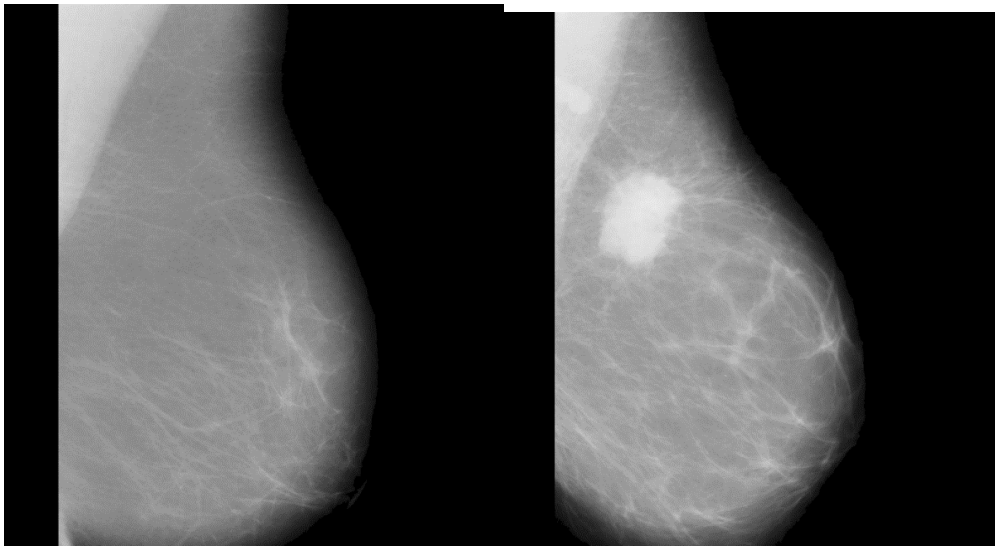
The relevant datasets that are used in most of the state-of-the-art studies are seven datasets, which are the Mammographic Image Analysis Society Digital Mammogram dataset (MIAS) [81], the Digital Database for Screening Mammography (DDSM) [82], the OPTIMAM database (OMI-DB) [83], the INbreast dataset [84], the Breast Cancer Digital Repository Film Mammography dataset (BCDR) [85], the Curated Breast Imaging subset of DDSM (CBIS-DDSM) [86], and the Digital Mammography Dream Challenge dataset [87]. These datasets will be discussed in the following sub-sections.

3.2.1 The MIAS Dataset

The MIAS dataset was created by an UK research organisation, namely the Mammographic Image Analysis Society (MIAS) [81]. The films have been digitised to a 50-micron pixel edge. All images are freely available online with a size of 1024 × 1024 [88]. Mammogram images are available through the pilot European image processing archive (PEIPA) at the University of Essex. The MIAS dataset contains 322 marked breast images in the portable grey map (PGM) format of 161 women in MLO view only [81]. A ground truth (GT) was prepared by experienced radiologists and confirmed using a biopsy test. The GT of the MIAS mammograms shows the location of the abnormality, the radius of the circle which contains the abnormal region, the characteristics of the background tissues, the breast density of each image such as fatty, glandular, or dense, and the severity of each abnormality either normal, benign or malignant [81]. A sample extracted from the MIAS dataset is shown in Figure 3.1.

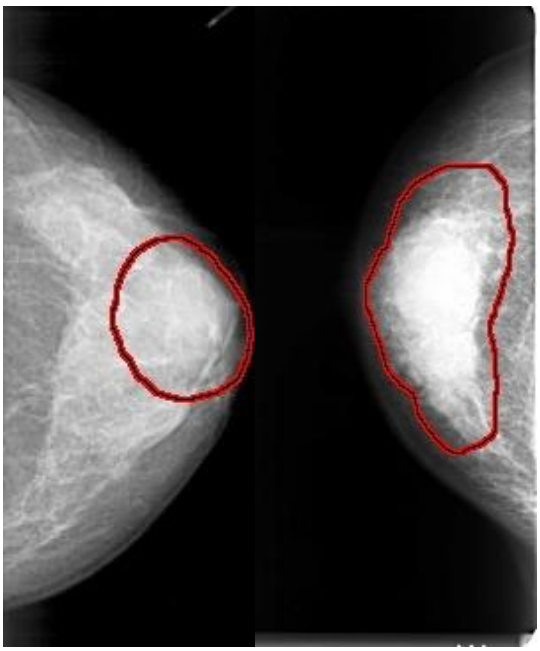
3.2.2 The DDSM Dataset

The Digital Database for Screening Mammography (DDSM) dataset is a well-known dataset released in the year 2001 [82], which is considered as one of the few publicly available online dataset [89] that is used for breast cancer detection and classification [82]. The DDSM is a collaborative effort involving the Massachusetts General Hospital, the University of South Florida, and the Sandia National Laboratories. Additionally, this dataset is collected from many sources: Wake Forest University's School of Medicine, Sacred Heart Hospital, Massachusetts General Hospital, and Washington University of St Louis (School of Medicine) [82]. It consists of 2620 scanned film mammography cases available in 43 volumes representing real breast data. The resolution of the mammogram is 50μm/pixel and the grey level depths are 12 bits and 16 bits. The volumes could be normal, benign, or malignant samples [82]. Each case involves four breast images, two of them are MLO views and the others are CC views of each breast. The images are in Joint Photographic Experts Group (JPEG) format. Benign and malignant masses in all mammograms are labelled and annotated by expert radiologists [82]. A benign and malignant mass tumour extracted from the DDSM dataset are shown in Figure 3.2.



(a) (b)

Figure 3.1: Samples extracted from the MIAS dataset (a) normal and (b) spiculated mass samples [81].



(a) (b)

Figure 3.2: A benign (left) and a malignant (right) mass tumour extracted from the DDSM dataset [82].

3.2.3 The OPTIMAM Database (OMI-DB)

The OMI-DB [83] is an extensive mammography image database of over 145,000 cases in over 2.4 million images, though not freely available. It is comprised of unprocessed and processed full-field digital mammogram (FFDM) from the National Breast Screening System (NBSS) in the UK, which also contains manually GT determined by domain experts and associated clinical data linked to the images [83]. The OMI-DB is composed of three sets of images, i.e., normal, benign, and malignant images, which have been selectively collected from three sites during the years of 2010 – 2019 [83].

3.2.4 The INbreast Dataset

The INbreast dataset was obtained from the S. João Hospital Centre in Porto [84] under the permission of both the Hospital's Ethics Committee and the National Committee of Data Protection [84]. The images were acquired between April 2008 and July 2010, and the acquisition equipment was the Mammo Novation Siemens FFDM, with a solid-state detector of amorphous selenium, a pixel size of 70 mm (microns), and 14-bit contrast resolution [84]. The INbreast dataset consists of 115 cases, of which 90 have two images (MLO and CC) of each breast and the remaining 25 cases are from 141 women who had a mastectomy, and two views of only one breast were included [84]. Therefore, in total there are 410 images available in the digital imaging and communication in medicine (DICOM) format. The sizes of the images are 3328 × 4084 or 2560 × 3328 pixels depending on the compression plates used in the acquisition process. All lesions including masses were assigned to a standardised Breast Imaging-Reporting and Data System (BI-RADS) category [90] by a radiologist after interpreting the corresponding mammogram. While the INbreast dataset should be obtainable via a request from, over the duration of this research projecting access has proven unsuccessful over a prolonged period. This dataset therefore appears to be hard to obtain at best [91].

3.2.5 The BCDR Dataset

The Breast Cancer Digital Repository Film mammography dataset # 3 (BCDR-F03) [85] dataset is extracted from the Breast Cancer Digital Repository, which is considered as a new dataset of film mammography. It is composed of 736 biopsy-proven lesions of 344 patients collected between March 2009 and March 2013. Each case includes clinical data for each patient, and both CC and MLO view mammograms, are available with the coordinates of the lesion's contours. BCDR-F03 is a binary class dataset composed of benign and malignant cases [85].

3.2.6 The CBIS-DDSM Dataset

The Curated Breast Imaging subset of the DDSM (CBIS-DDSM) is an updated and standardised version of the DDSM dataset released in the year 2017 by Lee et al. [86]. It is a free online dataset available in [92]. The CBIS-DDSM contains 6775 studies with 10,239 images in the standard DICOM format, which includes the original and segmented ROI images. The dataset contains 753 microcalcification and 891 mass cases in 1546 and 1318 images, respectively, for both MLO and CC views of the screened breasts [86]. Figure 3.3 shows samples of a benign and malignant mass.

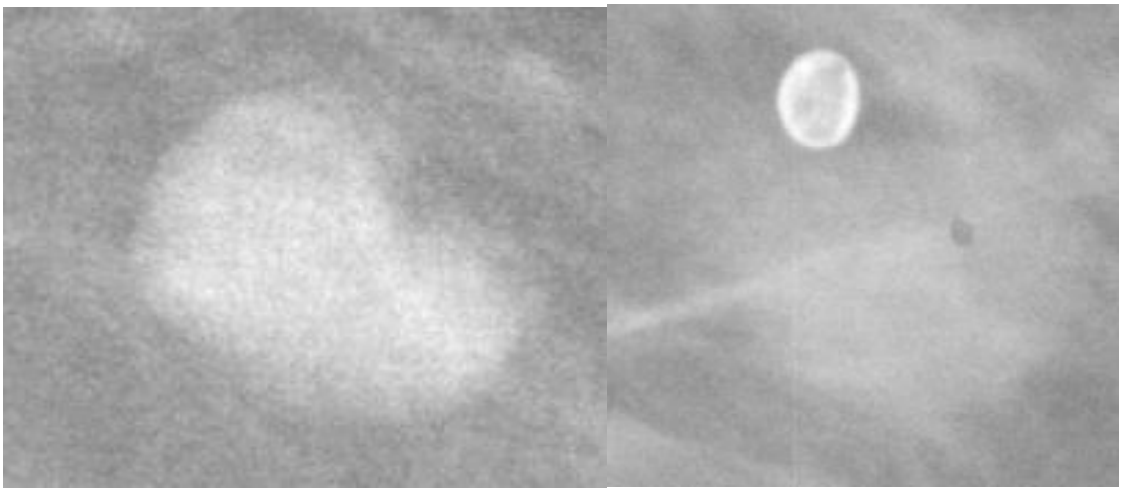


Figure 3.3: A benign (left) and a malignant (right) mass tumour extracted from the CBIS-DDSM dataset [86].

3.2.7 The Digital Mammography Dream Challenge Dataset

The Digital Mammography Dream Challenge [87] is a new dataset. The crowdsourcing coding competition offers a large monetary prize for the best algorithm in accurately predicting breast cancer on screening mammography [87]. This dataset consists of 34 and 466 abnormal and normal samples, respectively [93].

3.2.8 Comparative Investigation of the existing Datasets

In this work, the MIAS, DDSM, and CBIS-DDSM datasets are analysed due to their wide use. Furthermore, these datasets are used in order to compare our new proposed method with related studies on the same datasets. Since the OMI-DB, INbreast, and BCDR-F03 datasets are hard to access as they are subject to potentially protracted approval processes, they have been omitted from the comparison for fairness of access. On the other hand, the Digital Mammography Dream Challenge dataset is not normalised hence, it was not used in this work. The summary of the mammography datasets used in this thesis is shown in Table 3.1.

Table 3.1: The summary of the mammography datasets used in this thesis.

Dataset	Number of cases	Number of images	Available classes	Image format
MIAS	-	322	Normal, Benign, and Malignant	PGM
DDSM	2620	10480		JPEG
CBIS-DDSM	1644	3586		DICOM

3.3 Computer-Aided Diagnosis System

In the last two decades, the enormous growth in computer-aided diagnosis (CAD) research has occurred to support radiologists for the early detection of masses in mammograms [79]. However, these systems have limited effectiveness in terms of accuracy and are disposed to reduce the number of False Positives (FP) and False Negatives (FN) [94], [95]. CAD systems have been an essential part of screening mammograms, despite their effectiveness in the current form [96].

A breast cancer CAD system usually consists of five stages; (1) image pre-processing, (2) image segmentation, (3) feature extraction, (4) features classifications, and (5) system evaluation. The block diagram of a classical CAD system is shown in Figure 3.4 and it will be discussed in the following sub-sections.

3.3.1 Image Pre-processing

Image pre-processing is an important step in the process; it is a common name for operations with images at the lowest level of image processing [97]. The aim of pre-processing is an improvement of the image data that suppresses or enhances some image features important for further processing, although geometric transformations of images such as rotation, scaling, translation is classified among pre-processing methods [97]. The classifiers perform well when they can be training on a large number of samples. Since the amount of biomedical data available for this project is insufficient, data augmentation is applied in its different forms in order to increase this amount of data [98]. In the following sub-sections, the image enhancement and data augmentation techniques will be presented as they are considered in this work.

3.3.1.1 Image Enhancement

The image enhancement technique is used to suppress noise and improve the contrast of the image as it enlarges the intensity difference between objects and backgrounds [99]. The goal of the image enhancement techniques is to accentuate certain image features for the subsequent analysis or display [99]. Since the mammographic image has low contrast and it is hard to read the abnormalities, therefore, enhancement is applied to the mammographic image [100], [101]. There are several types of image enhancement techniques, among them the global histogram modification, local processing, and multi-scale processing approaches [99].

Figure 3.4: The main components of a CAD system.

The global histogram modification approach is to reassign the intensity values of pixels to make the new distribution of the intensity uniform. It is effective in enhancing the entire image with low contrast, but it cannot enhance the textural information and it only works with images having one object [102]–[104]. The most common global histogram technique is the adaptive histogram equalization (AHE). It is capable of improving an image's local contrast and bringing out more details in the image. However, it can also produce significant noise [105]. AHE is an excellent contrast enhancement method for both natural and medical images [106]. A generalization of AHE called contrast limited adaptive histogram equalization (CLAHE) was developed, where the histogram is calculated for the contextual region of the pixel [107]. One of the disadvantages of AHE is that it may over enhance the noises in the images due to the integration operation. Therefore, CLAHE is used as it uses a clip level to limit the local histogram to restrict the amount of contrast enhancement for each pixel [108]. The CLAHE algorithm can be summarised as follows: (1) divide the original image into contextual regions of equal size, (2) apply the histogram equalization on each region, (3) limit this histogram by the clip level, (4) redistribute the clipped amount among the histogram, and (5) obtain the enhanced pixel value by the histogram integration [108].

The local processing approach is based on non-linear mapping methods, such as adaptive histogram techniques [107]. Its implementation can be feature-based, which can be gained by edge detection or by using local statistic information such as local mean, standard deviation, etc. The local approach is effective in the local texture enhancement, but it cannot enhance the entire image well. On the other hand, the multi-scale processing approach is based on wavelet transformation [109]; it is flexible to select local features to be enhanced, and it can suppress the noise. However, it is difficult to determine the mother wavelet for transformation and weight modification functions [110].

3.3.1.2 Data Augmentation

Training the model with a quite small dataset leads to overfitting during the learning process. This means that the model remembers the details of the training set and it does not generalize it based on the validation or testing sets [111]. Therefore, to construct an efficient

classification model, a large dataset should be used [112]. However, the most frequently mentioned problem in the field of ML/DL is the lack of a sufficient amount of the training data or uneven class balance within the datasets [113]. This is because the biomedical datasets contain a slight number of samples due to limited patient volume. One of the ways of dealing with this problem is data augmentation [98]. It is a widely used technique in many ML tasks to virtually enlarge the training dataset size and avoid overfitting [114]. Data augmentation aims to increase the number of images; this is done by generating new images from original ones [115], [116]. Data augmentation has many forms, such as rotation, flipping, and transformation [117], [118].

3.3.2 Image Segmentation

Image segmentation is used to divide an image into parts having similar features and properties. The main aim of segmentation is to extract the regions of interest (ROIs) containing all the abnormalities in the breast and locate the suspicious regions from the ROI. This is achieved by simplifying the image by presenting it in an easily analysable way [6], [101]. There are several segmentation techniques, such as thresholding technique, region-based technique, and edge detection technique [119], [120].

3.3.2.1 Thresholding Techniques

Thresholding methods are the simplest methods for image segmentation. The image pixels are divided with respect to their intensity level. There are two threshold techniques: Global and local thresholding [119]. Global thresholding is one of the common techniques for image segmentation. It is based on global information such as the histogram of a mammogram, and it is not good for identifying ROIs because masses are often superimposed on the tissue of the same intensity level [6]. Therefore, global thresholding has good results when used as a primary step of some other segmentation techniques [121]. The global thresholding is done by setting an appropriate threshold value (T). This value of (T) will be constant for the whole image. Based on (T), the output image p (x,y) can be obtained from the original image q (x,y) as given in equation (3.1). However, local thresholding is slightly better than global thresholding; the threshold value is defined locally for each pixel based on the intensity values of its neighbour pixels [119]. Multiple pixels belonging to the same class are not always homogenous and may be represented by different feature values [121].

$$p(x,y) = \begin{cases} 1, if\ q(x,y) > T \\ 0, if\ q(x,y) \leq T \end{cases} \quad (3.1)$$

3.3.2.2 Region-Based Techniques

The region-based segmentation divides the image into different regions based on predefined criteria [122]. There are two main types for region-based segmentation, (1) region growing and (2) region splitting and merging. The region growing can remove a region from an image based on some predefined criteria such as the intensity. Region growing is an approach to image segmentation in which neighbouring pixels are examined and joined to a region class where no edges are detected [123]. The region growing is also classified as a pixel-based image segmentation method as it involves the selection of initial seed points. A seed pixel is chosen as a starting point from which the region iteratively grows and aggregates with neighbouring pixels that fulfil a certain homogeneity criterion [123]. On the other hand, the region splitting and merging method is the opposite of the region growing method as it works on the complete image [124].

3.3.2.3 Edge Detection Techniques

Edge detection methods transform original images into edge images benefits from the changes of grey tones in the image [125]. This is performed by dividing an image into an object and its background by observing the change in intensity or pixels of an image [122]. Edges are detected to identify the discontinuities in the image. Edges are local changes in the image intensity. Edges typically occur on the boundary between two regions [122]. The edges on the region are traced by identifying the pixel value and it is compared with the neighbouring pixels [126]. The major property of the edge detection technique is its ability to extract the exact edge line with good orientation. The main features can be extracted from the edges of an image and used by advanced computer vision algorithms [125]. Edge detection is used for object detection, which serves various applications like medical image processing [125]. There are many edge detection techniques, such as Sobel, Canny, and Robert's operators, the result of these methods is a binary image [125].

3.3.3 Feature Extraction

Feature extraction is a reliable tool to reduce the amount of irrelevant information, producing a robust feature vector [127]. The features that characterize a specific region are calculated from the ROI characteristics such as size, shape, density, etc. There are many techniques for feature extraction. Although, three main types of features could be extracted; intensity (colour), geometry (shape), and texture features [127]. The intensity or the colour feature is considered the simplest feature compared to other features. Images characterized by colour features have many advantages such as robustness, effectiveness, and simplicity. Colour

features can be derived from a histogram of an image. The weakness of colour histogram is that the colour histogram of two different things with the same colour can be the same. However, the geometric or shape feature is an important visual feature and one of the primitive feature for image content description. The shape of an object refers to its physical structure and profile. Shape content description cannot be defined exactly because it is difficult to measure the similarity between the shapes. The geometric feature extraction technique can include the moment, perimeter, and area of an image [128]. On the other hand, the texture features are based on the texture of the image. There are many techniques used by several researchers for texture feature extraction, such as wavelets, statistical features, and fractals. Texture analysis methods have been utilized in a variety of application domains such as medical image processing, document processing, remote sensing, and content-based image retrieval [129].

In recent years, deep learning (DL) techniques and especially the deep convolutional neural networks (DCNN) have attracted great attention due to their outstanding performance as a feature extractor. Consequently, the DCNN is used in this research to extract the features from the breast mammography images.

3.3.3.1 Deep Convolutional Neural Network

Deep learning (DL) can be defined as a set of methods that learn data representations using multiple levels of representation [130]. This can be obtained by composing simple, but non-linear models that transform the representation from one level starting with the raw input into increasing levels of representation [130]. The idea of representation learning is not new, but recently it emerged as a viable alternative due to the appearance and the popularization of the graphic processing units (GPUs), which are capable of delivering high computational at relatively low cost, achieved through their massively parallel architecture [20], [130], [131].

In recent years, DL based on a convolutional neural network (CNN) or deep convolutional neural network (DCNN) has attracted great attention in AI due to its successes in various research fields, such as pattern recognition, computer vision, natural language processing, and big-data analysis [132]–[134]. The convolutional process can simplify an image containing millions of pixels to a set of small feature maps, thereby reducing the dimension of input data while retaining the most-important difference features [23]. Moreover, DCNN achieved outstanding performance in recognition of natural images [23], [135], [136]. Additionally, DL techniques are one of the most often used algorithms for getting better, scalable, and accurate results from the data as compared to state-of-the-art methods of ML. DCNN is applied to the biomedical images to diagnose diseases with precisely tailored

treatment plans for improving the patient's health. DCNN is widely used in breast cancer classification problem due to its ability to extract useful features from images. Additionally, mammograms, MRI, etc. are the trending biomedical images for the diagnosis of patients by minimising the intervention of humans [22], [137]–[145] and [18], [19], [146], [147].

A DCNN consists of multiple trainable stages stacked on top of each other, followed by a supervised classifier and sets of arrays named feature maps to represent both the input and output of each stage [148]. There are three main types of layers used to build DCNN architectures, convolutional layer, pooling layer, and fully connected (FC) layer [149].

- **Convolution Layer**

The convolutional layer computes the output neurons connected to the local regions in the input. Each neuron computes a dot product between their weights and the region connected to the input volume. The set of weights that is convolved with the input is called filter or kernel. Every filter is small spatially (width and height), but extends through the full depth of the input volume. The distance between the applications of filters is called stride [149]. The output of a convolution layer is calculated as,

$$\text{The output size of the conv layer} = \left[\left(\frac{input - filter\ size + (2\ \times Padding)}{Stride}\right) + 1\right]. \quad (3.2)$$

- **Pooling Layer**

The pooling layer down samples the volume spatially, independently in each depth slice of the input volume to reduce the amount of computation and improve the robustness [150]. Thus, the pool operator resizes the input along width and height, discarding activations. The maximum-pooling function is applying a window function to the input patch and computes the maximum in that neighbourhood [149]. The output of the pooling layer is calculated using equation (3.3).

$$\text{The output size of the pooling layer} \ = \left[\left(\frac{output\ of\ conv - pool\ size}{Stride}\right) + 1\right] \quad (3.3)$$

- **Fully Connected Layer**

In a fully connected layer (FC), neurons have full connections to all activations in the previous layer and their activations can be computed using a matrix followed by a bias offset. This type of layer is standard with a regular neural network. The last fully-connected layer holds the net output, such as probability distributions over classes [149].

Additionally, there are other layers such as normalisation layer [151] and a rectified linear unit (ReLU) or activation function [152]. The normalisation layer normalises the output

of the previous activation layer by subtracting the batch mean and dividing by batch standard deviation. However, the ReLU is a piecewise linear function that outputs the input directly if it is positive otherwise outputs zero [153]. The input of the DCNN is the intensity volume (R, G, B) of an image. While the output produces a distribution of predicted probabilities over the 1000 classes for ImageNet classification [154].

There are many DCNN architectures between them CiFarNet [155], [156], AlexNet [23], VGGs nets [157], GoogleNet [24], Inception-v3 [25], the ResNet [26], Inception ResNet-v2 [27], and MobileNet-v2 [28]. Table 3.2 and Figure 3.5 show the summary and the timeline of the used DCNN architectures, respectively. Moreover, the full details of the networks used in this work are presented in Appendix A. All DCNNs are firstly trained using the ImageNet dataset [154], which contains 1.2 million natural images for the classification of 1000 classes.

3.3.3.1.1 Transfer Learning

Generally, the training process for DCNNs requires a large number of annotated samples to avoid overfitting to the training dataset. Researchers often address this issue as transfer learning, which is also known as domain adaptation. Transfer learning is considered to be an efficient methodology, in which the knowledge from one image domain can be transferred to another image domain [29], [31], [158]. Transfer learning is an efficient approach for dealing with small datasets by allowing pre-trained networks to be fine-tuned and adjusted to solve problems from a particular domain or imaging modality. The weights of the model are pre-initialized when utilizing a pre-trained version, as opposed to being randomly initialized while training from scratch [159]. Moreover, the transfer learning technique is performed on DCNN networks so that it could be used in any classification problem [160]. The DCNN is firstly pre-trained using the ImageNet dataset [154] with 1000 classes. Therefore, the last fully connected layer is replaced by a new layer for the classification of two classes either benign and malignant masses or normal and abnormal masses.

Some research papers discuss the importance of the transfer learning technique as in [161] the authors suggested that the success of any transfer learning approach highly depends on the extent of similarity between the datasets on which a DCNN is pre-trained and the database to which the image features are transferred. Suzuki et al. [150] introduced transfer learning in the DCNN achieving a sensitivity of 89.9% when differentiating between mass and normal lesions using the DDSM dataset [82]. Their study was the first demonstration of the DCNN mammography CAD applications. Huynah et al. [159] used transfer learning to extract and classify the DCNN features of 219 mammogram lesions from the University of Chicago

Medical Centre reaching an AUC equals to 0.81 (81%). Moreover, Tajbakash et al. [162] debated if the use of pre-trained DCNNs with sufficient fine-tuning could eliminate the need for training a DCNN from scratch. The authors analysed the influence of the choice of the training samples on the performance of DCNNs and concluded that there is no set rule to say if a shallow tuning or deep tuning is beneficial and that the optimal method is dependent on the type of the application [162]. Jiang et al. [163] explored the transfer learning to fine-tune the GoogleNet and AlexNet DCNNs architectures, to classify breast lesions. The AUC of the GoogleNet and AlexNet was 0.88 (88%) and 0.83 (83%), respectively. The authors were the first to introduce the new dataset BCDR-F03 to evaluate the results [85]. Samala et al. [164] introduced a multi-stage transfer learning for digital breast tomosynthesis (DBT) using deep neural networks. The knowledge learned from ImageNet was fine-tuned first with the mammography data and then with the DBT data. The AUC reached 0.91 (91%) for the DBT collected from the University of Michigan and the Massachusetts General Hospital. Chougrad et al. [165] used the transfer learning technique to fine-tuned the VGG-16 DCNN architecture to classify the lesions of the CBIS-DDSM [86], MIAS [81], INbreast [84], and BCDR-F03 datasets [85].

Table 3.2: A summary of the state-of-the-art DCNN architectures used in this thesis.

DCNN Architectures	Number of Layers	Input Size	Output Size
AlexNet	8	227 × 227	4096 × 2
MobilNet-v2	16	224 × 224	1280 × 2
ResNet-18	18	224 × 224	512 × 2
GoogleNet	22	224 × 224	1024 × 2
Inception-v3	48	229 × 229	2048 × 2
ResNet-50	50	224 × 224	2048 × 2
ResNet-101	101	224 × 224	2048 × 2
Inception ResNet-v2	164	299 × 299	1536 × 2

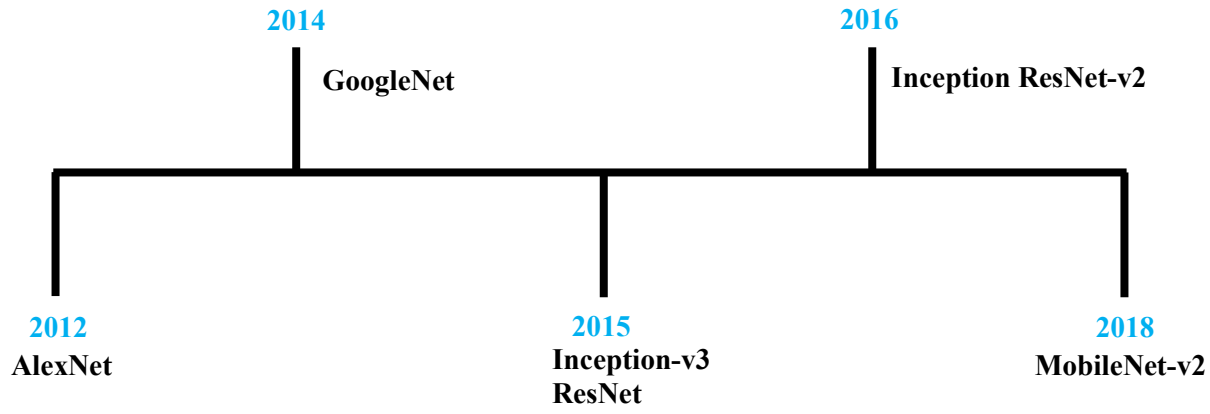


Figure 3.5: The timeline of the used state-of-the-art DCNN architectures.

3.3.4 Features Classifications

In this step, the features extracted are input into a classifier to classify the suspicious areas into normal tissues, benign, or malignant areas. There are many known classifiers among them k-nearest neighbour (k-NN) [166], [167], linear discriminant analysis (LDA) [168], [169], decision trees (DT) [170], [171], artificial neural network (ANN) [172]–[174], and support vector machine (SVM) [175]. However, the DT and SVM classifiers are used in this work and are explained briefly in the following sub-sections.

3.3.4.1 Decision Trees

Decision trees (DT) are commonly used in ML techniques. They are used extensively in medical applications such as breast cancer, ovarian cancer, and heart sound diagnosis [170]. This is due to their ability to visualise the reactions between data attributes. Visualization facilitates the doctors' understanding of how the classification decision is made [170]. DT can handle categorical and numeric attributes. They are also robust to outliers and missing values. A DT classifies data points in the training set based on rules or conditions to form a tree structure. A DT construction is like a tree with a root node whose leaves representing class labels, and branch nodes, which represent attributes and reasons, which lead to those class labels [170]. Nodes are connected by arcs, which represent the conditions on the attributes. The attribute splitting is determined by a metric such as information gain, gain ratio, or Gini index [171]. DT has several types of trees such as J48, random forest (RF), and random tree (RT).

J48 DT classifier uses top-down and greedy search through all probable nodes to construct a DT [171]. Furthermore, the RF is considered a strong classifier that achieves high classification accuracy with datasets with a huge number of features even without any feature selection [171]. Moreover, RF is capable of figuring out the important attributes of a dataset. Additionally, RT selects a random number of attributes to construct a DT and classify the data [170], [171].

3.3.4.2 Support Vector Machines

A support vector machine (SVM) is a tool that originated in modern statistical learning theory [175]. In recent years, SVM has found a wide range of real-world applications, including handwritten digit recognition, object recognition, speaker identification, face detection in images, and text categorization [12]. The SVM proved to be an efficient method for many real-world problems because of its high generalization performance without the need to add a priori knowledge. The SVM can map the input vectors into a high dimensional feature space through some non-linear mapping. In this space, an optimal separating hyper-plane is constructed [176].

The formulation of SVM is based on the principle of structural risk minimization. Instead of minimizing an objective function based on the training samples such as mean square error (MSE), the SVM attempts to minimize a bound on the generalization error [177]. This means that the error made by the learning machine on test data is not used during training. As a result, SVM tends to perform well when applied to data outside the training set. Indeed, it has been reported that SVM based approaches are significantly outperforming competing methods in many applications [176]. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These borderline training examples are called support vectors [177].

SVM performs classification by constructing N-dimensional hyper-planes that optimally separates the data into two categories. SVM models are closely related to neural networks [178]. In the parlance of SVM literature, a predictor variable is called an attribute, and a transformed attribute that is used to define the hyper-plane is called a feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case, a row of predictor values is called a vector [178]. Therefore, the goal of SVM is to find the optimal hyper-plane that separates clusters of a vector in such a way that cases with one category of the target variable are on one side of the plane, and cases with the other category are on the other side of the plane [178]. In other words, SVM aims to devise a computationally efficient way of learning separating hyper-planes in a high dimensional feature space. The vectors near the hyper-plane are the support vectors [178] as shown in Figure 3.6. The performance of SVM depends on kernel methods. The classification problem can be restricted to consideration of the two-class problem without loss of generality. In this problem, the goal is to separate the two classes by a function and to produce a classifier that generalizes well. Consider the example, in Figure 3.7, where many possible linear kernels can separate the data, but only one that maximizes the margin. This linear classifier is termed the optimal separating hyper-plane. Generally, the SVM is an implementation of the structural risk minimization principle, whose object is to minimize the upper bound on the generalization error [177], [179]. There are two cases for SVM: Linear SVM and non-linear SVM. The linear SVM has a linear kernel function as shown in Figure 3.7, however, the non-linear SVM kernels could be cubic, quadratic, and Gaussian as shown in Figure 3.8.

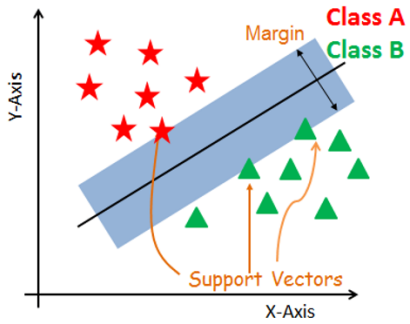


Figure 3.6: The SVM process showing the separation between two classes A and B. Adopted from [180].

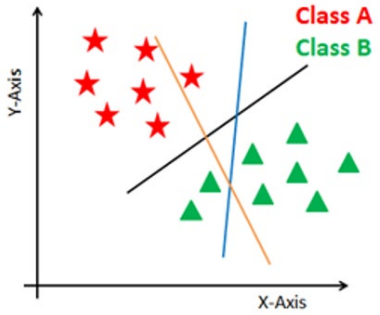


Figure 3.7: The optimal separating hyper-plane with a linear kernel function. Adopted from [180].

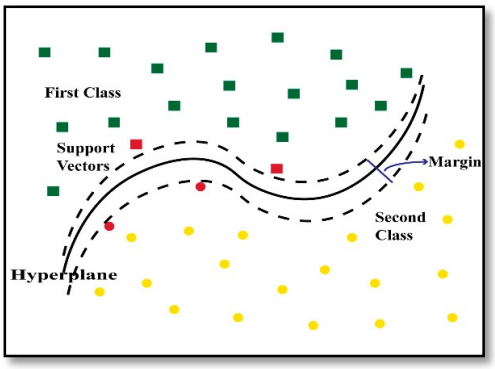


Figure 3.8: An example of a non-linear SVM. Adopted from [180].

3.3.5 Performance Metrics

The goal of classification learning is to build a classifier from a set of training samples with class labels such that the classifier can well predict the unseen testing samples [181]. The predictive ability of the classification algorithm is typically measured by its predictive accuracy. However, accuracy is not enough. Therefore, there are several evaluation tools to assess a classifier amongst them, the confusion matrix, the sensitivity, the specificity, the receiver-operating curve (ROC), the area under the ROC curve (AUC), the precision, and the F1-score [181]. Moreover, cross-validation has been widely used to evaluate ML algorithms. In addition, statistical analysis is commonly used to measure the statistical significance of the results of CAD systems. Therefore, these performance metrics are presented in the following sub-sections.

Table 3.3: An example of a confusion matrix.

Class Label	Predicted Class Label	
	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

3.3.5.1 The Confusion Matrix

The confusion matrix is a specific table visualizing the performance of the classifier [182]. Usually, in the field of ML, a confusion matrix is known as the error matrix. An image region is said to be positive or negative, depending on the data type. Furthermore, a decision for the detected result can be either correct (true) or incorrect (false). Therefore, the decision will be one of four possible categories: True positive (TP), true negative (TN), false positive (FP), and false-negative (FN) [182]. The correct decision is the diagonal of the confusion matrix. Table 3.3 provides an example of the confusion matrix for two classes' normal and abnormal classification. FN and FP are two kinds of errors [182].

An FN error implies that a true abnormality was not detected and an FP error occurs when a normal region was falsely identified as an abnormality. On the other hand, a TP decision is a correct judgment of an actual abnormality and a TN decision means a normal region was correctly labelled [6]. For years, the performance of diagnostic systems has been measured and reported in terms of correct percent, which is the percentage of diagnostic decisions that proved to be correct [5].

3.3.5.2 Accuracy

Accuracy is a measure used to determine how many instances the classifier has correctly classified from the whole data [181]. Thus, it indicates the ability of the classifier to perform well. The accuracy is defined as in equation (3.4).

$$accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{3.4}$$

3.3.5.3 The Receiver Operating Characteristics

The receiver operating characteristic (ROC) analysis is a well-known evaluation method for detection tasks [183]. It is based on statistical decision theory and it is developed in signal detection theory. ROC analysis was first used in medical decision making, subsequently, it was used in medical imaging. A ROC curve is a graph representing the true positive rate (TPR) as a function of the false positive rate (FPR) [183]. The TPR is called sensitivity or recall while the true negative rate (TNR) is called the specificity and they are defined as in equations (3.5) and (3.6), respectively.

$$sensitivity\ (TPR) = \frac{TP}{TP + FN} \tag{3.5}$$

$$specificity\ (TNR) = \frac{TN}{TN + FP} \tag{3.6}$$

Additionally, FPR and FNR indicate the false positive rate and false-negative rate, respectively, as in equations (3.7) and (3.8). FPR and FNR criteria represent the system error. FPR is a false alarm rate indicating the percentage of class (1) that has been incorrectly classified as class (2). Whereas FNR provides the percentage of class (2) that has been incorrectly classified as class (1) by the classifier [184].

$$FPR\ = 1 - TNR \tag{3.7}$$

$$FNR\ = 1 - TPR \tag{3.8}$$

3.3.5.4 The Area under the ROC Curve

The area under the ROC curve (AUC) is used in medical diagnostic systems. The AUC provides an approach for evaluating models based on an average of each point on the ROC curve. Since the AUC score is always between '0' and '1,' therefore, the model with a higher AUC value gives a better classifier performance [185].

3.3.5.5 Precision

The precision is the ratio of correctly predicted positive observations of the total predicted positive observations [186]. High precision relates to low FPR [181]. The precision is calculated using the following equation,

$$Precision\ = \frac{TP}{TP + FP} \tag{3.9}$$

3.3.5.6 F1-Score

The F1-score is the harmonic mean of precision and recall. It is used as a statistical measure to rate the performance. F1-score reaches its best score at value one. This score takes both false positives and false negatives into account [181], [186]. F1-score is defined as in equation (3.10).

$$\mathrm{F1} - score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{3.10}$$

3.3.5.7 Cross-Validation

The cross-validation is used to indicate how well the model will predict unseen data. Cross-validation is done by partitioning a dataset and using a subset to train the algorithm and the remaining data to test it [187]. The common cross-validation techniques are the k-fold cross-validation and hold out. In the k-fold cross technique, the data are randomly sorted and

divided into k folds. One of the folds is used for testing and the remaining folds for training the algorithm. This scenario is repeated k times [187]. On the other hand, the holdout technique divides the data into exactly two subsets of a specified ratio for training and testing [187].

3.3.5.8 Statistical Analysis

A result is statistically significant when a p-value is less than the significance level α. The p-value is the probability of obtaining at least as extreme results given that the null hypothesis is true [188]. The significance level α is the probability of rejecting the null hypothesis given that it is true. Statistical methods can be parametric or non-parametric. Parametric methods have several parameters such as the one-way analysis of variance (ANOVA) test. Whereas, non-parametric methods are statistics not based on parameterized families of probability distributions [188].

3.4 State-of-the-Art Review for Breast Cancer Classifications

Recently, several researchers studied and proposed methods for detecting and classifying breast abnormality in mammography images using ML and DL techniques [189]–[192]. The following sub-sections will discuss the most relevant studies for breast cancer classification using conventional ML and DL frameworks.

3.4.1 Machine Learning-based Frameworks

Machine learning (ML) methods are used extensively in medical applications this is because they have a remarkable capability for extracting and classifying the valued data from the medical datasets [193]. Researchers discussed some ML techniques to classify lesions of breast cancer giving a promising accuracy as in [194]–[203] and [204], [205], [214], [206]–[213].

Through the years 2005 till recently CAD systems using classical ML evolved; Fu et al. [215] extracted and selected using a sequential forward search (SFS) strategy 61 features for each MC sample from both the spatial and spectral domain. The selected features were classified by SVM and general regression neural network (GRNN) classifiers achieving 0.98 (98%) and 0.978 (97.8%) of AUC, respectively. The approach proposed was evaluated on the database of Nijmegen University Hospital located in the Netherlands [215]. Liu et al. [216] used the SVM classifier to classify the masses of the DDSM dataset [82] yielding to 0.7 (70%) of AUC. Sharkas et al. [217] used the discrete wavelet transform (DWT), the contourlet

transform, and the principal component analysis (PCA) methods for extracting features. The system was able to detect and classify normal and abnormal tissues in addition to benign and malignant MC tumours of the DDSM [82] dataset. They achieved an accuracy of 98%. Al Sharkawy et al. [218] extracted and classified mass lesions of the DDSM [82] dataset using DWT and SVM, respectively. The accuracy achieved was 92%. Anitha et al. [219] extracted the grey level co-occurrence matrix (GLCM) features of the MIAS [81] dataset, then classified them using SVM achieving 95% accuracy. Ragab et al. [220] used the DWT as a feature extraction technique to detect mass abnormalities of the DDSM [82] dataset. The authors employed SVM and ANN for classifying normal, abnormal tissues, benign, and malignant MCs tumours. The accuracy achieved was 96% and 98% for ANN and SVM, respectively.

Zheng et al. [221] constructed a hybrid of K-means and SVM algorithms to extract and classify the features of the Wisconsin diagnostic breast cancer (WDBC) dataset giving an accuracy of 97.38%. Beura et al. [222] used the GLCM and DWT to extract the texture features from the MIAS [81] and the DDSM [82] datasets. The authors classified the ROI using a feed-forward back-propagation multilayer neural network (BPNN) achieving an accuracy of 98% and 98.8% for MIAS and DDSM datasets, respectively. Abdel-Zaher et al. [223] developed a CAD scheme for the detection of breast cancer using a deep belief network unsupervised path followed by a backpropagation supervised path. The technique was tested on the Wisconsin breast cancer dataset (WBCD) dataset giving an accuracy of 99.68% [223]. Kim et al. [224] classified the mass samples of the DDSM [82] dataset using DT, LDA, and SVM. The highest AUC achieved was 0.95 (95%) for the SVM classifier.

Additionally, in 2016, Saad et al. [225] segmented the MC samples extracted from the DDSM [82] and MIAS [81] datasets using Otsu's algorithm. The segmented samples went through two classification techniques: ANN and adaptive boosting (AdaBoosting), which is defined as a strong classifier formed by a combination of different weak classifiers. The authors achieved an overall accuracy of 98.68% [225]. Zhang et al. [226] use Fourier transforms and PCA, followed by SVM to classify the samples of the MIAS dataset [81] achieving 92.16% accuracy. Pawar and Talbar [227] used a wrapper method for feature selection. The features of the MIAS dataset were extracted using the wavelet co-occurrence features from the four decomposition levels. The accuracy achieved was 89.47%. Xie et al. [228] proposed a novel CAD system for the diagnosis of breast cancer mammograms extracted from DDSM [82] and MIAS [81] datasets based on extreme learning machine (ELM) and SVM. The system achieves a better average accuracy of 96.02%.

Furthermore, in 2017, Khan et al. [229] used a bank of Gabor filters to extract the directional textural features and the SVM classifier to classify normal and abnormal samples of the MIAS dataset [81]. The accuracy achieved was ranged from 68% to 98%. Al-Salihy et al. [230] classified 569 mass samples (357 benign and 212 malignant) extracted from the WBCD dataset using decision tree classifiers such as RF DT, J48 DT, Function Tree, decision stump, Alternating DT, and best first tree algorithm. The highest and lowest accuracies achieved were 97.17% and 88% for the Function Tree and decision stump, respectively [230].

3.4.1.1 The Most Relevant ML Frameworks

In the year 2018, Seryasat and Haddadnia [231] extracted the morphological and texture features from the MIAS [81] and DDSM [82] datasets. The relevant features were selected and then classified using ensemble classifier achieving an accuracy of 94.8% and 92% for MIAS [81] and DDSM [82] datasets, respectively. Tasdemir et al. [232] extracted the features using wavelet transform, Haralick, and histogram of oriented gradient (HOG) descriptors. A wrapper-based feature selection method was implemented to reduce the number of dimensions and eliminate irrelevant features. In addition, several ML classifiers such as SVM and RF DT were used to classify the samples of the digital mammography dream challenge dataset [87]. The highest accuracy achieved was 87.5% using the RF DT classifier [232]. Wang et al. [233] used the SVM based ensemble learning algorithms to classify the samples of two datasets, (1) WDBC dataset including 569 instances (357 benign and 212 malignant) collected from the University of Wisconsin Hospitals, and (2) Surveillance, Epidemiology, and End Results (SEER) breast cancer dataset collected from National Cancer Institute's SEER program. The authors achieved an accuracy of 97.89% [233].

Additionally, in the year 2019, Dhahri et al. [234] employed the ensemble of the following classifiers, SVM, k-NN, DT, gradient boosting classifier (GB), RF DT, logistic regression (LR), AdaBoosting classifier, Gaussian Naïve Bayes (GNB), and LDA to classify the lesions of WBCD dataset. The accuracy achieved was 97.34% [234]. Ragab et al. [235] used image processing techniques to remove the pectoral muscle of the MIAS [81] and the digital mammography dream challenge dataset [87]. The authors extracted the features using some statistical metrics and classified them using single and multiple classifiers. The highest accuracy achieved was 99.7% [235].

Moreover, in the year 2020, Assiri et al. [236] proposed an ensemble classifier based on a majority voting mechanism. The performance of different state-of-the-art ML classification algorithms was evaluated for the WBCD dataset achieving accuracy of 99.42% [236].

Mohammed et al. [237] proposed an approach that improves the accuracy and enhances the performance of three different classifiers: J48 DT, Naïve Bayes (NB), and Sequential Minimal Optimization (SMO). The experiments were performed on the WBCD dataset achieving 99.56% using the SMO classifier [237].

Furthermore, recently, Mangukiya et al. [238] used the SVM, Decision Tree, NB, k-NN, Adaboost, extreme gradient boosting (XGBoost), and Random Forest to detect and classify breast cancer. The authors conducted their experiments on the WBCD dataset. The experiments showed that XGboost classifier offered the highest accuracy 98.24% [238]. Table 3.4 summarizes the ML techniques used recently for mammogram breast cancer classifications.

3.4.2 Deep Learning-based Frameworks

During the past few years, various contributions have been made in literature regarding the application of deep learning [17], [130] and exclusively through deep convolutional neural networks (DCNN) techniques for breast cancer diagnosis [21], [239]–[242]. This is because DCNN has surpassed the accuracy of almost all other traditional classification methods and even the human ability [243]. Moreover, DCNN is used to classify the breast tumours not only in mammogram modalities [244], [245] but also in MRI scans as in [246]–[250] and microscopic images such as the histopathological breast images as in [251]–[255] and [256]–[264].

Furthermore, several researchers studied the diagnoses of breast cancer in ultrasound images as in [265]–[267]. Moon et al. [268] proposed a CAD system for tumour diagnosis in ultrasound images using an image fusion method combined with different image content representations and an ensemble of different DCNN architectures. The DCNN-based methods include VGG Net, ResNet, and DenseNet [268]. The experiments were performed on private and open datasets [268]. The private dataset has 1687 total tumours including 953 benign and 734 malignant tumours. However, the open dataset has 697 total tumours including 437 benign lesions, 210 malignant tumours, and 133 normal images. When using the private dataset, the results for accuracy, sensitivity, specificity, precision, F1-score, and the AUC were 91.10%, 85.14%, 95.77%, 94.03%, 89.36%, and 0.9697 (96.97%), respectively [268]. On the other hand, the accuracy, sensitivity, specificity, precision, F1-score, and the AUC achieved, when using the open dataset were 94.62%, 92.31%, 95.60%, 90%, 91.14%, and 0.9711 (97.11%), respectively [268].

Table 3.4: Classification results for breast cancer using different ML algorithms.

Paper	Year	Classification	Dataset	Result
[215]	2005	SVM GRNN	Database of Nijmegen University Hospital	AUC = 0.98 – 0.978
[216]	2010	SVM	DDSM	AUC = 0.7 (70%)
[217]	2011	SVM	DDSM	Accuracy = 98%
[218]	2012	SVM	DDSM	Accuracy = 92%
[219]	2012	SVM	MIAS	Accuracy = 95%
[220]	2013	ANN – SVM	DDSM	Accuracy = 96% – 98%
[221]	2014	SVM	WBCD	Accuracy = 97.38%
[222]	2015	BPNN	MIAS – DDSM	Accuracy = 98% – 98.8%
[223]	2015	ANN	WDBC	Accuracy = 99.68%
[224]	2015	SVM	DDSM	AUC = 0.95 (95%)
[225]	2016	ANN – AdaBoosting	DDSM – MIAS	Accuracy = 98.68%
[226]	2016	SVM	MIAS	Accuracy = 92.16%
[228]	2016	SVM ELM	DDSM MIAS	Accuracy = 96.02%
[229]	2017	SVM	MIAS	Accuracy = 68% – 98%
[230]	2017	DT	WBCD	Accuracy = 88% – 97.17%
[231]	2018	Ensemble classifier	DDSM – MIAS	Accuracy = 92% – 94.8%
[232]	2018	SVM – RF DT	Digital mammography dream challenge dataset	Accuracy = 87.5%
[233]	2018	SVM	WDBC	Accuracy = 97.89%
[234]	2019	Ensemble classifier	WBCD	Accuracy = 97.34%
[235]	2019	Single and MCS	MIAS - Digital mammography dream challenge	Accuracy = 99.7%
[236]	2020	Ensemble classifier	WBCD	Accuracy = 99.42%
[237]	2020	J48 DT – NB – SMO	WBCD	Accuracy = 99.56%
[238]	2022	XGBoost	WBCD	Accuracy = 98.24%

Pi et al. [269] considered the heterogeneity of breast lesions and perform diagnoses using a multiple-instance learning method. This was performed by extracting and fusing the deep features of ResNet-50, Inception-v3, and DenseNet-169, followed by feature aggregation operators such as maximum, mean, and quantile [269]. The authors achieved an accuracy of 96.41% and 0.983 (98.3%) of AUC for classifying benign and malignant lesions of ultrasound images collected and obtained from the People's Hospital of DeYang City [269]. Wang et al. [270] proposed a CAD system based on DCNN to classify breast cancer lesions in automated breast ultrasound imaging. The proposed DCNN adopts a modified Inception-v3 architecture to provide efficient feature extraction in breast ultrasound imaging. The proposed DCNN was trained and evaluated on 316 breast lesions (135 malignant and 181 benign). The authors achieved an AUC, sensitivity, and specificity values of 0.9468 (94.68%), 0.886, (88.6%) and 0.876 (87.6%), respectively with five-fold cross-validation [270]. Irfan et al. [271] developed an algorithm based on dilated semantic segmentation network with morphological operation to segment ultrasonic breast lesions. The feature vectors acquired from DenseNet-201 and 24 layer DCNN employing parallel fusion were fused to categorize the nodules. These feature vectors are merged with the SVM classification and attained an accuracy of 98.9% [271].

Since the work presented in this thesis depends on classifying the breast lesions in mammogram image modalities. Therefore, in the following sub-sections, the most relevant classification results depending on the classification of individual features and deep feature fusion will be discussed.

3.4.2.1 Classification of Individual Features in Mammogram Images

This section provides a comprehensive survey of recently published works related to breast cancer classification using DCNN based CAD system. In the year 2016, Arevalo et al. [272] proposed a DCNN model that learns features for mammography mass lesions before feeding them to the classification stage. The researchers achieved an accuracy of 86% using 736 images of the BCDR-F03 [85] dataset. Wichakam and Vateekul [273] extracted the features using DCNN, then classify them using the SVM classifier. The sensitivity achieved was 98.44% using the INbreast dataset [84]. Jain and Levy [274] used AlexNet DCNN to classify benign and malignant masses in mammograms of the DDSM dataset [82] achieving an accuracy of 66% [274]. Yi and Dunnmon [275] used the GoogleNet architecture to classify benign and malignant breast mammogram tumours of a subset of the DDSM dataset [82]. The accuracy and AUC achieved were 85% and 0.91 (91%), respectively.

Moreover, in the year 2017, Jadoon et al. [276] proposed a model that classifies three classes; normal, benign, and malignant. The model proposed two methods for feature extraction, namely convolutional neural network discrete wavelet (CNN-DW) and convolutional neural network curvelet transform (CNN-CT). The authors classified the samples of the image retrieval in medical applications (IRMA) dataset using the SVM classifier. They achieved accuracy rated from 81.83% to 83.74% [276]. Carneiro et al. [277] developed an automated DL model to examine the two views of the mammogram CC and MLO of the two datasets DDSM [82] and INbreast [84]. The authors achieved AUC 0.9 (90%) and 0.7 (70%) for both semi-automated and fully automated techniques, respectively [277]. Zhu et al. [278] proposed an end-to-end trained AlexNet DCNN for mass classification based on the whole mammogram image and not the ROI. The results of the proposed method were validated on INbreast dataset [84]. The accuracy achieved was 84% for the end-to-end AlexNet DCNN. The authors then employed the idea of extracting the deep features and classify them using RF DT. In this case, the accuracy increased to 91% [278]. Zhang et al. [279] developed and evaluated DCNN models for whole mammography image classification introducing transfer learning and data augmentation techniques. The authors collected the images from the department of radiology, University of Kentucky. The AUC scored 0.73 (73%) [279]. Dhungel et al. [280]

proposed a multi-view deep residual neural network for the fully automated classification of mammograms. The authors achieved an AUC of 0.8 (80%) on classifying malignant and benign samples of the INbreast dataset [84]. Shen [281] used DCNN to classify the lesions of DDSM [82] and INbreast [84] datasets achieving an AUC of 0.91 (91%) and 0.96 (96%), respectively. Zhang et al. [282] extracted and classified the features of the DDSM dataset [82] using a fine-tuned AlexNet DCNN. The AUC and the sensitivity achieved were 0.8 (80%) and 72%, respectively. Duraisamy and Emperumal [283] extracted the features using a DCNN algorithm then classified the features using a well-known fully complex-valued relaxation network (FCRN) classifier. The authors used two datasets (1) MIAS [81] and (2) BCDR-F03 [85] datasets to perform their experiments. The accuracy and AUC achieved were 99% and 0.9815 (98.15%), respectively [283]. Platania et al. [284] proposed a framework for automated breast cancer detection and diagnosis, providing automated ROI detection and diagnosis using DCNN. The framework presented was tested on the DDSM [82] dataset achieving 90% and 93.5% for detection and classification accuracies, respectively [284].

In the year 2018, Al Hussien and Salem [285] proposed a DCNN architecture model to classify the samples of the INbreast dataset [84]. The accuracy and AUC achieved with 5-fold cross-validation were 80% and 0.78 (78%), respectively. Xi et al. [286] classified the mass and calcifications samples of the CBIS-DDSM dataset [86] using several fine-tuned DCNN architectures. The authors achieved 91.23%, 92.53%, 91.1%, and 91.8% of accuracy for AlexNet, GoogleNet, VGGNet, and ResNet, respectively [286]. Ribli et al. [287] used the fast R-CNN for the classification of benign and malignant lesions of the INbreast [84] and Digital Mammography Dream Challenge [87] datasets giving 0.95 (95%) and 0.85 (85%) of AUC, respectively. Al-Antari et al. [288] proposed a fully integrated CAD system including three DL stages detection, segmentation, and classification. First, an automatic DL You-Only-Look-Once (YOLO) detector was used to detect the breast lesion from the entire mammogram images extracted from INbreast [84] dataset. Second, the ROI was segmented using a full resolution convolutional network. Finally, a DCNN classifier was presented to distinguish between benign or malignant detected and segmented masses. The results were evaluated by fourfold cross-validation achieving 98.96% of accuracy [288]. Chougrad et al. [289] extracted and classified the lesions of DDSM [82], INbreast [84], and BCDR-F03 [85] datasets using the fine-tuned VGG-16, ResNet-50, and Inception-v3 DCNN architectures. The highest accuracy and AUC achieved were for Inception-v3 DCNN architecture, achieving 97.35%, 0.98 (98%), 95.5%, 0.987 (97%), and 96.67%, 0.96 (96%) for DDSM, INbreast, and BCDR-F03 datasets,

respectively. However, when merging the three datasets to increase the training set the accuracy increased to 98.94% for Inception-v3 DCNN as well [289].

Moreover, in the year 2019, Ting et al. [290] presented an algorithm called convolutional neural network improvement in breast cancer classification (CNNI-BCC) to assist medical experts in the classification of breast cancer lesions through the implementation of DCNN. The proposed method classified the lesions of the MIAS dataset [81] giving 90.5%, 89.47%, 90.71%, and 0.9 (90%) for the accuracy, sensitivity, specificity, and AUC, respectively [290]. Mechira et al. [291] used the AlexNet architecture to detect breast cancer from 8000 mammogram images extracted from the DDSM dataset [82]. The authors achieved 89.23% for accuracy, 91.11% for sensitivity, and 87.75% for specificity. Li et al. [292] proposed a new modification for DenseNet architecture by replacing the first convolution layers with an Inception model naming it DenseNet-II. Additionally, the authors classified the lesions of mammography breast cancer using AlexNet, VGGNet, GoogleNet, and DenseNet-II, giving an average accuracy of 94.55%. The authors performed their experiments on a dataset collected from the First Hospital of Shanxi Medical University [292]. Mendel et al. [293] extracted and classified the features of 78 mammogram lesions using pre-trained DCNN VGG-19 and SVM, respectively. The AUC achieved was 0.81 (81%). Agarwal et al. [294] used the transfer learning technique to fine-tune the pre-trained DCNN architectures; VGG-16, ResNet-50, and Inception-v3 to classify breast tumours in mammograms. The DCNNs were trained on the public dataset CBIS-DDSM [86] and tested on the INbreast dataset [84]. The accuracy achieved was 83.69% for both VGG-16 and ResNet-50 DCNN, however, it increased to 84.16% for Inception-v3 DCNN [294]. Shamy and Dheeba [295] initialized the K-means Gaussian mixture model and DCNN for the detection and classification of breast cancer. The first stage was to identify the ROI. The second stage was ROI texture extraction and feature optimization with the optimized feature selection algorithm. Finally, the predicted anomalies were classified as malignant or benign through DCNN. The proposed method was evaluated using MIAS [81] dataset achieving 95.8% accuracy [295]. Lotter et al. [296] cropped the mammogram samples into patches and labelled them with '1' and '0' for with and without lesion, respectively. A large number of patches were generated and then trained using the ResNet-50 DCNN architecture. The detection model outputs bounding boxes with corresponding classification scores, then trained using end-to-end RetinaNet on the full image. They used 2D and 3D mammogram samples extracted from DDSM [82] and a dataset collected from the UK named the UK-based OPTIMAM [83] mammography image database collected from many sites

[296]. McKinney et al. [297] constructed an AI system for breast cancer prediction consisting of an ensemble of three DL models including ReTinaNet, MobileNet-v2, and ResNet-50 architectures each operating on a different level of analysis. Each model produces a cancer risk score between '0' and '1' for the entire mammography case. The authors evaluated the performance of the AI system using two large, clinically representative datasets from the UK and USA. They compared the predictions of the system to those made by readers in routine clinical practice and show that performance exceeds that of individual radiologists. These observations were confirmed by an independently conducted reader study [297].

Moreover, in the year 2020, Alkhaleefah et al. [298] generated new training samples from the CBIS-DDSM [86] dataset by performing several augmentation techniques including rotation, flipping, zooming, changing brightness, and contrast. They classified the lesions with the fine-tuned DCNN VGG-19 using the transfer learning technique achieving accuracy and an AUC of 90.4% and 0.941 (94.1%), respectively [298]. Agarwal et al. [299] presented a mass detection framework based on the faster region-based convolutional neural network (Faster-RCNN) and applied it to the OPTIMAM mammography image database [83]. The TPR obtained was 0.93 (93%). The authors used the transfer learning technique to fine-tune the Faster-RCNN model [299]. Wu et al. [300] labelled the breast images using two types of labels (1) breast-level labels indicating whether there is a benign or malignant finding in each breast and (2) pixel-level labels indicating the location of biopsies malignant and benign findings. Moreover, they cropped each image view to a fixed size. The authors proposed a deep multi-view DCNN architecture based on four columns of ResNet-22 DCNN for training. Then they applied a sliding window to create two heat maps for each image containing an estimated probability of a benign or malignant finding for each pixel. The dataset used included 229,426 digital screening mammography exams, approved from the institutional review board (IRB), and was compliant with the health insurance portability. The AUC achieved was 0.895 (89.5%) [300].

Al-Antari et al. [301] used the YOLO detector to detect the breast lesion from the entire mammogram images extracted from DDSM and INbreast datasets. The lesions were classified using three fine-tuned end-to-end DCNN architectures ResNet-50, Inception ResNet-v2, and feedforward DCNN. The Inception ResNet-v2 DCNN achieved the highest accuracy of 97.5% and 95.32% for DDSM and INbreast datasets, respectively. However, the accuracies achieved for ResNet-50 were 95.83% and 92.55% for DDSM [86] and INbreast [84] datasets, respectively [301]. Arefan et al. [302] extracted the deep features using an end-to-end fine-

tuned GoogleNet DCNN architecture and classify them using the LDA classifier. The proposed approach was evaluated on a dataset consisting of 226 patient mammograms approved by the IRB. The AUC achieved for the end-to-end GoogleNet and GoogleNet-LDA were 0.67 (67%) and 0.73 (73%), respectively when classifying normal and abnormal lesions [302]. Wessels and Van der Haar [303] used the fine-tuned VGG-16 and GoogleNet DCNN architectures classify mammogram samples extracted from DDSM [82] and MIAS [81] datasets. The accuracies achieved to classify the normal and abnormal lesions from the DDSM dataset were 59.17% and 76.19% for VGG-16 and GoogleNet DCNN architectures, respectively. On the other hand, when distinguishing between the samples extracted from the MIAS dataset the accuracies increased yielding to 97.68% and 85.5% for VGG-16 and GoogleNet DCNN architectures, respectively [303]. Agnes et al. [304] proposed an end-to-end supervised DCNN approach named multiscale all convolutional neural networks (MA-CNN) for classifying mammogram images from the MIAS [81] dataset. MA-CNN model implements multilevel dilated convolutions that could extract both low-level and high-level contextual features from the image. The proposed system consists of two major phases: Context feature extraction and classification of mammograms into normal and abnormal patterns. The system achieved accuracy and an AUC of 96.47% and 0.99 (99%), respectively [304].

Recently, Saber et al. [305] employed the transfer learning technique to fine-tuned the pre-trained DCNN architectures to extract the deep features and then classify them using the SVM classifier. They used the VGG-16, VGG-19, ResNet-50, Inception-v3, and Inception ResNet-v2 architectures achieving an accuracy of 98.96%, 95.84%, 97.11%, 98.15%, and 94.23%, respectively [305]. The authors evaluated their results on the MIAS dataset [305]. Hamed et al. [306] used the ResNet, VGG, Xception, and Inception-v3 DCNNs to distinguish between benign and malignant tumours of the INbreast dataset [84]. The authors achieved 95% accuracy [306]. Malebary and Hashmi [307] proposed a novel technique based on k-mean clustering, Long Short-Term Memory (LSTM) network of Recurrent Neural Network (RNN), RF DT, and boosting techniques to classify the breast mass into benign, malignant, and normal [307]. The authors performed their experiments on the two datasets; DDSM and MIAS datasets achieving 96% and 95% for accuracy, respectively [307]. Table 3.5 summarizes the recent DL techniques using individual deep features of breast cancer classifications in mammograms. Moreover, the deep learning methods for individual breast cancer classification features were mentioned as in [308]–[317] and [318]–[325].

3.4.2.2 Classification of Deep Features Fusion in Mammogram Images

Recently, several researchers led to fusing the deep features to improve the classification accuracy and this was proved in many research papers [326]–[328]. Wang et al. [329] fused the deep features with morphological features, texture features, and density features; then classify them using an ELM classifier. The experiments were performed on 400 mammogram images extracted from a private hospital achieving accuracy and an AUC of 86.5% and 0.923 (92.3%), respectively. Khan et al. [330] proposed a multi-view feature fusion (MVFF) based CAD system using a feature fusion technique of four views for the classification of mammograms. The authors performed their experiments on the CBIS-DDSM [86] and MIAS [81] datasets. The deep features were extracted and fused from four fine-tuned DCNN architectures VGG-16, VGG-19, GoogleNet, and ResNet-50. The fused deep features were classified according to a three stages strategy for classifications: (1) classification of the mammogram (normal/abnormal), (2) classification of abnormality (mass/calcification), and (3) classification of pathology (benign/malignant). For the first classification stage the accuracy, AUC, sensitivity, and specificity achieved were 96.66%, 0.934 (93.4%), 96.31%, and 90.47%, respectively. On the other hand, the accuracy, AUC, sensitivity, and specificity for the second and third stages classifications achieved 92.29% and 80.56%, 0.923 (92.3%) and 0.769 (76.9%), 93.37% and 81.82%, 91.17% and 72.02%, respectively [330]. Song et al. [331] proposed a new CAD system to classify three classes, normal, benign, and malignant samples of the DDSM dataset [82]. The authors fused the deep features of GoogleNet, Inception-v2, and Inception with n × n convolution with handcrafted features. The handcrafted features included scoring features, GLCM, and histogram of oriented gradient (HOG) features. The features were classified by SVM and XGBoost classifiers. The authors first classified the end-to-end DCNN features achieving an accuracy of 82.84%; however, when classifying the fused features the accuracy increased reaching 92.8%. The results achieved by the XGBoost proved to be higher than those achieved by the SVM classifier [331].

Zhang et al. [332] proposed a simple and effective model called DE-Ada* which is an organic integration of multi-feature fusions, for breast mass classification. DE-Ada* consists of the following state-of-the-art technologies: “D” represents the discriminant correlation analysis (DCA) algorithm, “E” represents the modified effective range-based gene selection (ERGS) algorithm, “Ada” represents the well-known adaptive boosting (AdaBoosting) algorithm, and “*” represents two ensemble learning strategies. The authors fused the image features including Gist, the scale-invariant feature transform (SIFT), the HOG, and the local

binary pattern (LBP) features, with the deep features extracted from the fine-tuned DenseNet, VGGNet, and ResNet DCNN architectures. The authors used single and MCS classifiers to classify the fused features such as k-NN, SVM, DT, NB, AdaBoosting, and XGBoost. The experiments were performed on the CBIS-DDSM [86] and INbreast [84] datasets, achieving accuracies of 90.91% and 87.93%, respectively [332].

Yu et al. [333] designed a framework to fuse the deep features of the fine-tuned VGG-16 and VGG-19 DCNN architectures. The authors classified the lesions of the MIAS [81] dataset achieving an accuracy of 89.06%. Arora et al. [334] employed the transfer learning technique to extract and fuse the deep features of AlexNet, VGG-16, GoogleNet, ResNet-18, and Inception ResNet DCNN architectures from the CBIS-DDSM [86] dataset. The deep features were classified using ANN classifier achieving equal accuracy and an AUC of 0.88 (88%) [334]. Maqsood et al. [335] proposed a framework CAD system to identify breast cancer in mammogram. They extracted the deep features by determining the best layers which enhance the classification accuracy of the following DCNNs architectures: Inception ResNet-v2, Inception-v3, VGG-16, VGG-19, GoogleNet, ResNet-18, ResNet-50, and ResNet-101. Then they fused all the extracted feature vectors using the convolutional sparse image decomposition approach [335]. The proposed approach employed on DDSM, INbreast, and MIAS datasets and attained an average accuracy of 97.49% [335]. Zahoor et al. [336] extracted and fused the deep features of the fine-tuned architectures MobileNet-v2 and NasNet Mobile. They optimized the fused deep features using the modified entropy whale optimization algorithm. The authors tested their proposed technique on the three publicly datasets: INbreast, MIAS, and CBISDDSM, achieving an accuracy of 99.7%, 99.8%, and 93.8%, respectively [336]. Table 3.6 summarizes the recent DL techniques using deep features fusion for breast cancer classifications in mammograms. Furthermore, DCNN can be used to classify tumours in other organs such as blood [337], liver [338], brain [339], etc.

3.5 Summary

With the advancement in digital computing technology, many researchers have combined image processing and pattern recognition to develop CAD systems to assist radiologists in the diagnosis process. Thus, in this chapter, the five steps used to build-up a CAD system were discussed. Firstly, the images go through the enhancement step to improve their contrast and suppress the noise. Therefore, some image enhancement techniques were presented. Moreover, in order to enhance the performance of the CAD system; data

augmentation was applied to increase the data, thus, a brief description of data augmentation was given. Secondly, the ROI was determined from the image. Then, some features were extracted from the ROI to remove irrelevant information from it. Afterward, these features were classified according to their type. Finally, the evaluation of the diagnostic test is immensely important in modern medicine for confirming not only the presence of disease but to rule out the disease in healthy samples. Therefore, some scores such as accuracy, AUC, sensitivity, specificity, precision, and F1-score were defined to assess a classifier.

DL had been successfully used in several AI applications such as computer vision and pattern recognition. Recently, DCNN had gained attention in the medical area and especially for classifying mammography breast cancer. Therefore, in this work, DCNN was used as a feature extractor and classification. The structure of the DCNN was discussed, which consists of convolutional, pooling, and fully connected layers. Furthermore, the transfer learning technique was introduced to be able to use DCNN architectures in various applications. Finally, this chapter presented the results of the state-of-the-art of the traditional ML and DCNNs techniques for breast cancer classification. Besides, this chapter presented and discussed the available mammogram breast datasets that were used in the state-of-the-art techniques.

Table 3.5: A comparative view of the classification results of breast cancer using individual deep features.

Reference	Year	Feature Extraction	Classification	Dataset	Result
[272]	2016	DCNN		BCDR-F03	Accuracy = 86%
[273]	2016	DCNN	SVM	INbreast	Sensitivity = 98.44%
[274]	2016	AlexNet		DDSM	Accuracy = 66%
[275]	2016	GoogleNet		DDSM	Accuracy = 85%, AUC = 0.91
[276]	2017	CNN-DW-CNN-CT	SVM	IRMA	Accuracy = 81.83% - 83.74%
[277]	2017	DCNN		DDSM INbreast	AUC = 0.9 - 0.7
[278]	2017	AlexNet DCNN		INbreast	Accuracy = 84%
		AlexNet DCNN	RF DT		Accuracy = 91%
[279]	2017	DCNN		University of Kentucky	AUC = 0.73
[280]	2017	ResNet		INbreast	AUC = 0.8
[281]	2017	DCNN		DDSM INbreast	AUC = 0.91- 0.96
[282]	2017	AlexNet		DDSM	AUC = 0.8, Sensitivity = 72%
[283]	2017	DCNN	FCRN	MIAS BCDR-F03	Accuracy = 99% AUC = 0.9815
[284]	2017	DCNN		DDSM	Accuracy = 93.5%
[285]	2018	DCNN		INbreast	Accuracy = 80%, AUC = 0.78
[286]	2018	AlexNet, **GoogleNet**, VGGNet, ResNet		CBIS-DDSM	Accuracy = 92.53%
[287]	2018	R-CNN		INbreast Digital Mammography Dream Challenge	AUC = 0.95 - 0.85
[288]	2018	DCNN		INbreast	Accuracy = 98.96%
[289]	2018	VGG-16, ResNet-50, **Inception-v3**		DDSM, INbreast, BCDR-F03	Accuracy = 98.94%
[290]	2019	CNNI-BCC		MIAS	Accuracy = 90.5%, AUC = 0.9
[291]	2019	AlexNet		DDSM	Accuracy = 89.23%
[292]	2019	DesNet-II, AlexNet, VGGNet, GoogleNet		First Hospital of Shanxi Medical University	Accuracy = 94.55%
[293]	2019	VGG-19	SVM	Private dataset	AUC = 0.81
[294]	2019	VGG-16, ResNet-50, and Inception-v3		CBIS-DDSM INbreast	Accuracy = 84.16%
[295]	2019	DCNN		MIAS	Accuracy = 95.8%
[298]	2020	VGG-19		CBIS-DDSM	Accuracy = 90.4%, AUC = 0.941
[299]	2020	Faster-RCNN		OPTIMAM	TPR = 93%
[300]	2020	ResNet-22		Private dataset	AUC = 0.895
[301]	2020	ResNet-50 – Inception ResNet-v2		DDSM INbreast	Accuracy = 95.83% - 92.55% Accuracy = 97.5% - 95.32%
[302]	2020	GoogleNet		Private dataset	AUC = 0.67
		GoogleNet	LDA		AUC = 0.73
[303]	2020	VGG-16 and GoogleNet DCNNs		DDSM MIAS	Accuracy = 59.17% - 97.68% Accuracy = 76.19% - 85.5%
[304]	2020	MA-CNN		MIAS	Accuracy = 96.47% AUC = 0.99
[305]	2021	VGG-16, VGG-19, ResNet-50, Inception-v3, Inception ResNet-v2	SVM	MIAS	Accuracy = 98.96%, 95.84%, 97.11%, 98.15%, and 94.23%
[306]	2021	ResNet, VGG, Xception, and Inception-v3		INbreast	Accuracy = 95%
[307]	2021	LSTM of RNN, RF DT, Boosting		DDSM – MIAS	Accuracy = 96% - 95%

Table 3.6: A comparative view of the classification results of breast cancer using deep feature fusion.

Reference	Year	Feature Extraction	Classification	Dataset	Result
[329]	2019	Fused deep features with morphological features	ELM	Private dataset	Accuracy = 86.5%, AUC = 0.923
[330]	2019	Deep features fusion of VGG-16, VGG-19, GoogleNet, ResNet-50		CBIS-DDSM MIAS	Accuracy = 96.6%, AUC = 0.934
[331]	2020	GoogleNet Inception-v2	XGBoost	DDSM	Accuracy = 92.8%
[332]	2020	The fusion of Gist, SIFT, HOG, LBP, VGG, ResNet, and DenseNet features	SVM XGBoost NB k-NN DT AdaBoosting	CBIS-DDSM INbreast	Accuracy = 90.91% - 87.93%
[333]	2020	The fusion of VGG-16 and VGG-19		MIAS	Accuracy = 89.06%
[334]	2020	Deep features fusion of AlexNet, VGG-16, GoogleNet, ResNet-18, Inception ResNet	ANN	CBIS-DDSM	AUC = 0.88
[335]	2022	Deep features fusion of Inception ResNet-v2, Inception-v3, VGG-16, VGG-19, GoogleNet, ResNet-18, ResNet-50, ResNet-101		DDSM INbreast MIAS	Accuracy = 97.49%
[336]	2022	Deep features fusion of Mobilenet-v2 and NasNet Mobile		INbreast MIAS CBISDDSM	Accuracy = 99.7%, 99.8%, 93.8%

Chapter 4

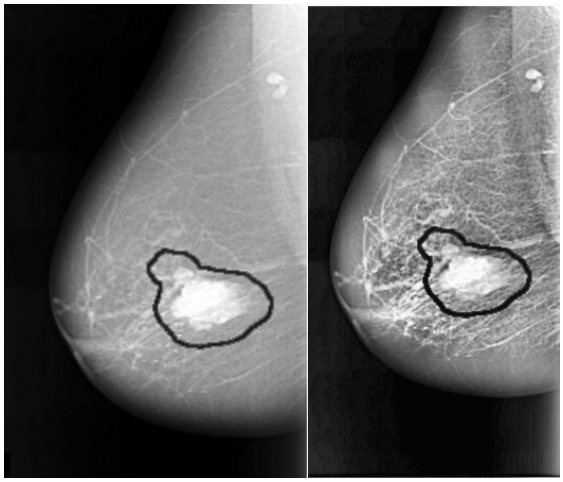
AlexNet-SVM: A Framework for Breast Cancer Segmentation and Classification

4.1 Introduction

In this chapter, a new methodology for classifying benign and malignant mass tumours using DL and some segmentation techniques are introduced. Two segmentation approaches are used; the first one involves determining the ROI using circular contours, while the second uses the adaptive threshold method. The AlexNet DCNN is used as a feature extractor and a classifier. However, it is fine-tuned to classify two classes instead of 1000 classes. Moreover, the last FC layer of the DCNN is connected to SVM to obtain better classification results. The proposed framework was trained and tested on the DDSM [82] and the CBIS-DDSM [86] datasets.

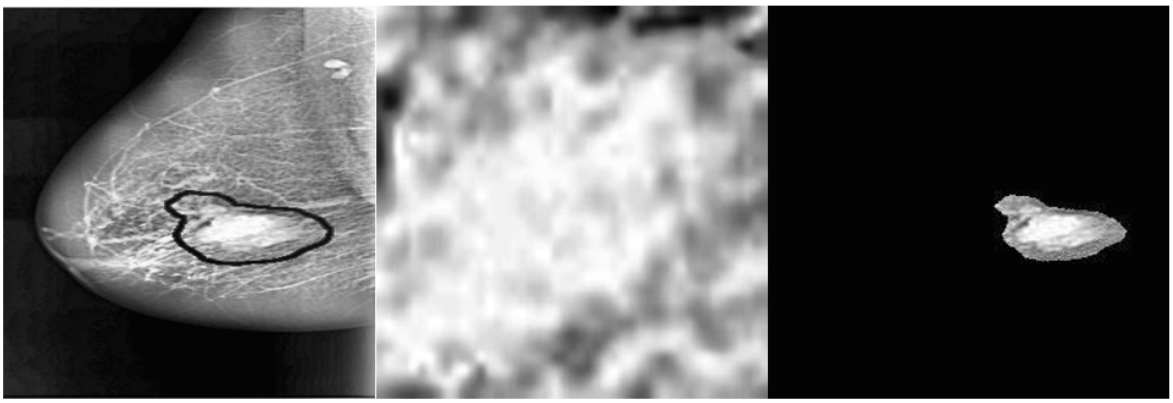
4.2 Methodology

As stated in Chapter 3, the CAD system consists of five main modules, (1) image enhancement, (2) image segmentation, (3) feature extraction, (4) feature classification, and finally, (5) evaluating the classifier. Consequently, for the proposed framework, the images are enhanced using the CLAHE method as shown in Figure 4.1. The ROI is then extracted from the original mammogram image by circular contours and adaptive threshold method. As the tumours in the DDSM [82] dataset are labelled with a red contour as illustrated in Figure 4.2-(a). Accordingly, these contours are determined by examining the pixel values of the tumour and use them to extract the ROI as shown in Figure 4.2-(b). However, for the adaptive threshold method, the first step to extract the ROI is to determine the tumour region by a threshold value, which is a value determined with respect to the red colour pixel. After some trials, the threshold was set to 76 for all the images regardless of the size of the tumour. Then, the biggest area within this threshold along the image was determined and the tumour was cropped as shown in Figure 4.2-(c).



(a) (b)

Figure 4.1: An example of image enhancement; (a) Original malignant mass case extracted from DDSM [82] dataset and (b) enhanced image using CLAHE method.



(a) (b) (c)

Figure 4.2: The image segmentation techniques; (a) Original enhanced malignant mass case extracted from DDSM [82], (b) enhanced ROI extracted using circular contours, and (c) enhanced ROI extracted by the adaptive threshold method.

The steps for the adaptive threshold method can be summarised as follows:

1) Convert the original mammogram greyscale image into a binary image using the adaptive threshold technique.

2) Binary image objects are labelled and the number of pixels is counted.

All binary objects are removed except for the largest one, which is the tumour with respect to the threshold. The largest area is enclosed within the red contour labelled around the tumour.

3) After the algorithm checks all pixels in the binary image, the largest area pixels within the threshold are set to "1", otherwise all other pixels are set to "0."

4) The resulting binary image is multiplied with the original mammogram image to get the final image without taking into consideration the rest of the breast region or any other artefacts.

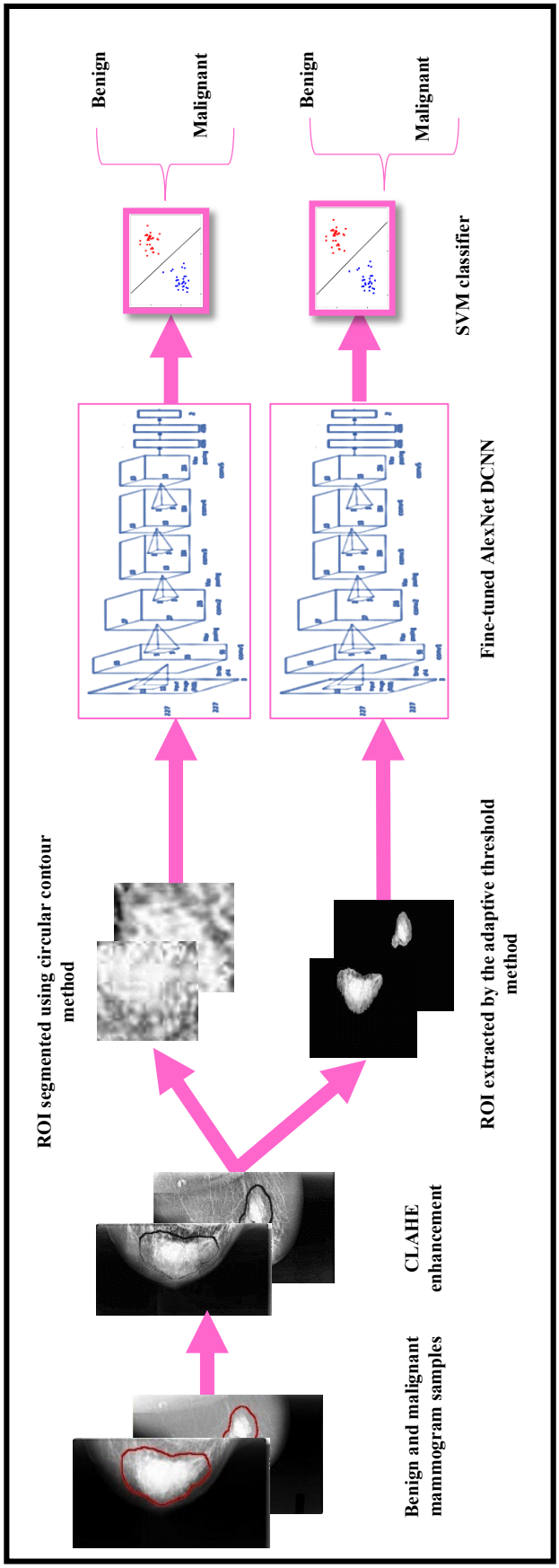


Figure 4.3 : The proposed framework.

In the feature extraction step, the features are extracted using a fine-tuned AlexNet DCNN architecture. This is performed by replacing the last FC layer with a new layer for the classification of two classes, benign and malignant masses. The proposed framework uses two classifiers fine-tuned AlexNet DCNN and SVM. The SVM is used because it achieved high classification rates in the breast cancer classification problem. The proposed framework is shown in Figure 4.3.

4.3 Experimental Setup

All the experiments were carried out using MATLAB R2018a with an academic license provided by the University of Strathclyde. Experiments were applied to the DDSM [82] and the CBIS-DDSM [86] mammogram datasets.

4.3.1 Augmentation

Due to the insufficient amount of samples and the classifiers perform well, giving high accuracy rates when training on a huge number of samples. Therefore, data augmentation technique and especially the rotation form was applied to the samples. Each original image was rotated by four angles, which are 0, 90, 180, and 270 degrees. Accordingly, each original image was augmented to four images.

4.3.2 Parameter Setting

To train the fine-tuned AlexNet DCNN, numerous network hyper-parameters were tuned whereas others were kept unchanged from their default values as the performance of AlexNet was insensitive to the parameter selection. These network hyper-parameters were the mini-batch size, the number of epochs, the initial learning rate, and the validation frequency. The maximum number of epochs was set to 20, however, to avoid overfitting an early stopping criteria was set in which epoch number 12 proved to be the best epoch. Moreover, the L2-regularization or the weight decay was set to 0.0005. The L2-regularization was considered the most common type of regularization and it was used to avoid overfitting. The mini-batch size was chosen to be 10, which was selected to fits the memory of the GPU. In addition, the initial learning rate was chosen to be 10^{-5}, which achieved the highest classification accuracy while lessening the training time. Additionally, the validation frequency was set to 128. These configurations were to confirm that the parameters were fine-tuned for the diagnosis of medical data. The optimization algorithm used was the stochastic gradient descent with momentum (SGDM) [15] learning algorithm.

4.4 Results and Discussions

This framework presented two approaches for segmentation techniques to classify mass tumours in a mammogram using DCNN. The input layer of the AlexNet architecture requires that the size of the image is 227 × 227 × 3. Therefore, there was a pre-processing step to convert all the input images regardless of their sizes to the size required by the AlexNet. The features went through the DCNN and SVM for classification, in which the last FC layer was connected to SVM with different kernel functions to obtain better results. When using the end-to-end DCNN seventy percent of the images were used for training and the rest for testing and validation as this is the common ratio used in the classification problem. The number of training, testing, and validation samples for each segmentation technique is shown in Table 4.1. However, when classifying the deep features using SVM five fold cross-validation was used giving a ratio of 80%:20% for training and testing samples. This means that SVM with different kernels classifiers were taught with four folds and verified by the remaining fold. Thus, the model was taught five times and the testing classification accuracy was calculated for each time then averaged. The results computed for the two used datasets will be presented and discussed in the following sub-sections.

4.4.1 DDSM Dataset

A subset from the DDSM was extracted to apply the proposed methods. For the first segmentation technique; the ROI was cropped using the circular contour method that was already labelled with a red contour. The accuracy of the new-trained end-to-end AlexNet DCNN was only 71.01%. Whereas, when classifying the extracted deep features using SVM the results were better. The accuracy with the linear kernel function was 78.8%, which was the highest value compared to the other kernels. Moreover, the AUC, sensitivity, specificity, precision, and F1-score reached 0.88 (88%), 0.770 (77%), 0.810 (81%), 0.822 (82.22%), and 0.796 (79.6%), respectively, which proved to be the highest values compared to the other kernels too, as it was obvious from Table 4.2. Figure 4.4-(a) and Figure 4.4-(b) demonstrate the SVM classification accuracy between benign and malignant tumours samples and the ROC curve computed in this case.

Table 4.1: The number of training, testing, and validation samples for all the datasets used.

	Training	Testing	Validation	Total
DDSM (ROI cropped using circular contour)	1580	338	338	2256
DDSM (ROI using adaptive threshold)	1288	276	276	1840
CBIS-DDSM	3690	791	791	5272

Furthermore, for the second segmentation technique; the adaptive threshold was determined using the red contour surrounding the tumour area. The accuracy of the end-to-end DCNN reached only 69.2%. However, the accuracy of the SVM classifier with linear kernel function, increased to 81.2% with AUC equals 0.880 (88%). This was clear in Figure 4.4-(c) and in the computed ROC curve shown in Figure 4.4-(d). The SVM with linear kernel function revealed to be the highest values compared to the others as well as clear in Table 4.3. Moreover, the sensitivity, specificity, precision, and F1-score achieved 0.783 (78.3%), 0.845 (84.5%), 0.860 (86%), and 0.820 (82%), respectively. The deep features extracted from the DCNN is visualised in Figure 4.5. In this figure, the features from the first and fifth convolutional layers were visualised as an example.

Therefore, for the DDSM samples, the accuracy of the end-to-end new-trained AlexNet DCNN architecture for the first segmentation method was higher than that of the second method. Additionally, when classifying the benign and malignant masses features extracted from AlexNet DCNN by SVM, the accuracy for the adaptive threshold method was higher than the cropped ROI using the circular contour method. However, the SVM with linear kernel function achieved the highest accuracy for both segmentation techniques compared to the other kernel functions. Furthermore, the AUC for both segmentation methods were the same. One can easily notice this from the ROC curves shown in Figure 4.4-(b) and Figure 4.4-(d) of the first and second segmentation techniques, respectively. On the other hand, when calculating the sensitivity, specificity, precision, and F1-score for each SVM kernel function for both segmentation techniques, it was proved that the kernel with the highest accuracy has all the other scores high as well as it was clear in Table 4.2 and Table 4.3. Finally, all the results obtained for the classification of benign and malignant masses for both segmentation techniques for the DDSM dataset were summarised in Table 4.4.

Table 4.2: The classification scores of SVM with different kernel functions for cropping the ROI using circular contour for the DDSM dataset.

SVM Kernel Functions	**Cropping ROI using Circular Contour**					
	Accuracy (Std)	**AUC (Std)**	**Sensitivity (Std)**	**Specificity (Std)**	**Precision (Std)**	**F1-score (Std)**
Linear	**78.8%** (0.004)	**0.880** (0.001)	**0.770** (0.01)	**0.810** (0.009)	**0.822** (0.014)	**0.796** (0.004)
Quadratic	77% (0.01)	0.857 (0.011)	0.765 (0.016)	0.787 (0.016)	0.796 (0.017)	0.779 (0.013)
Cubic	77.6% (0.004)	0.860 (0.001)	0.767 (0.003)	0.789 (0.007)	0.796 (0.009)	0.782 (0.005)
Medium Gaussian	78.2% (0.011)	0.874 (0.009)	0.772 (0.011)	0.795 (0.009)	0.802 (0.008)	0.787 (0.009)
Coarse Gaussian	75.8 % (0.02)	0.850 (0.017)	0.729 (0.03)	0.802 (0.006)	0.830 (0.014)	0.776 (0.013)

Table 4.3: The scores of SVM with different kernel functions for the adaptive threshold method for the DDSM dataset.

SVM Kernel Functions	Adaptive Threshold Segmentation					
	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
Linear	**81.2%** (0.005)	**0.880** (0.005)	**0.783** (0.007)	**0.845** (0.007)	**0.860** (0.007)	**0.820** (0.005)
Quadratic	78.4% (0.008)	0.854 (0.009)	0.771 (0.025)	0.81 (0.019)	0.823 (0.022)	0.793 (0.013)
Cubic	77.4% (0.004)	0.839 (0.008)	0.762 (0.006)	0.786 (0.006)	0.795 (0.009)	0.779 (0.005)
Medium Gaussian	78.7% (0.008)	0.858 (0.007)	0.762 (0.007)	0.813 (0.016)	0.828 (0.02)	0.794 (0.01)
Coarse Gaussian	78.2% (0.014)	0.866 (0.013)	0.745 (0.014)	0.826 (0.015)	0.850 (0.013)	0.795 (0.013)

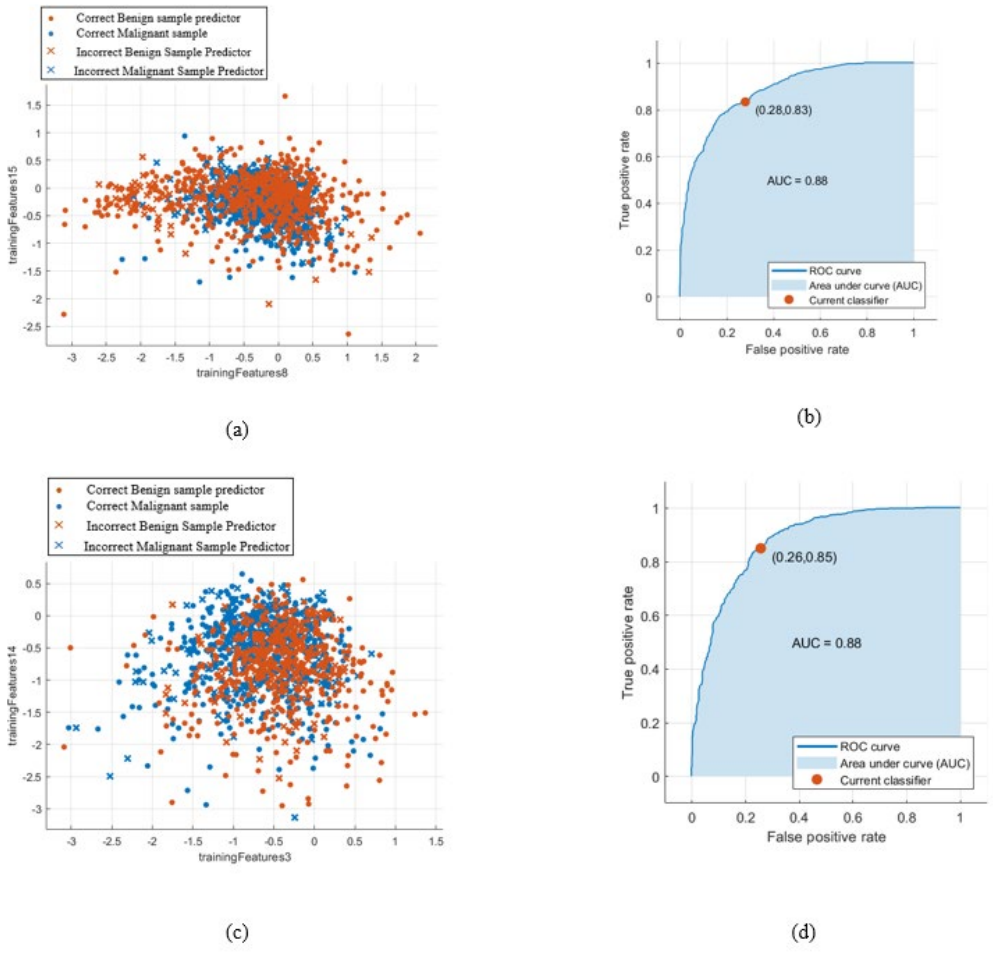
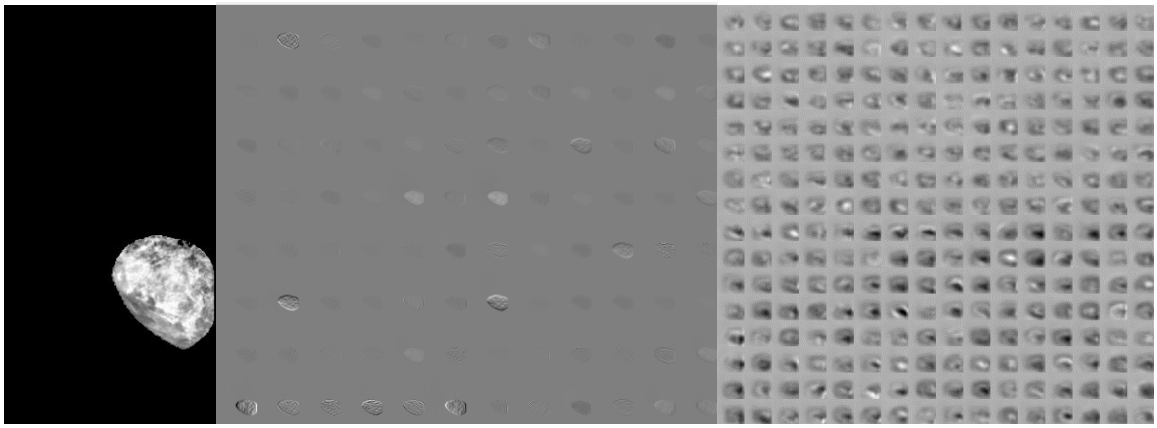


Figure 4.4: The SVM classification between benign and malignant masses and the computed ROC for both segmentation techniques for the DDSM dataset.

(a) The SVM classification between benign and malignant masses segmented by the circular contour method, (b) The computed ROC for the circular contour method, (c) SVM classification between benign and malignant masses segmented by the adaptive threshold method, and (d) The computed ROC for the adaptive threshold method.



(a) (b) (c)

Figure 4.5: Visualising the deep features of the AlexNet DCNN segmented by the adaptive threshold method.

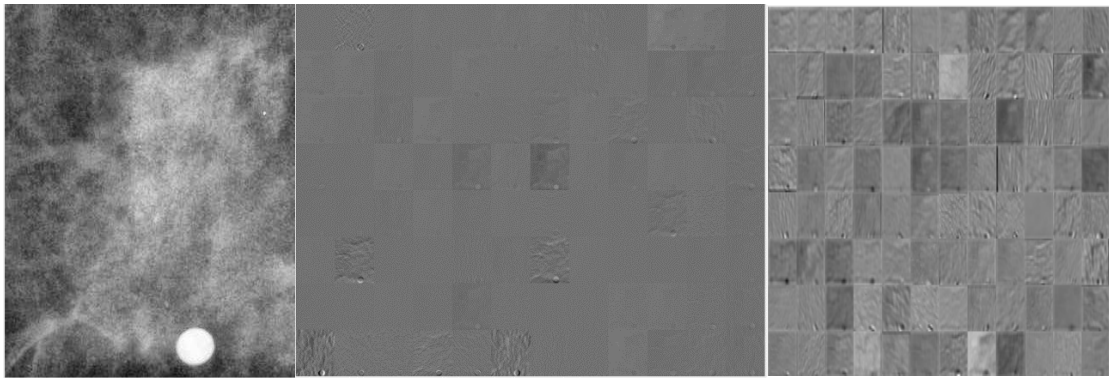
(a) Benign ROI image extracted from DDSM dataset, (b) the deep features from the first convolutional layer, and (c) the deep features from the fifth convolutional layer.

Table 4.4: The summary of the results obtained to classify benign and malignant masses for the DDSM dataset.

	Segmentation Techniques	
	Cropping ROI using Circular Contour	**Adaptive Threshold**
Trained DCNN Accuracy	71.01%	69.2%
SVM Accuracy	**78.8%**	**81.2%**
AUC	**0.880**	**0.880**
Sensitivity	0.770	**0.783**
Specificity	0.810	**0.845**
Precision	0.820	**0.860**
F1-score	0.796	**0.820**

4.4.2 CBIS-DDSM Dataset

For this dataset, the samples were only enhanced using CLAHE, this was because the samples of this dataset were already segmented. The accuracy achieved for the end-to-end fine-tuned AlexNet DCNN became 73.6%, which was higher than the DDSM samples. Then the deep features were extracted by AlexNet DCNN as visualised in Figure 4.6. In this figure, the first and second convolutional layers were visualised as an example. Additionally, when classifying the deep features extracted from the DCNN using the SVM classifier the accuracy reached 87.2% using the medium Gaussian kernel function as illustrated in Table 4.5. Besides, the AUC was 0.94 (94%) as shown in the ROC curve in Figure 4.7. This time the SVM with the Medium Gaussian achieved the highest values for all the scores compared to other kernel functions as demonstrated in Table 4.5. The sensitivity, specificity, precision, and F1-score reached 0.863 (86.3%), 0.878 (87.8%), 0.880 (88%), and 0.872 (87.2%), respectively. As it was obvious that all the values achieved for the CBIS-DDSM were higher than that of the DDSM dataset, this was because the data of the CBIS-DDSM were already segmented.



(a) (b) (c)

Figure 4.6: Visualising the deep features of the AlexNet DCNN for the CBIS-DDSM dataset.

(a) Malignant ROI image, (b) the deep features from the first convolutional layer, and (c) the deep features from the second convolutional layer.

Table 4.5: The evaluation scores calculated for SVM with different kernel functions for the CBIS-DDSM dataset.

SVM Kernel Functions	CBIS-DDSM Dataset					
	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
Linear	86.3% (0.004)	0.940 (0.001)	0.854 (0.002)	0.872 (0.005)	0.874 (0.006)	0.865 (0.004)
Quadratic	84.1% (0.012)	0.914 (0.015)	0.842 (0.013)	0.85 (0.013)	0.851 (0.012)	0.846 (0.012)
Cubic	83% (0.014)	0.902 (0.014)	0.830 (0.013)	0.831 (0.014)	0.831 (0.014)	0.831 (0.014)
Medium Gaussian	**87.2%** (0.004)	**0.940** (0.001)	**0.863** (0.001)	**0.878** (0.006)	**0.880** (0.007)	**0.872** (0.004)
Coarse Gaussian	86.2% (0.003)	0.937 (0.005)	0.860 (0.005)	0.877 (0.001)	0.880 (0.001)	0.868 (0.003)

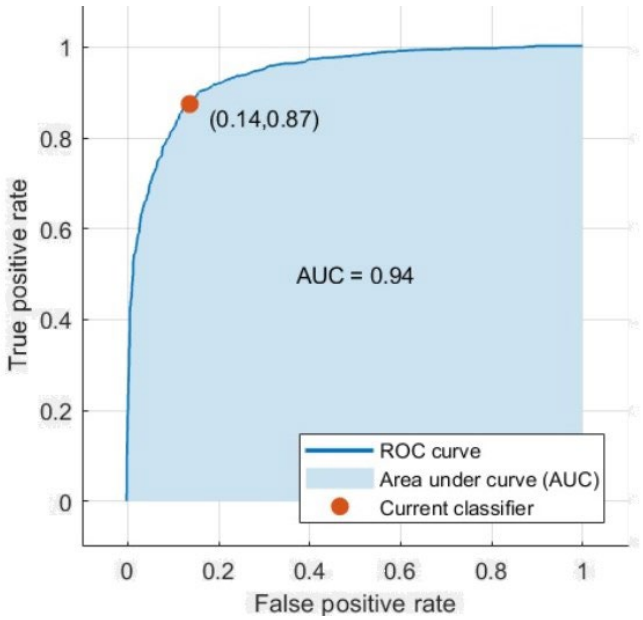


Figure 4.7: The computed ROC curve for the CBIS-DDSM dataset.

Table 4.6: The ANOVA test details for the different kernel functions of the SVM classifier for the circular contour method of the DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00535	4	0.00134	11.09	< 0.001
Error	0.00543	45	0.0012		
Total	0.01079	49			

Table 4.7: The ANOVA test details for the different kernel functions of the SVM classifier for the adaptive threshold method of the DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00848	4	0.00212	30.54	< 0.001
Error	0.00312	45	0.00007		
Total	0.0116	49			

Table 4.8: The ANOVA test details for the different kernel functions of the SVM classifier for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.01159	4	0.0029	41.92	< 0.001
Error	0.00311	45	0.00007		
Total	0.01471	49			

To test and validate the statistical significance of the results, a one-way analysis of variance (ANOVA) test was performed on all the results obtained from the repeated five fold cross-validation process. The null hypothesis Ho for all classification was that the mean accuracies of all SVM kernel classifiers were the same. This test was performed on all the accuracy results of the different kernels of the SVM classifier to test the statistical significance between them. Table 4.6 and Table 4.7 show the ANOVA test for the first and second segmentation techniques for the DDSM dataset, respectively. On the other hand, Table 4.8 shows the ANOVA for the accuracies obtained from different SVM kernels for the CBIS-DDSM dataset. It can be observed from Tables 4.6–4.8 that the p-values achieved were lower than α, where $\alpha = 0.05$. Therefore, it can be concluded that there was a statistically significant difference between the accuracies of the classifiers.

Finally, the proposed framework has been compared with other state-of-the-art that have the same conditions to prove the efficiency of the proposed method as discussed in Chapter 3. Regarding the DCNN AlexNet architectures, the results have shown that the framework proposed recorded the highest AUC, which was equal to 0.94 (94%) for the CBIS-DDSM dataset compared to Huynah et al. [159] and Jiang et al. [163]. The former achieved AUC 0.81 (81%) while the latter achieved 0.83 (83%). Huynah et al. [159] applied their experiments on 219 breast lesions collected from the University of Chicago medical center. The AlexNet with the transfer learning method was also used. However, Jiang et al. [163] used the BCDR-F03 dataset. They performed their tests on 736 mass cases. The ROI was extracted using the Otsu

segmentation algorithm. Besides, the transfer learning was used to classify two classes instead of 1000 like in this proposed framework.

4.5 Summary

The goal of this work was to classify benign and malignant lesions in mammograms by proposing a new framework. Two segmentation techniques were suggested and applied only to the DDSM dataset. However, for the CBIS-DDSM dataset, the data provided was already segmented so, therefore, no need for the segmentation step.

In the first technique, the ROI was cropped from the original image using circular contour. This was because the tumours in the DDSM dataset were labelled with a red contour. Whereas, in the second technique, the adaptive threshold method was used by setting a threshold, which was found to be equal to 76, and then determining the largest area including this threshold. In the feature extraction step, the DCNN was used. The AlexNet was retrained to distinguish between two classes and its parameters were changed to classify medical images.

The accuracy of the end-to-end DCNN of the circular contour method was higher than that of the adaptive threshold method by 1.8% using the DDSM dataset. To achieve better accuracy, the last fully connected layer in the DCNN was replaced by SVM. When comparing the two segmentation techniques for the DDSM dataset it was found that the SVM with linear kernel function for the adaptive threshold method provided promising results. Moreover, when using the samples of the CBIS-DDSM dataset, the accuracy of the end-to-end DCNN process increased to those of the DDSM samples. In addition, the accuracy and AUC of the extracted features classified by the SVM with medium Gaussian kernel function increased as well.

Chapter 5

Deep Features Fusion Framework for Breast Cancer Classification

5.1 Introduction

The novelty of this chapter lies in the design of an efficient framework based on the extraction, fusion, and classification of different deep features using DL techniques. This framework is evaluated using two datasets, CBIS-DDSM [86] and MIAS [81]. The samples of the datasets are segmented; hence segmenting the ROI as performed during the preprocessing for the method in Chapter 4 is no longer required. Several papers in the literature employed individual DCNNs to classify breast cancer in their CAD systems [274], [275], [286], [291], [293], [294], and [301] where the classification accuracies between 68% – 94% were not sufficient for a reliable and powerful CAD system. However, other papers proposed the use of feature fusion from several DCNNs of different architectures [257], [330], [331]. Although the fusion techniques could improve the accuracy to 92.8% – 97.67%, however, the combination of deep features contributed the most to the improved performance was not analysed. Moreover, they did not investigate how to reduce the computational cost of the CAD system. To tackle these drawbacks, a framework is proposed to explore the fusion of various features extracted from different DCNNs for choosing the best combination of the features, which improves the accuracy of the framework. Moreover, the proposed framework used the principal component analysis (PCA) to reduce the feature dimension as well as the associated computational cost. These procedures are made through four different scenarios as discussed in the following sections.

5.2 Methodology

First, the images are enhanced using the CLAHE enhancement technique and the ROIs are determined. The samples of the datasets used were already segmented using the information provided by the dataset. Then a framework is constructed using four different scenarios. For the first scenario, five end-to-end pre-trained DCNNs are used; AlexNet [23], GoogleNet [24], the ResNet-18, ResNet-50, and ResNet-101 [26]. The input layer of each of the five DCNN architectures constructed requires a specific image size. For example, the input layer of the GoogleNet architecture needed was 224 × 224 × 3. Thus, a pre-processing step is required to

change all image sizes to the size required for each DCNN architecture. Afterward, in the second scenario, the deep features of the DCNNs are extracted and fed to an SVM classifier with different kernel functions. This is performed to test if classifying the deep features with another classifier would enhance the accuracy of the end-to-end DCNN. In the third scenario, four sets of deep features are generated. These features include a different combination of deep features extracted from the five DCNNs architecture. This scenario is implemented to investigate if fusing deep features could enhance the accuracy of the SVM classifiers. It also selects the best combination of deep features, which improves the classification results. Deep feature fusion leads to a large feature space, therefore, in the fourth scenario, the effect of using a feature reduction method such as PCA is tested. PCA is applied to the feature sets to remove the irrelevant features and reduce the feature vector length. Additionally, the PCA reduces the computational cost as well as. The number of principal components is chosen in a sequential forward strategy. Figure 5.1 shows the proposed framework.

The PCA technique reduces the number of observed variables to a small number of principal components that still contain most of the information of the large set [340]. PCA is performed using the variance-covariance structure of a set of variables through linear combinations. It is used when variables are highly correlated, and it is suitable for data sets in multiple dimensions [340]. PCA provides a powerful tool for data analysis and pattern recognition. It is used frequently in signal and image processing [340]. The PCA technique can be summarised in the following steps: (1) subtract the mean from each of the data dimensions, (2) calculate the covariance matrix, (3) calculate the eigenvectors and the eigenvalues of the covariance matrix, (4) choose the components and form a feature vector, and (5) generate the new dataset [340].

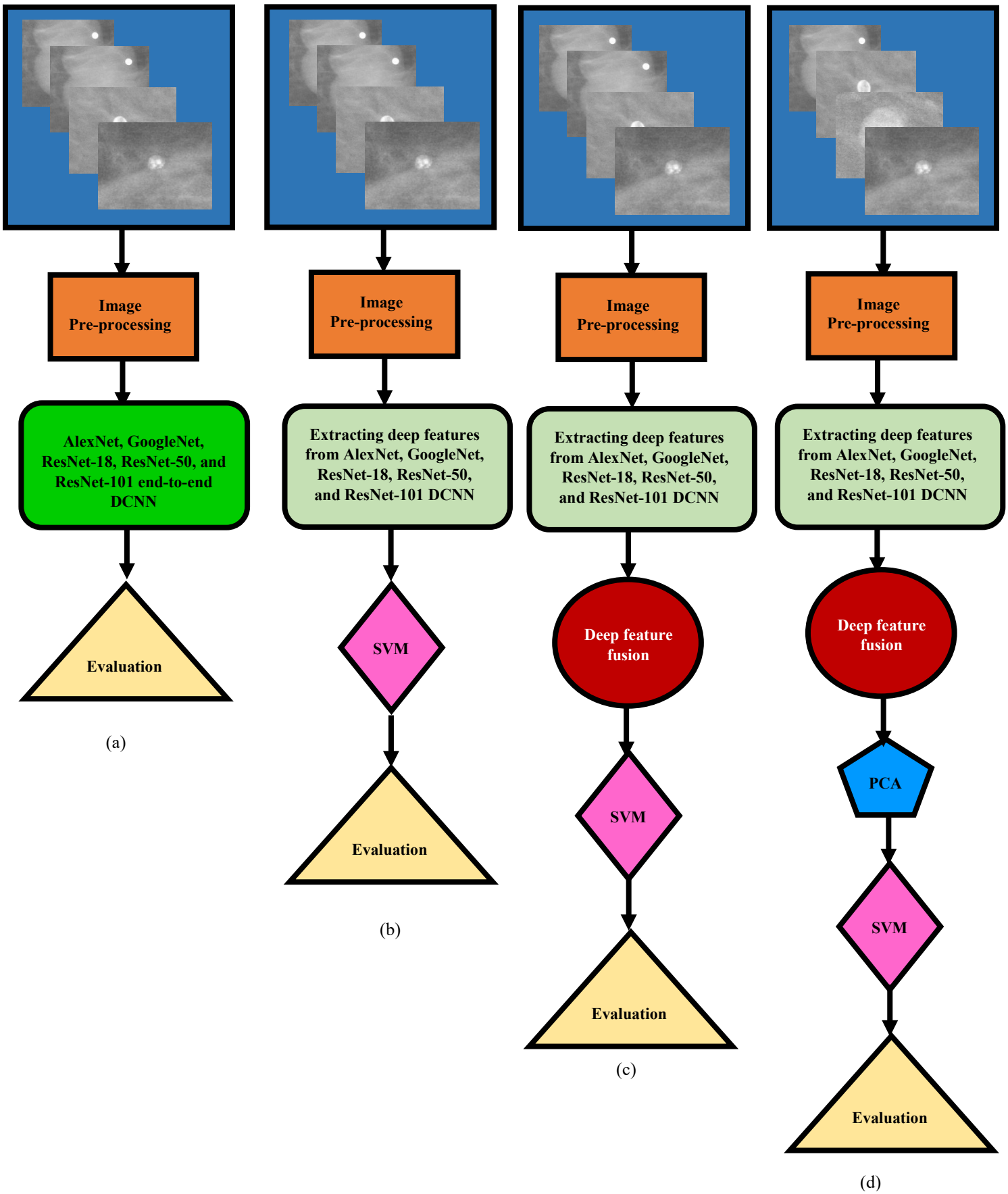


Figure 5.1: The proposed framework.

(a) Scenario (1), (b) scenario (2), (c) scenario (3), and (d) scenario (4).

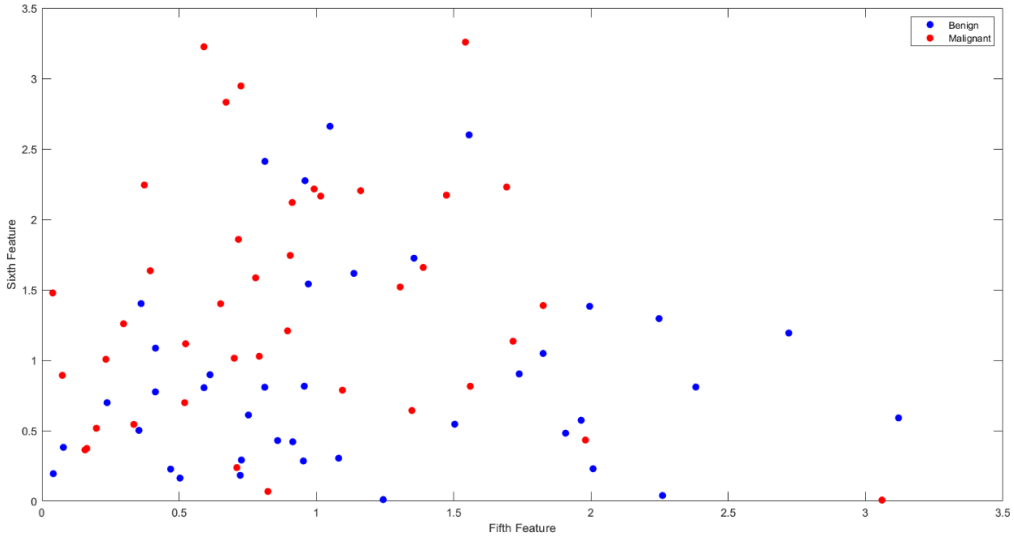


Figure 5.2: The fifth feature values versus the sixth feature values for the first 10 samples of the CBIS-DDSM images and their rotated versions for the ResNet-18 DCNN architecture features.

5.3 Experimental Setup

All the experiments were performed on the Intel® CORE™ I7 processor and NVIDIA GeForce 940MX, Windows 10, 64 bit with 8 GB of random-access memory (RAM). The software used to implement the experiments was MATLAB R2018b with an academic license provided by the University of Strathclyde. The experiments were applied on the CBIS-DDSM [86] and the MIAS [81] datasets to test the performance of the proposed framework. Moreover, the transfer learning technique is performed on all the DCNNs architectures to classify two classes instead of 1000.

5.3.1 Augmentation

Due to the lack of mammogram images, therefore, it was necessary to perform data augmentation techniques. The rotation method was applied to the training samples to increase the data. This was performed by rotating each original image by four angles (0, 90, 180, and 270 degrees). Therefore, each image was augmented to four images. The two-dimensional scatter plot based on the feature vectors for benign and malignant samples of the CBIS-DDSM breast cancer dataset is shown in Figure 5.2. This figure represents the fifth feature versus the sixth feature as an example for the features of ResNet-18 DCNN architecture for the first 10 samples of the CBIS-DDSM dataset images and their orientations with a total of 40 images for each class.

5.3.2 Parameter Setting

For training the five DCNNs, some parameters were adjusted while others were not altered from their default values as the adjustment of those parameters did not affect the DCNN performance. The hyper-parameters adjusted in the DCNNs were the minibatch size which was the number of samples included in each sub-epoch weight change and was chosen to be 10. This number was chosen to fit the memory of the GPU, as increasing it lead to an "out of memory" problem and the optimisation problem became numerically untractable. The learning rate defined as the stride size at every iteration whilst turning on the way to a minimum of a loss function. It was selected to be 10^{-5}, which achieved the highest accuracy while minimizing the training time. It is well-known that increasing the number of epochs increases the training time. The sum of epochs was modified to 10 as rising this number did not enhance the performance. The L2-regularization was set to 0.0005 to avoid the overfitting. These arrangements were to approve that the parameters were modified for classifying medical breast cancer mammogram images. The optimization algorithm SGDM was applied as well as in the framework presented in Chapter 4.

5.4 Results and Discussions

The proposed framework performs four scenarios. For the first scenario, seventy percent of the images were used for training and the rest for testing and validation, as this was the common ratio used in the classification problem. Although, for the rest of the scenario, the ratio of the training and testing was 80%:20%. This was because these scenarios were validated using five fold cross-validation. This means that the classifiers were trained with four folds and confirmed by the remaining fold. Consequently, the scenarios were taught five times and the testing classification accuracy was calculated for each time then averaged.

For the CBIS-DDSM dataset benign and malignant mass samples from the two mammogram views: CC and MLO were extracted and used in the four scenarios of the CAD system. Images of this dataset were already segmented and the breast cancer lesion was shown. Therefore, they do not need to be segmented. The samples were only enhanced using the CLAHE method. On the other hand, for the MIAS dataset, the number of samples for normal, benign, and malignant was not normalized. Therefore, we differentiated between only two classes, i.e. normal and abnormal, which is consistent with most existing works on the MIAS dataset as reported in [165], [229], [290], [303], [341], [342]. Therefore, the images were enhanced and the ROI was cropped using the co-ordinates of the centre and the radius of the

abnormality provided by the dataset. Table 5.1 shows the numbers of training, testing, and validation samples used for the CBIS-DDSM and MIAS datasets. In the following sub-sections, the results of the four scenarios will be presented and discussed for the CBIS-DDSM and MIAS datasets.

5.4.1 Scenario (1)

In this scenario, five end-to-end DCNNs including AlexNet, GoogleNet, ResNet-18, ResNet-50, and ResNet-101 were constructed. For the CBIS-DDSM dataset, the classification accuracy ranged from (71.09% – 76.01%) with the best accuracy achieved using the end-to-end GoogleNet as illustrated in Table 5.2. Moreover, the training time for the five networks varied between (6:30 – 62 hours) which is quite significant. On the other hand, the accuracy of DCNN networks in the case of the MIAS dataset ranged from (59.69% – 74.40%) with the highest accuracy achieved using GoogleNet architecture as well, as shown in Table 5.3. Table 5.2 and Table 5.3 show the classification accuracy for the different DCNN architecture to classify benign and malignant lesions and normal and abnormal lesions for the CBIS-DDSM and the MIAS datasets, respectively.

5.4.2 Scenario (2)

To improve the classification accuracy of DCNNs constructed in the first scenario, the deep features were extracted from each network. These deep features were used separately to train and test SVM classifiers with different kernel functions. The deep features obtained for each DCNN for the CBIS-DDSM dataset are visualised in Figures 5.3 – 5.7. In these figures, the first and second convolutional layers were visualised for AlexNet, GoogleNet, ResNet-18, ResNet-50, and ResNet-101, respectively. Moreover, the classification accuracies for the CBIS-DDSM dataset increased and ranged between 85.2% and 93.7% as obvious from Table 5.4. The scores obtained from the deep features of the ResNet-18 proved to be the highest compared to the other networks. Furthermore, when comparing the different SVM kernels constructed using ResNet-18 deep features the best accuracy was for the medium Gaussian kernel function. The accuracy was 93.7% and the AUC scored 0.98 (98%) as shown in Figure 5.8. Additionally, the sensitivity and specificity were 0.940 (94%) and 0.931 (93.1%), respectively. Table 5.4 shows the accuracy, AUC, sensitivity, and specificity of the SVM classifiers with different kernels constructed with the five deep feature sets. Additionally, Figure 5.8 displays the ROC curve and the AUC computed for the ResNet-18 deep features with medium Gaussian kernel function SVM.

Table 5.1: The total number of samples used for the CBIS-DDSM and MIAS datasets.

			Training	Testing	Validation	Total
CBIS-DDSM	**Benign**	2728	3690	791	791	5272
	Malignant	2544				
MIAS	**Normal**	836	900	194	194	1288
	Abnormal	452				

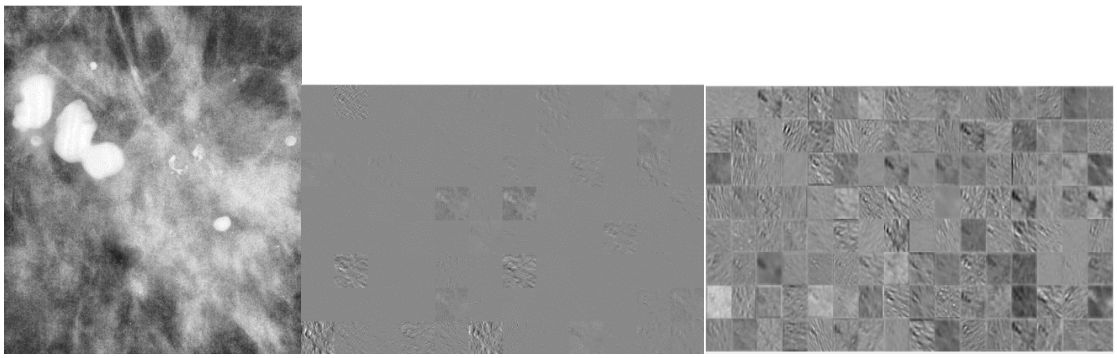
Table 5.2: The accuracy and the trained time of the end-to-end DCNN architectures for the CBIS-DDSM dataset.

DCNN Architectures	DCNN Accuracy	Training Time
AlexNet	74.68%	6 hours, 30 min
GoogleNet	**76.01%**	12 hours
ResNet-18	72.23%	14 hours
ResNet-50	71.09%	33 hours
ResNet-101	71.47%	62 hours

Table 5.3: The accuracy of the end-to-end DCNN architectures for the MIAS dataset.

DCNN Architectures	DCNN Accuracy
AlexNet	59.69%
GoogleNet	**74.40%**
ResNet-18	68.22%
ResNet-50	68.73%
ResNet-101	67.44%

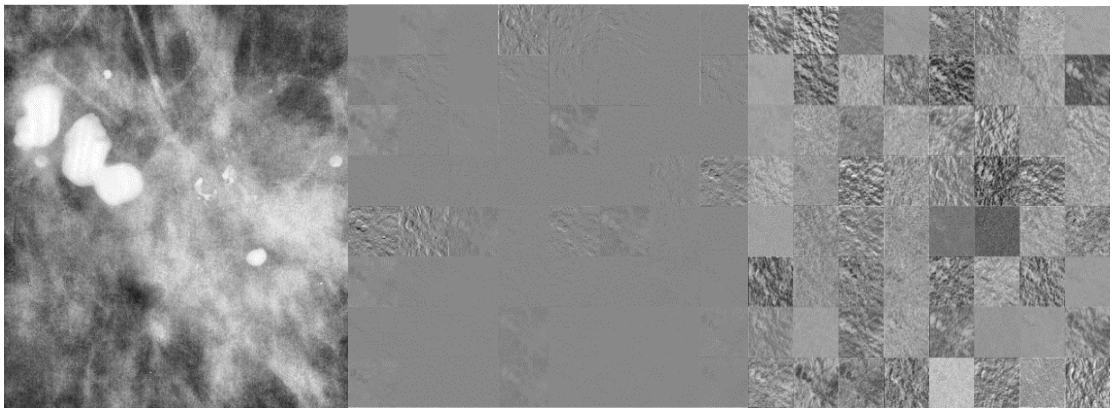
On the other hand, for the MIAS dataset, the accuracies of the SVM classifiers constructed using each deep feature of the DCNN have also increased to reach a range of (71% – 95.4%). This time the highest accuracy was achieved using the deep features of the ResNet-50 architecture. Furthermore, the quadratic kernel SVM constructed using these deep features ranked the first accuracy, which was 95.4% compared to the other kernels. The sensitivity and specificity of the quadratic SVM in this case, were 0.966 (96.6%) and 0.921 (92.1%), respectively. The classification accuracies of these models are illustrated in Table 5.5. Moreover, the AUC calculated from the ROC curve was 0.990 (99%) as shown in Figure 5.9.



(a) (b) (c)

Figure 5.3: Visualising the deep features for the fine-tuned AlexNet DCNN architecture.

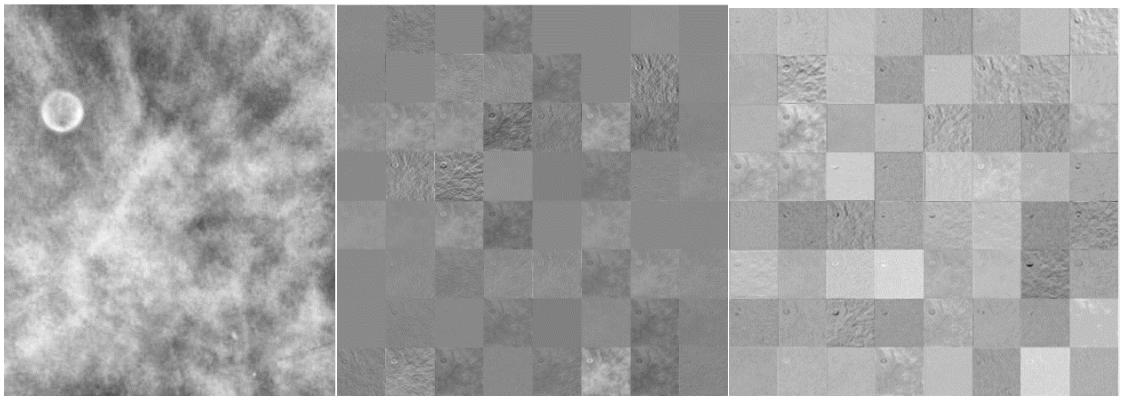
(a) Malignant ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.



(a) (b) (c)

Figure 5.4: Visualising the deep features for the fine-tuned GoogleNet DCNN architecture.

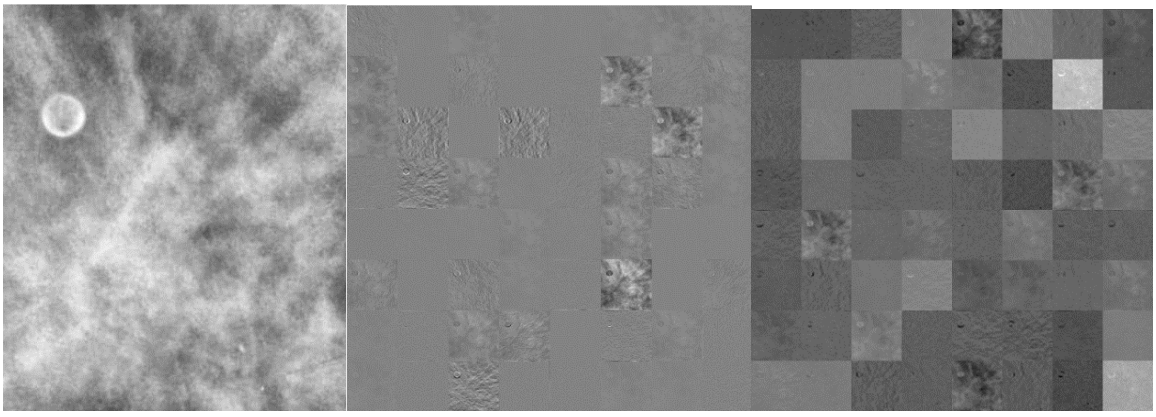
(a) Malignant ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.



(a) (b) (c)

Figure 5.5: Visualising the deep features for the fine-tuned ResNet-18 DCNN architecture.

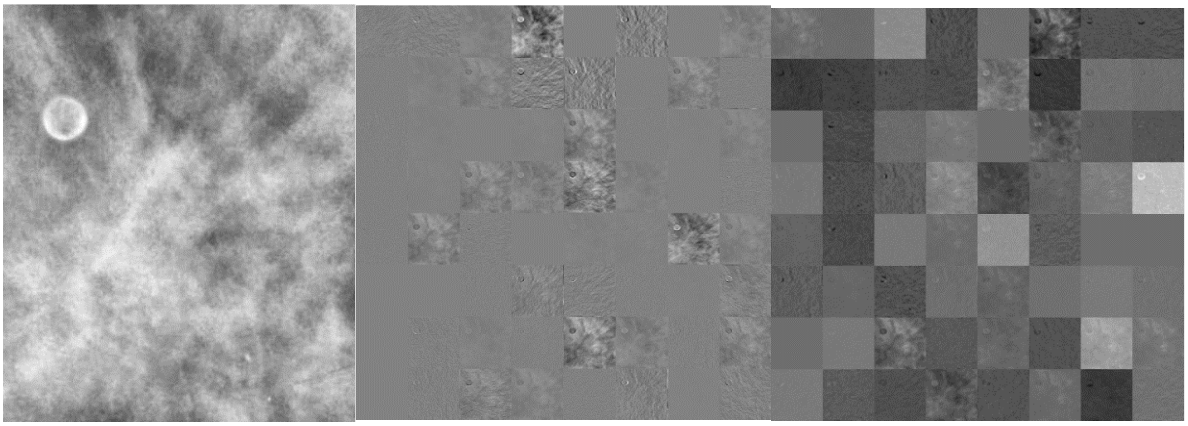
(a) Benign ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.



(a) (b) (c)

Figure 5.6: Visualising the deep features for the fine-tuned ResNet-50 DCNN architecture.

(a) Benign ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.



(a) (b) (c)

Figure 5.7: Visualising the deep features for the fine-tuned ResNet-101 DCNN architecture.

(a) Benign ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.

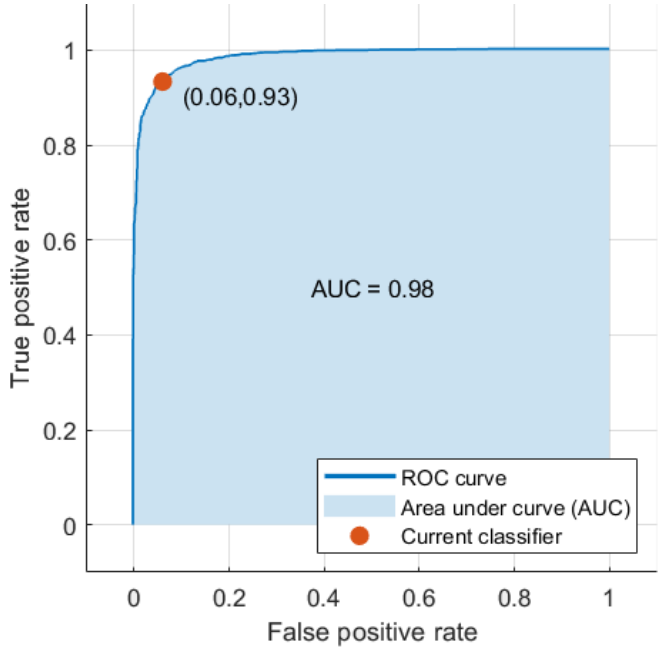


Figure 5.8: The computed ROC for the ResNet-18 with medium Gaussian kernel function SVM of the CBIS-DDSM dataset.

Table 5.4: The calculated scores of the DCNN deep features for the CBIS-DDSM dataset classified by the SVM classifier with different kernels.

DCNN	Different Kernels	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)
AlexNet	Linear	91.3% (0.001)	0.970 (0)	0.918 (0.005)	0.911 (0.001)
	Quadratic	91.0% (0.002)	0.960 (0.004)	0.909 (0.006)	0.911 (0.001)
	Cubic	90.9% (0.002)	0.960 (0.001)	0.904 (0.004)	0.913 (0.005)
	Medium Gaussian	91.1% (0.001)	0.970 (0)	0.910 (0.001)	0.910 (0.001)
	Coarse Gaussian	89.2% (0.001)	0.960 (0.001)	0.884 (0.003)	0.899 (0.001)
GoogleNet	Linear	90.1% (0.002)	0.970 (0)	0.900 (0.001)	0.900 (0.001)
	Quadratic	89.4% (0.003)	0.970 (0.004)	0.900 (0.001)	0.900 (0.001)
	Cubic	88.7% (0.004)	0.960 (0.004)	0.883 (0.004)	0.891 (0.004)
	Medium Gaussian	87.9% (0.002)	0.950 (0.004)	0.859 (0)	0.905 (0)
	Coarse Gaussian	88.6% (0.004)	0.950 (0.004)	0.898 (0.001)	0.876 (0.006)
ResNet-18	Linear	93.5% (0.002)	0.980 (0.001)	0.931 (0.001)	0.939 (0.004)
	Quadratic	93.1% (0.002)	0.980 (0.001)	0.930 (0.003)	0.931 (0.004)
	Cubic	93.0% (0.001)	0.980 (0.001)	0.930 (0.001)	0.930 (0.001)
	Medium Gaussian	**93.7%** (0)	**0.980** (0.013)	**0.940** (0.001)	**0.931** (0.003)
	Coarse Gaussian	93.4% (0.001)	0.980 (0.001)	0.932 (0.003)	0.940 (0.001)
ResNet-50	Linear	87.2% (0.002)	0.950 (0.005)	0.864 (0.003)	0.879 (0.001)
	Quadratic	88.4% (0.003)	0.950 (0)	0.880 (0.008)	0.887 (0.005)
	Cubic	87.8% (0.001)	0.950 (0)	0.873 (0.001)	0.884 (0.005)
	Medium Gaussian	87.3% (0.01)	0.950 (0)	0.861 (0.006)	0.894 (0.004)
	Coarse Gaussian	85.2% (0.005)	0.930 (0.004)	0.833 (0.006)	0.874 (0.001)
ResNet-101	Linear	89.5% (0.001)	0.96 (0.001)	0.878 (0.021)	0.906 (0.007)
	Quadratic	89.3% (0.002)	0.96 (0.001)	0.893 (0.006)	0.900 (0.006)
	Cubic	89.1% (0.001)	0.95 (0.004)	0.89 (0.003)	0.893 (0.005)
	Medium Gaussian	89.3% (0.002)	0.96 (0.001)	0.877 (0.006)	0.908 (0.001)
	Coarse Gaussian	87.9% (0.004)	0.94 (0.004)	0.853 (0.006)	0.905 (0.001)

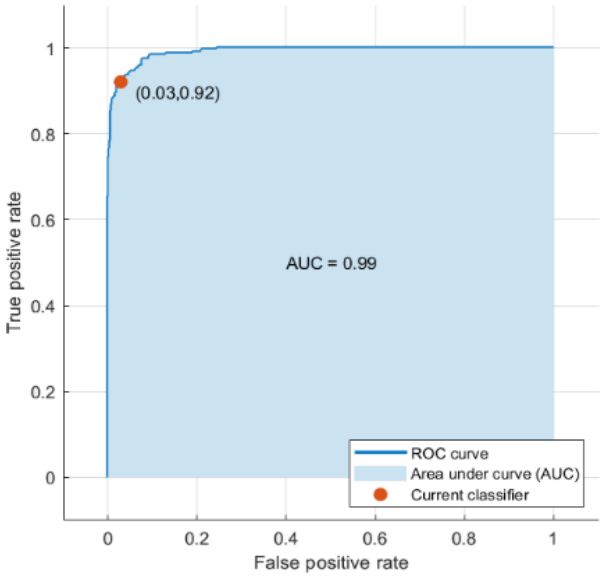


Figure 5.9: The computed ROC for the ResNet-50 with quadratic kernel SVM function of the MIAS dataset.

Table 5.5: The calculated scores of the different SVM kernel functions for the DCNN deep features for the MIAS dataset.

DCNN	Different Kernels	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)
AlexNet	Linear	79.7% (0.002)	0.860 (0.001)	0.871 (0.007)	0.684 (0.003)
	Quadratic	80.8% (0.007)	0.880 (0.004)	0.841 (0.04)	0.723 (0.016)
	Cubic	80.0% (0.002)	0.860 (0.001)	0.84 (0.004)	0.725 (0.006)
	Medium Gaussian	78.8% (0.007)	0.860 (0.006)	0.884 (0.006)	0.662 (0.007)
	Coarse Gaussian	72.7% (0.004)	0.820 (0.001)	0.902 (0.003)	0.573 (0.002)
GoogleNet	Linear	76.6% (0.007)	0.830 (0.004)	0.841 (0.011)	0.65 (0.005)
	Quadratic	76.4% (0.007)	0.830 (0.004)	0.802 (0.044)	0.667 (0.009)
	Cubic	77.2% (0.007)	0.840 (0.005)	0.817 (0.011)	0.688 (0.002)
	Medium Gaussian	77.3% (0.004)	0.830 (0.007)	0.888 (0.002)	0.641 (0.003)
	Coarse Gaussian	71% (0.003)	0.790 (0)	0.916 (0.004)	0.557 (0.003)
ResNet-18	Linear	87.8% (0.002)	0.940 (0.001)	0.937 (0.004)	0.792 (0.001)
	Quadratic	89.1% (0.002)	0.940 (0.001)	0.921 (0.037)	0.821 (0.011)
	Cubic	87.9% (0.003)	0.940 (0.004)	0.921 (0.005)	0.810 (0.001)
	Medium Gaussian	85.3% (0.001)	0.930 (0.004)	0.944 (0.001)	0.740 (0.002)
	Coarse Gaussian	73.7% (0.002)	0.910 (0.004)	0.964 (0.001)	0.574 (0.002)
ResNet-50	Linear	94.4% (0.004)	0.990 (0)	0.968 (0.004)	0.894 (0.007)
	Quadratic	**95.4%** (0.001)	**0.990** (0)	**0.966** (0.015)	**0.921** (0.011)
	Cubic	94.6% (0.002)	0.990 (0)	0.971 (0.005)	0.908 (0.001)
	Medium Gaussian	92% (0.001)	0.980 (0.001)	0.976 (0.001)	0.838 (0.001)
	Coarse Gaussian	79.1% (0.002)	0.950 (0.004)	0.977 (0.01)	0.627 (0.002)
ResNet-101	Linear	93.1% (0.003)	0.980 (0.001)	0.976 (0.004)	0.860 (0.001)
	Quadratic	93.6% (0.002)	0.980 (0.001)	0.961 (0.026)	0.887 (0.01)
	Cubic	93.6% (0.003)	0.980 (0.001)	0.966 (0.004)	0.882 (0.003)
	Medium Gaussian	91% (0.004)	0.970 (0.004)	0.974 (0.004)	0.817 (0.003)
	Coarse Gaussian	79.1% (0.001)	0.960 (0.001)	0.977 (0.001)	0.627 (0.001)

5.4.3 Scenario (3)

This scenario was conducted to determine if combining deep features would enhance the performance of the SVM classifiers; consequently, four feature sets representing different combination deep features were produced. To generate these sets, the classification accuracies produced in scenario (1) were used as a ranking method to order the deep features extracted from each DCNN in descending order. Subsequently, this ranking was employed in sequential forward feature set selection to identify the best combination of deep features.

For the CBIS-DDSM dataset, it was clear from Table 5.2 that the GoogleNet and AlexNet features achieved the highest accuracies compared to the others. Therefore, these features were combined to produce a single feature vector with 5120 features named feature set (1). The linear kernel SVM achieved the highest accuracy, which was 94.4%. Moreover, the ResNet-18 features were added to the feature set (1). This set was named feature set (2) containing 5632 features. The highest accuracy achieved was 96.9% for the linear kernel SVM as well. Additionally, the features of ResNet-101 were added to feature set (2) producing a feature

vector of 7680 features in length named feature set (3). The accuracy increased to 97.5% for the linear kernel SVM. Finally, all the deep features were combined to produce a feature vector with 9728 features in length named feature set (4). Figure 5.10 shows a comparison for the accuracies of SVM classifiers of different kernels for the four feature sets. This figure reveals that increasing the number of deep features will increase the classification accuracy of the SVM classifiers. It was obvious that feature set (4) had improved the accuracy to reach 97.9% using the quadratic and cubic kernels. This was higher than the 93.7% of the linear SVM classifier constructed using only the deep features of ResNet-18 in scenario (2). Thus, Table 5.6 shows the accuracy, AUC, sensitivity, and specificity for the SVM classifiers of different kernels for feature set (4). Table 5.6 also indicated that the sensitivity and specificity were both equal to 0.98 (98%), which were higher than those achieved in the scenario (2). Additionally, the AUC of the cubic and quadratic SVM classifiers achieved the highest accuracy yielding to 1.00 (100%) as shown in the ROC curve in Figure 5.11.

Conversely, for the MIAS dataset, four sets of different combinations of deep features were generated in the same manner as those produced in the CBIS-DDSM. Feature set (1) represents those extracted from the GoogleNet and ResNet-50 as these DCNNs achieved the highest accuracies compared to the others as in Table 5.3. The feature-length of feature set (1) was 3072 features. The highest accuracy achieved was 95% using the quadratic kernel SVM. On the other hand, feature set (2) consists of feature set (1) plus ResNet-18 producing 3584 feature-length. The accuracy was 96.3% achieved using the quadratic kernel SVM as well. The features of the ResNet-101 were added to the feature set (2) and named feature set (3). The highest accuracy achieved was 97.4% for the quadratic kernel SVM. Finally, adding the AlexNet features to feature set (3) produces feature set (4). This time the accuracy decreased to 96.6%. Figure 5.12 shows the accuracies of the SVM classifiers constructed using different kernels for the four sets of features. Figure 5.12 reveals that feature set (3) had the highest classification accuracy 97.4% using the quadratic kernel SVM classifier. This was higher than the 95.4% accuracy achieved using the deep features of ResNet-50 only in the second scenario. Hence, Table 5.7 shows different scores for the feature set (3) which achieved the highest accuracy using the quadratic SVM classifier. Furthermore, the sensitivity and the specificity increased to 0.987 (98.7%) and 0.949 (94.9%), respectively, which were higher than 0.966 (96.6%) and 0.921 (92.1%), achieved in the second scenario. Additionally, the AUC increased from 0.99 (99%) to 1.00 (100%) as clear in Figure 5.13. The ROC curve and the AUC for the quadratic kernel SVM function of feature set (3) are shown in Figure 5.13.

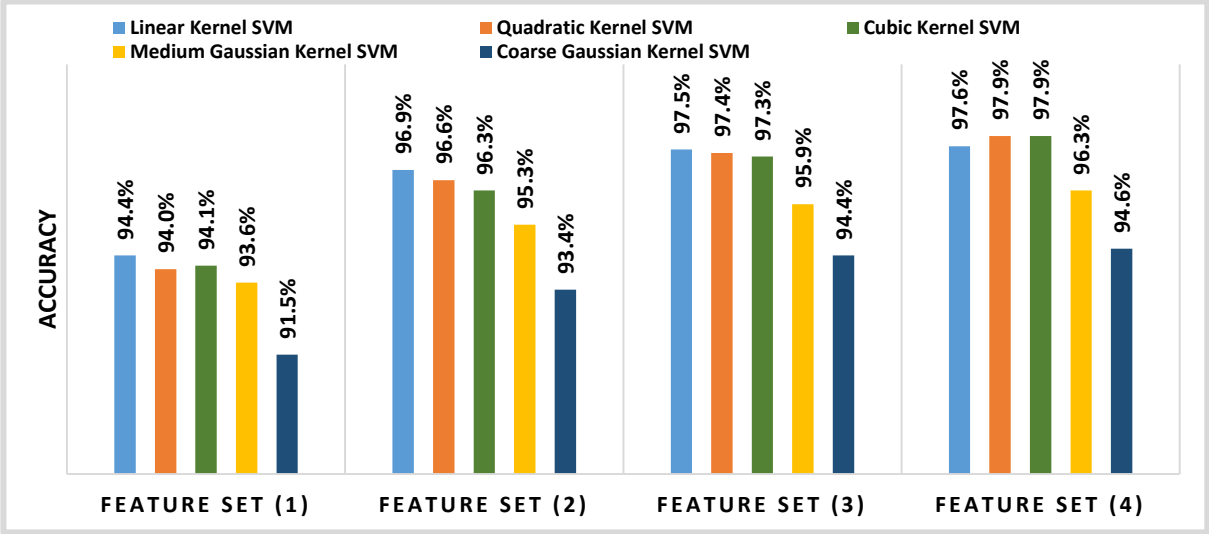


Figure 5.10: The accuracies of different SVM kernel functions for the different deep features combination of the CBIS-DDSM dataset.

Feature Set (1) = GoogleNet and AlexNet.
Feature Set (2) = GoogleNet, AlexNet, and ResNet-18.
Feature Set (3) = GoogleNet, AlexNet, ResNet-18, and ResNet-101.
Feature Set (4) = GoogleNet, AlexNet, ResNet-18, ResNet-101, and ResNet-50.

Table 5.6: The calculated scores of the different SVM kernel functions for feature set (4) for the CBIS-DDSM dataset.

GoogleNet, AlexNet, ResNet-18, ResNet-101, and ResNet-50 DCNN Features				
	Accuracy (Std)	**AUC (Std)**	**Sensitivity (Std)**	**Specificity (Std)**
Linear	97.6% (0.001)	0.999 (0)	0.971 (0.001)	0.980 (0.001)
Quadratic	**97.9%** (0.001)	**1.000*** (0)	**0.980** (0.003)	**0.980** (0.001)
Cubic	**97.9%** (0.001)	**1.000*** (0)	**0.980** (0.001)	**0.980** (0.001)
Medium Gaussian	96.3% (0.001)	0.990 (0)	0.951 (0)	0.970 (0)
Coarse Gaussian	94.6% (0)	0.990 (0)	0.950 (0)	0.950 (0)
* rounded values				

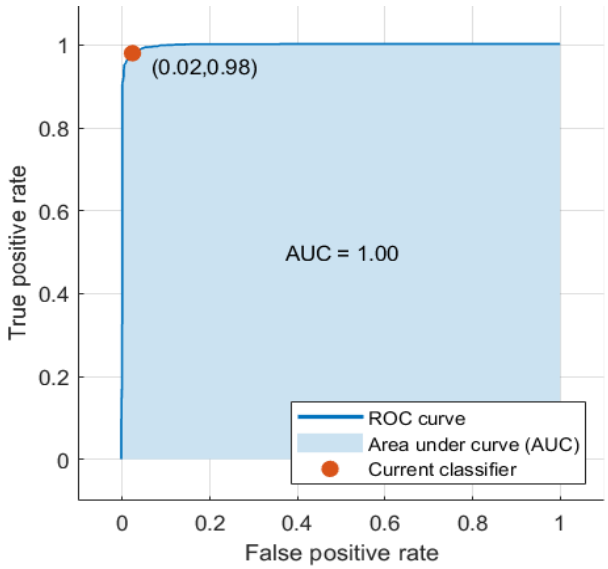


Figure 5.11: The computed ROC for all combined DCNN features for the cubic and quadratic SVM kernel functions of the CBIS-DDSM dataset.

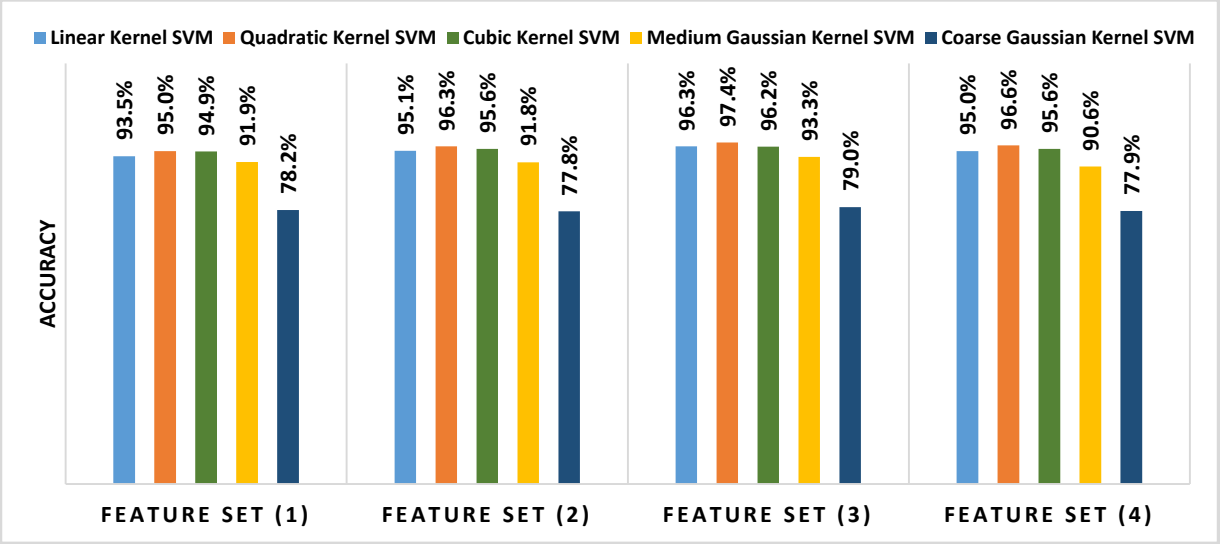


Figure 5.12: The accuracies of different SVM kernel functions for the different deep features combination of the MIAS dataset.

Feature Set (1) = GoogleNet and ResNet-50.
Feature Set (2) = GoogleNet, ResNet-50, and, ResNet-18.
Feature Set (3) = GoogleNet, ResNet-50, ResNet-18, and ResNet-101.
Feature Set (4) = GoogleNet, ResNet-50, ResNet-18, and ResNet-101, and AlexNet.

Table 5.7: The calculated scores of the different SVM kernel functions for feature set (3) of the MIAS dataset.

GoogleNet, ResNet-50, ResNet-18, and ResNet-101 DCNN Features				
	Accuracy (Std)	**AUC (Std)**	**Sensitivity (Std)**	**Specificity (Std)**
Linear	96.3% (0.001)	0.990 (0)	0.990 (0)	0.917 (0)
Quadratic	**97.4%** (0)	**1.000*** (0)	**0.987** (0.013)	**0.949** (0.012)
Cubic	96.2% (0)	0.999 (0)	0.990 (0.001)	0.926 (0.001)
Medium Gaussian	93.3% (0)	0.990 (0)	0.989 (0.001)	0.854 (0.001)
Coarse Gaussian	79% (0)	0.960 (0.001)	1.000 (0.001)	0.629 (0.001)
* rounded values				

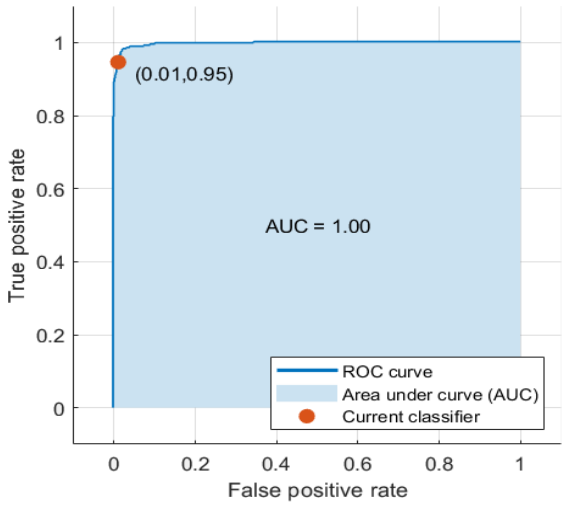


Figure 5.13: The computed ROC for the feature set three for the fusion of deep features for the quadratic SVM kernel function of the MIAS dataset.

5.4.4 Scenario (4)

PCA was applied in this scenario to reduce the large feature space of the feature sets generated in the third scenario. In order to select the number of principal components that achieve the highest accuracy, a sequential forward strategy was used. It started with 50 principal components and added the components iteratively.

For the CBIS-DDSM dataset, when reducing the features of sets (1) and (2) the accuracies reached 93.6% and 96.5% with 300 principal components only. However, for feature sets (3) and (4), the accuracies became 97.4% and 97.8% with 400 principal components only. Therefore, the highest classification accuracy in this scenario was achieved using feature set (4) as shown in Figure 5.14. This figure represents the classification accuracy, using the PCA reduction for the four feature sets. Moreover, the execution time for training decreased from 287.48 s to 57.85 s.

Whereas for the MIAS dataset, when the PCA method was applied to feature sets (1) and (2) the accuracies yielded to 94.7% and 95.3%, respectively. This was achieved using 150 principal components. Furthermore, the accuracies of feature sets (3) and (4) became 96.8% and 95.2% with 200 principal components. Thus, the highest accuracy achieved for the MIAS dataset in this scenario was for feature set (3) as shown in Figure 5.15. Furthermore, the operating time decreased from 40.77 s to 1.97 s when applying PCA on feature set (3).

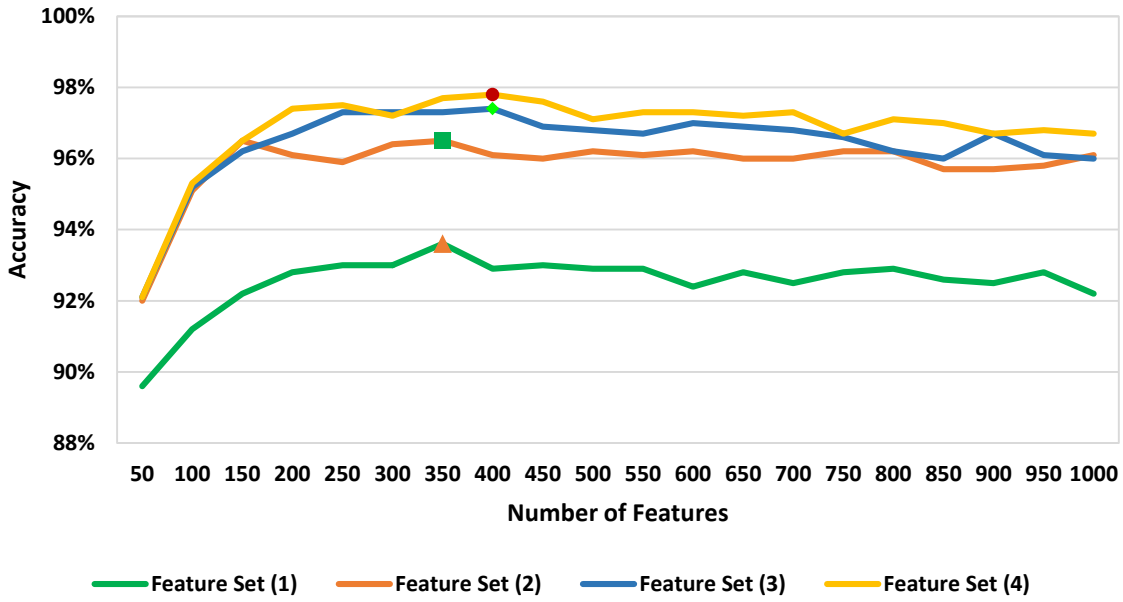


Figure 5.14: A comparison of the classification accuracy of different combined deep features after PCA feature reduction for the CBIS-DDSM dataset.

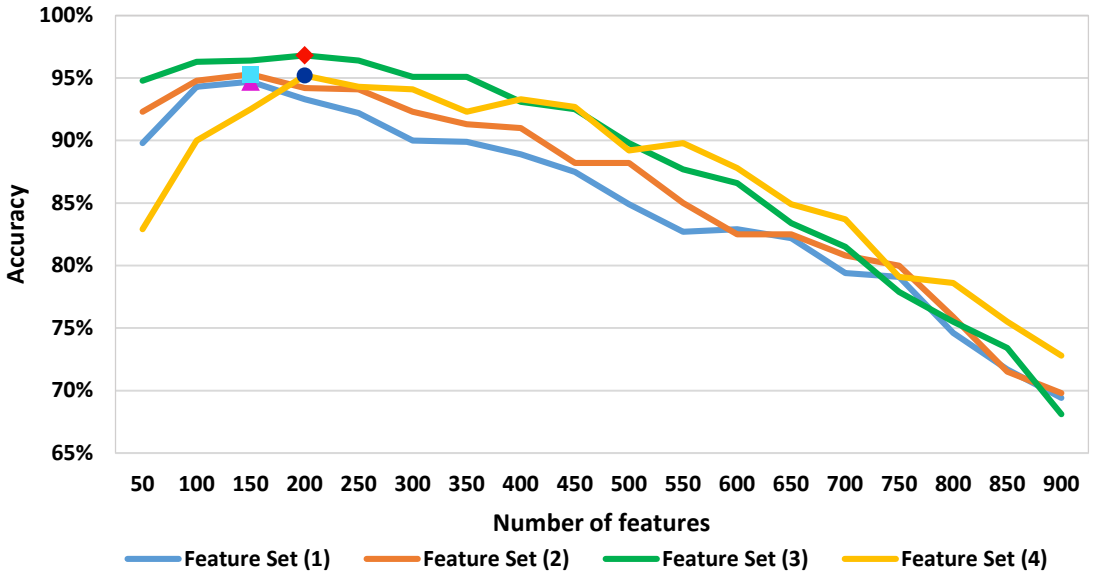


Figure 5.15: A comparison of the classification accuracy of different combined deep features after PCA feature reduction for the MIAS dataset.

To validate the statistical significance of the results obtained in all scenarios, an ANOVA test was performed on all the results by a five fold cross-validation repetition method. The null hypothesis Ho for all classification was that the mean accuracies of all SVM kernel functions classifiers were the same. Tables 5.8 – 5.12 and Tables 5.13 – 5.17 show the ANOVA test for the deep features of the five DCNN architectures constructed in scenario (2) for the CBIS-DDSM and MIAS datasets, respectively. For scenario (3), since the highest accuracy was achieved using feature set (4) and feature set (3) for the CBIS-DDSM and MIAS datasets, respectively, therefore, the ANOVA test was computed to these sets. Table 5.18 and Table 5.19 show the ANOVA test for the feature set (4) and feature set (3) performed in the scenario (3) for CBIS-DDSM and MIAS datasets, respectively. From these tables, it was revealed that the p-values achieved were lower than α, where $\alpha = 0.05$. Consequently, it can be concluded that there was a statistically significant difference between the accuracies of the classifiers.

Table 5.8: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of AlexNet DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00298	4	0.00074	532.99	< 0.001
Error	0.00006	45	0		
Total	0.00304	49			

Table 5.9: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of GoogleNet DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00267	4	0.00067	100.86	< 0.001
Error	0.0003	45	0.00001		
Total	0.00296	49			

Table 5.10: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-18 DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00037	4	9.33×10^{-5}	68.29	< 0.001
Error	0.00006	45	1.3667×10^{-6}		
Total	0.00043	49			

Table 5.11: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-50 DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00605	4	0.00151	60.45	< 0.001
Error	0.00113	45	0.00003		
Total	0.00717	49			

Table 5.12: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-101 DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00173	4	0.00043	120.51	< 0.001
Error	0.00016	45	0		
Total	0.00189	49			

Table 5.13: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of AlexNet DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.04284	4	0.01071	509.95	< 0.001
Error	0.00095	45	0.00002		
Total	0.04379	49			

Table 5.14: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of GoogleNet DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.02842	4	0.00711	241.35	< 0.001
Error	0.00132	45	0.00003		
Total	0.02975	49			

Table 5.15: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-18 DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.15896	4	0.03974	11332.57	< 0.001
Error	0.00016	45	0		
Total	0.15912	49			

Table 5.16: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-50 DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.18644	4	0.04661	16208.77	< 0.001
Error	0.00013	45	0		
Total	0.18657	49			

Table 5.17: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-101 DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.15574	4	0.03894	8375.22	< 0.001
Error	0.00021	45	0		
Total	0.15595	49			

Table 5.18: The ANOVA test details for the different kernel functions of the SVM classifier for the feature set (4) in scenario (3) for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00795	4	0.00199	99391	< 0.001
Error	0	45	0		
Total	0.00795	49			

Table 5.19: The ANOVA test details for the different kernel functions of the SVM classifier for the feature set (3) in scenario (3) for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.23117	4	0.05779	9.3699×10^{16}	< 0.001
Error	0	45	0		
Total	0.23117	49			

Finally, the proposed framework has been compared with the applicable state-of-the-art frameworks presented in Chapter 3 to prove the efficiency of the proposed method. The results reveal that the proposed framework has outperformed others. Regarding the CBIS-DDSM dataset, the results have shown that the proposed framework recorded slightly high classification accuracy and AUC compared to Khan et. al [330]. This was obvious as Khan et. al [330] achieved 96.6% for accuracy and 0.934 (93.4%) for AUC. However, Khan et al. [330] fused the deep features of VGG-16, VGG-19, GoogleNet, and ResNet-50 DCNN. Moreover, it was found that the accuracy increased compared to the first framework proposed in Chapter 4 as well [33]. Moreover, lately, in 2020, Zhang et al. [332] fused some handcrafted features with deep features and classify them using several classifiers. However, the accuracy achieved was lower than the one achieved by feature set (4) generated in the fourth scenario by 6.99%. Whereas for the MIAS dataset, the accuracy achieved was higher than that in Hepsag et al. [342] and Tan et al. [317] as well. This was clear as the highest classification accuracy and AUC achieved were 97.4% and 0.99 (99%).

On the other hand, when comparing the usage of different DCNN architectures, it was obvious that the scores achieved by the proposed scenarios were high as well. In 2017, Jiang et al. [163] and Zhang et al. [282] achieved an AUC of 0.83 (83%) and 0.8 (80%), respectively using the AlexNet DCNN. However, they evaluated the approaches on different datasets BCDR-03 and DDSM, respectively. Moreover, Jiang et al. [163] used the GoogleNet DCNN to achieve a better AUC compared to the one achieved using the AlexNet although, it was still

lower than that of the proposed framework. Moreover, in 2020, Song et al. [331] extracted and classified the deep features of GoogleNet and Inception-v2 using the XGBoost classifier achieving an accuracy of 92.8%. Besides, there was a very small slight difference between the accuracy achieved in this proposed work with the work offered by Khan et al. [257]. They fused the deep features of VGGNet, GoogleNet, and ResNet DCNN. However, Khan et al. [257] applied their experiments on microscopic samples.

5.5 Summary

In this chapter, various fusion methods were presented by executing four scenarios. The first scenario composed of constructing five end-to-end pre-trained fine-tuned DCNN networks of different architectures. In order to increase and enhance the classification accuracy of scenario (1), the second scenario was constructed by extracting the deep features of the DCNNs constructed in scenario (1). These deep features were used separately to design SVM classifiers with different kernel functions. The results showed that the classification accuracies in scenario (2) were higher than those of scenario (1). On the other hand, the third scenario was devised to test if combining deep features would enhance the accuracy of the SVM classifiers. In this scenario, four feature sets were generated by ranking the accuracies of the deep features achieved in scenario (1) in descending order. These feature sets were used to train the SVM classifiers. The results showed that combining more deep features increased the performance of the SVM classifiers in both datasets. This demonstrated that the feature fusion using feature set (4) for the CBIS-DDSM dataset and feature set (3) for the MIAS dataset had improved the accuracy to reach 97.9% and 97.4%, respectively. Finally, in scenario (4), PCA was used to reduce the large dimension of the feature space produced in scenario (3). The results showed that PCA had reduced the feature space to 400 and 200 principal components for the CBIS-DDSM and the MIAS datasets, respectively. In addition, the classification accuracies were the same as in scenario (3). However, the computational cost decreased when applying PCA to the feature fusion. This was clear as the execution time for the classification process reduced from 287.48 s to 57.85 s and from 40.77 s to 1.9794 s for the CBIS-DDSM and MIAS datasets, respectively.

Chapter 6

A Breast Cancer Classification Framework Based on the Decision-Levels Fusion using Multiple DCNNs

6.1 Introduction

This chapter presents a novel framework to investigate different fusion techniques based on decision-level fusion rather than feature fusion. This framework is built-up using six scenarios to help radiologists to classify breast cancer lesions in mammograms. The first one is creating end-to-end pre-trained fine-tuned DCNN networks. In the second one, the deep features of the DCNNs are extracted and fed to an SVM classifier with different kernel functions. The third experiment is performed by deep features fusion to prove that combining deep features would enhance the accuracy of the SVM classifiers. The fourth scenario is extracting the outcomes of the samples instead of the features from the DCNN architectures of the first scenario and fed them separately to individual and MCS classifiers. In the fifth scenario, the outcomes of each DCNN are fused and fed into individual and MCS as well. Finally, the sixth scenario is constructed by fusing the outcomes of each individual and multiple classifiers for all DCNN architectures. All the scenarios are performed on a new mammogram dataset namely the "DAR-Breast" dataset collected from the Armed Forces Hospital located in Alexandria, Egypt.

6.2 The DAR-Breast Dataset

Patients' mammogram images were collected from the Armed Forces Hospital located in Alexandria, Egypt named the DAR-Breast dataset. To the best of our knowledge, this is the first time mammogram breast cancer scans are acquired and analysed for the purpose of breast cancer diagnosis in Egypt in addition to the Middle East. DAR-Breast consists of seventy-six cases, including 268 images being categorised as normal, benign, malignant masses, suspicious calcifications, and architectural distortion samples as illustrated in Figure 6.1. All the samples were in the DICOM format and labelled by expert radiologists from the hospital. As the number of samples is not normalised, we grouped the benign, malignant masses, suspicious calcifications, and architectural distortion samples as abnormal samples. Thus, we differentiated between only two classes, i.e. normal and abnormal. Figure 6.2 shows a normal

and abnormal mammogram sample extracted from the dataset. The mammograms were captured using a Siemens MAMMOMAT Inspiration device with a detector of amorphous selenium.

The MAMMOMAT Inspiration in Figure 6.3 is provided with progressive reconstruction intelligently minimizing exposure, (prime) technology. This technology allows the dose to be reduced by up to 30% without impairing image quality. Combining the multiple dose-saving features, MAMMOMAT Inspiration is the reference for a low dose [343]. Additionally, the MAMMOMAT Inspiration device increases the diagnostic certainty with high definition breast tomosynthesis, with the widest angle of 50° and unique EMPIRE Technology. It obtains highly defined tissue and lesion morphology in unprecedented detail and gains deeper insights in both 2D and 3D – leading to more accurate and earlier detection [343].

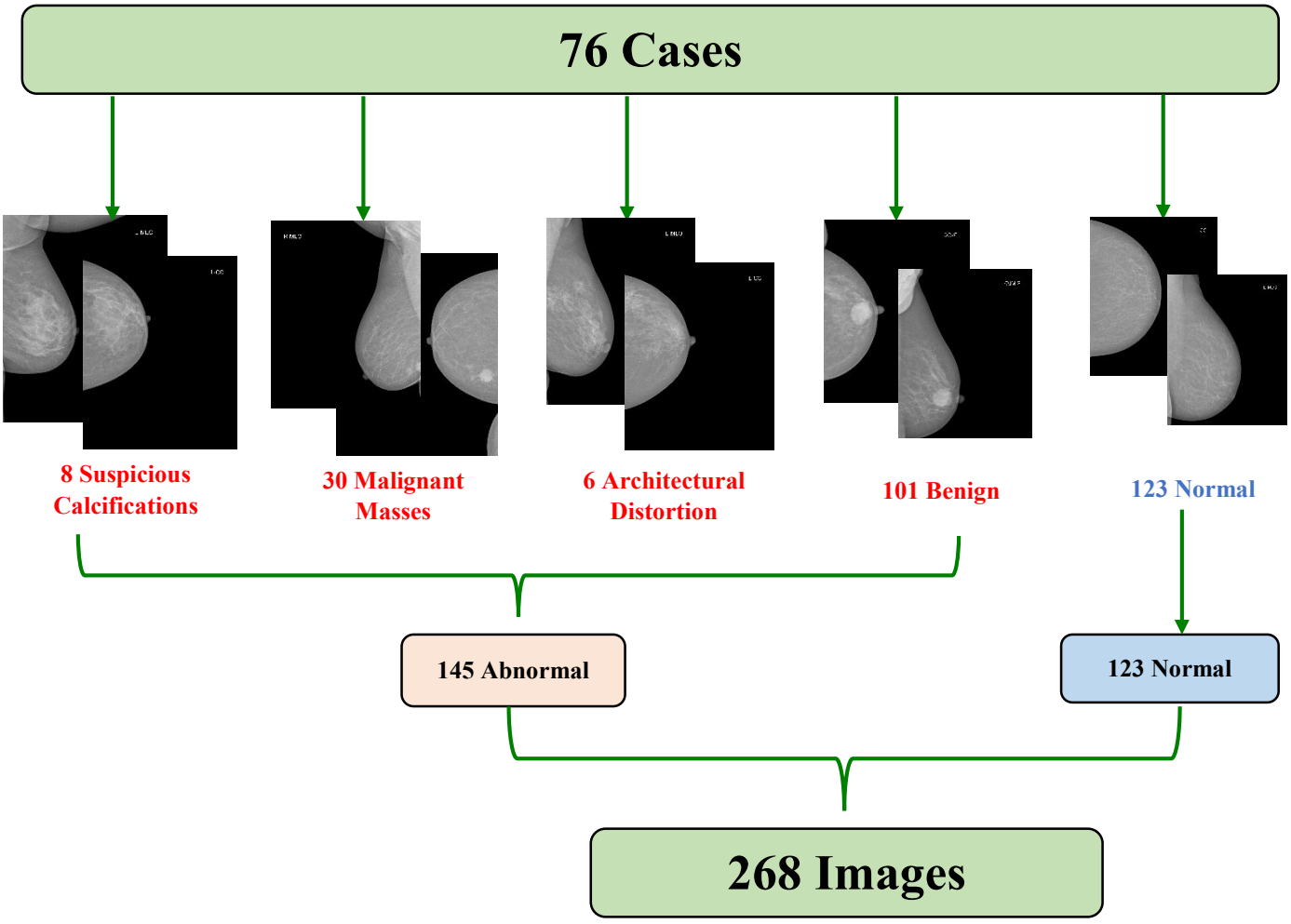


Figure 6.1: The distribution of the number of cases from the DAR-Breast dataset.

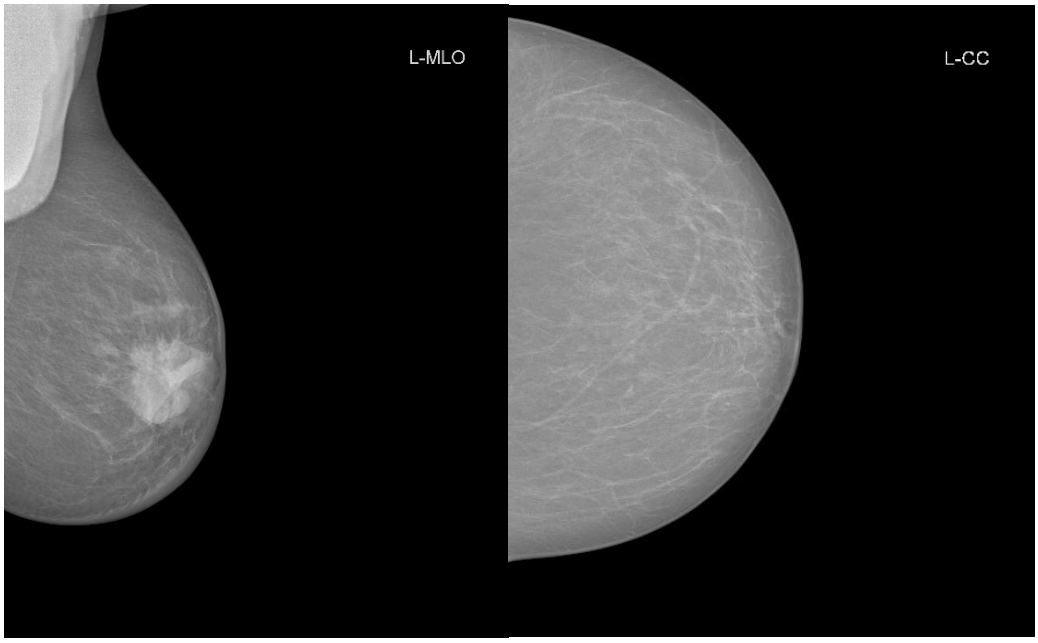


Figure 6.2: A mammogram sample extracted from the DAR-Breast dataset, an abnormal (left) and a normal (right) samples.

Figure 6.3: The MAMMOMAT Inspiration mammogram device located in the Armed Forces Hospital. Captured from Armed Forces Hospital.

6.3 Methodology

Firstly, the ROI is extracted by suppressing the breast from the images and removing any artefacts. This is performed using the following steps;

1) Orient all the mammogram samples in the same direction to avoid the situation of applying different methods for the left and the right-oriented MLO mammograms. Therefore, for example, flip all the RMLO view to look like the LMLO view samples.

2) Eliminate the mammogram image from any radiopaque artefacts such as labels by using thresholding and morphological operations [344]. A global threshold with a value of 18 was found to be the most suitable threshold for transforming the greyscale images into binary [0, 1] format [344]. Figure 6.4 shows the mammogram image sample with artefact suppression.

The proposed framework as shown in Figure 6.5 performs six scenarios:-

1) An end-to-end DCNN process: The pre-trained architectures; AlexNet, GoogleNet, Inception-v3, Inception ResNet-v2, MobileNet-v2, ResNet-18, ResNet-50, and ResNet-101 are used.

2) Deep features extraction: Classify the deep features individually using several SVM classifiers with different kernel functions.

3) Deep features fusion.

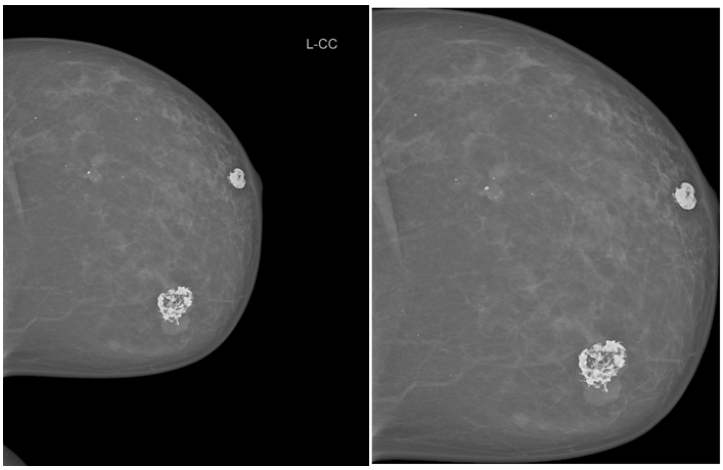
4) Decision-levels extraction: The decisions of the samples from scenario (1) are extracted and classified using individual and MCS classifiers.

5) Decision-levels fusion: Fusing and classifying the decision-level in scenario (4) by the same individual and MCS classifiers.

6) A multi-stage classifier: The decision-levels for each individual and MCS classifiers constructed in scenario (4) are fused individually and classified again using the same individual and MCS classifiers.

As it is obvious that the first three scenarios are constructed in the same manner as the ones generated in Chapter 5. This was performed to validate the classification accuracy obtained on the two publicity datasets the CBIS-DDSM [86] and MIAS [81] in Chapter 5 with the DAR-Breast dataset. On the other hand, for scenarios (4) – (6) novel fusion techniques are proposed. The individual classifier used in these scenarios is the DT with different trees such as J48 DT, RF DT, and RT DT.

The multiple classifier systems (MCS) are a hybrid method that fuses the classification results of several classifiers connected by a combiner. The MCS adds the strength of each classifier that usually exceeds the performance of each classifier. This is corresponding to medical applications in cases where the diagnosis of a specific illness is made by taking decisions from various doctors to come to a more confident final decision [345]. The MCS can avoid the possibility of poor results that are generated from a certain unsuitably selected model. The MCS has two structures: Cascaded and Parallel. In the cascaded structure, several classifiers are connected in series such as the AdaBoosting ensemble [345]. Each classifier in the ensemble attempts to improve the performance of the previous weaker classifier. It uses a class weighting resampling technique to train the next classifier in the ensemble. Instances that are not correctly classified with the first classifier in the ensemble are given higher weights and then these resampled instances enter the next classifier [345]. This procedure is repeated until all classifiers of the ensemble are processed. On the other hand, in the parallel structure, some classifiers are connected in parallel and their predictions are fused using either majority voting, maximum probability, minimum probability, or averaging methods. These classifiers may be of the same or different types, such as the bagging ensemble [345]. The bagging ensemble stands for bootstrap aggregation. It depends on the bootstrap resampling method to generate several data subsets from the original data randomly. These subsets are used to build several classifiers of the same type, such as decision trees [345]. Hence, the MCS used in scenarios (4) – (6) is constructed using three different structures. The first structure includes AdaBoosting with J48 DT, RF DT, and RT DT. The second structure is bagging with J48 DT, RF DT, and RT DT. In addition, the third structure is an MCS constructed with J48 DT, RF DT, RT DT, and combined using averaging fusing technique.



(a) (b)

Figure 6.4: An example of an abnormal sample and its ROI; (a) the original abnormal sample extracted from DAR-Breast dataset and (b) suppressed image from artefacts.

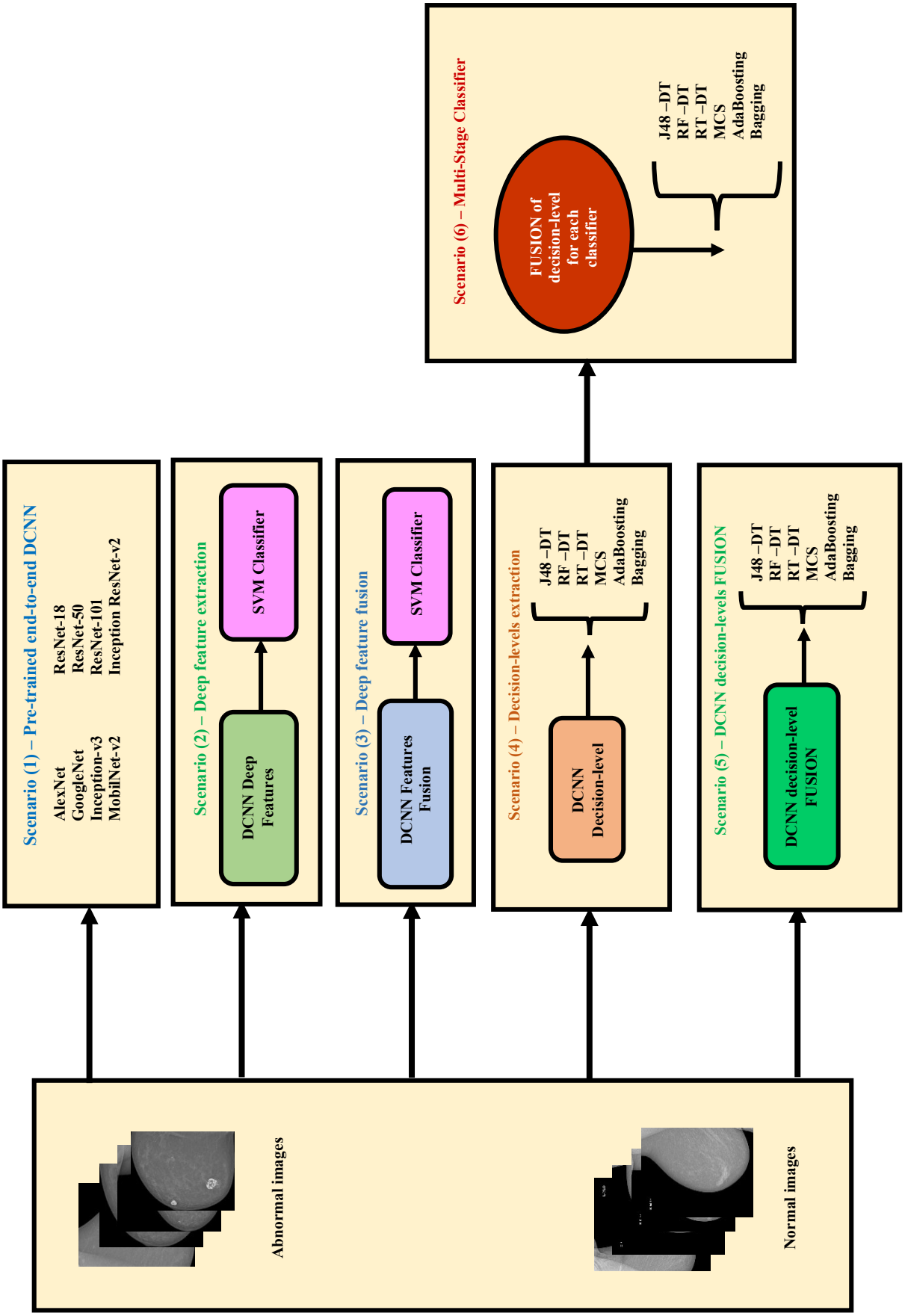


Figure 6.5: The proposed framework.

6.4 Experimental Setup

The proposed framework was applied to the mammogram images to distinguish between normal and abnormal classes. The scenarios were performed on the Intel® CORE™ I7 processor and NVIDIA GeForce 940MX, Windows 10, 64 bit with 8 GB of RAM. The software used to implement the scenarios was Matlab R2019b with an academic license provided by the University of Strathclyde. In addition, a computationally efficient tool called Waikato Environment for Knowledge Analysis (WEKA) [346] was used to perform scenarios (4), (5), and (6). WEKA is open-source software, which consists of a collection of ML algorithms for data mining tasks. The scenarios were applied to the DAR-Breast. After being confirmed by the experts' radiologists in the Hospital, the samples were categorized into normal and abnormal lesions and the ROI was determined. Additionally, all image sizes were changed according to each DCNN input image size as the input layer for each DCNN architecture was different. For example, the input layer of the Inception-v3 architecture required is 229 × 229 × 3. Therefore, the images were modified to the specific size of the network.

6.4.1 Augmentation

Various data augmentation techniques were applied to all images, including rotation, flipping horizontally, and randomly varying the image contrast and brightness by a factor between (0.5 and 2) and (-20 and 20), respectively. Firstly, the original images were rotated by four angles (0, 90, 180, and 270 degrees), and each rotated image was flipped horizontally. Therefore, in this stage, each original image was augmented to 8 images. Afterward, the image contrast and brightness were changed and the resulted images were rotated by the four angles and flipped horizontally as well. Therefore, each image was augmented to 24 images. The two-dimensional scatter plot based on the feature vectors for normal and abnormal samples is shown in Figure 6.6. This figure presents the first feature values versus the fourth feature values as an example for the features of ResNet-18 DCNN architecture for the first 10 samples of images and their augmented versions with a total of 240 images for both classes.

6.4.2 Parameter Setting

Some parameters were adjusted after fine-tuning the eight DCNNs architectures and applying the transfer learning technique to the fully connected layer. The maximum number of epochs was set to 20, although, to avoid overfitting an early stopping criteria was set. Therefore, for AlexNet, GoogleNet, ResNet-18, ResNet-50, ResNet-101, Inception-v3, Inception ResNet-v2, and MobileNet-v2 DCNNs, epoch number 13, 12, 13, 6, 10, 9, 8, and 5,

respectively proved to be the best epoch. Furthermore, the initial learning rate and the momentum for the DCNNs were 10^{-5} and 0.9, respectively. Moreover, the weight decay which is L2-regularization was used to avoid overfitting and was set to 5×10^{-4}. The mini-batch size was set to 4. Additionally, the validation frequency was set to the maximum number of iterations per epoch this is done to validate the result at the end of each epoch. The maximum number of iterations per epoch was calculated by dividing the number of training images by the mini-batch size. According to this rule, the maximum number of iterations per epoch is inversely proportional to the minibatch size. It means that when decreasing the mini-batch size, the maximum number of iterations per epoch increases, which gives more accurate results for each epoch. The other parameters were left with their default values. These configurations were to confirm that the parameters were fine-tuned for the medical breast cancer diagnosis. The optimization algorithm used was the SGDM.

6.5 Results and Discussions

The proposed framework performs six scenarios. The first three scenarios were the same as the scenarios introduced in Chapter 5 to validate the concept proposed on the new dataset DAR-Breast used in this chapter. However, in scenarios (4) – (6), new fusion procedures were presented. The transfer learning technique was offered in the first scenario to replace the last FC layer in the pre-trained DCNNs with a new one to classify two classes instead of 1000. Additionally, for the first scenario, seventy percent of the images were used for training and the rest for testing and validation as illustrated in Table 6.1. Although, for the rest of the scenarios, the ratio of the training and testing was 80%:20%. This was because these scenarios were validated using five fold cross-validation. Subsequently, the classifiers were trained with four folds and authenticated by the remaining fold. Consequently, the scenarios were trained five times and the testing classification accuracy was calculated for each time then averaged. Moreover, the nested cross-fold validation was conducted for the prune overfitting of DT classifiers. The nested cross-validation is a well-known procedure and used to reduce error pruning. It was used to overcome overfitting and over-optimistic results that may occur during model construction and parameter. Additionally, it produces a stable model which refers to how different models produced from the model selection algorithm were close to each other when training misclassification cost occurs [347]. Therefore, five-folds nested cross-validation was used for parameter selection. Similarly, five fold cross-validation was used to validate the performance of the models and avoid overfitting. Moreover, pruning was performed for tree

classifiers. For the RF DT, the number of trees generated was 10. In the following sub-sections, the results of the six scenarios will be presented and discussed.

Table 6.1: The number of training, testing, and validation samples.

		Training	Testing	Validation	Total
Normal	2952	4502	965	965	6432
Abnormal	3480				

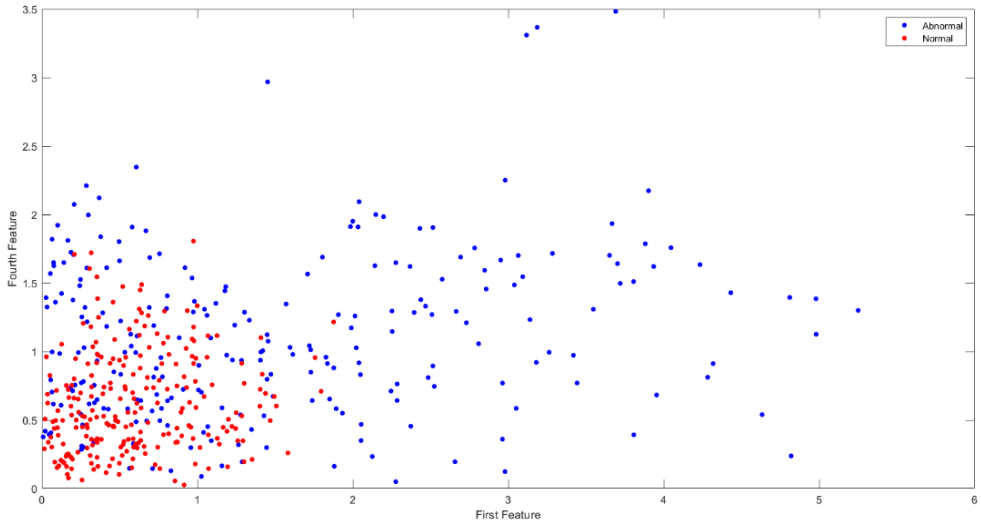


Figure 6.6: The first feature values versus the fourth feature values for the first 10 samples of images and their augmented versions for the ResNet-18 DCNN architecture features.

6.5.1 Scenario (1)

This scenario was formed via an end-to-end pre-trained DCNN of 8 different architectures; AlexNet, GoogleNet, ResNet-18, ResNet-50, ResNet-101, Inception-v3, Inception ResNet-v2, and MobileNet-v2. Table 6.2 shows the calculated accuracy, AUC, sensitivity, specificity, precision, F1-score, and the training time for each end-to-end DCNN architecture arranged in descending order according to the accuracy values. As it was obvious from Table 6.2 that the classification accuracy fluctuated from (79.79% – 92.12%) with the best accuracy achieved using ResNet-18. The AUC computed for the ResNet-18 was 0.983 (98.3%) as clear in the ROC curve shown in Figure 6.7. Additionally, the other calculated scores such as sensitivity, specificity, precision, and F1-score for the fine-tuned ResNet-18 achieved 0.976 (97.6%), 0.858 (85.8%), 0.873 (87.3%), and 0.922 (92.2%), respectively. On the other hand, the end-to-end MobileNet-v2 DCNN achieved the lowest accuracy and AUC scores of 79.79%, 0.88 (88%), respectively. In addition, it achieved 0.767 (76.7%), 0.836 (83.6%), 0.824 (82.4%), and 0.795 (79.5%) for the sensitivity, specificity, precision, and F1-score, respectively. Furthermore, the training time for the 8 networks varied between (4 – 55 hours). The Inception ResNet-v2 took a very long time to train it took 55 hours. This was

because it consists of 164 layers as shown in Table 6.2 which were the biggest number of layers compared to the other networks.

Table 6.2: The calculated scores for the end-to-end DCNN architectures.

DCNN Architectures	Accuracy	AUC	Sensitivity	Specificity	Precision	F1-score	Training Time	Number of Layers
ResNet-18	**92.12%**	**0.983**	**0.976**	**0.858**	**0.873**	**0.922**	**4 hours**	18
GoogleNet	91.04%	0.968	0.931	0.886	0.891	0.911	5 hours	22
ResNet-101	88.6%	0.955	0.926	0.840	0.853	0.889	19 hours	101
ResNet-50	87.88%	0.956	0.835	0.932	0.924	0.878	9 hours	50
AlexNet	84.97%	0.939	0.793	0.918	0.906	0.846	5 hours	8
Inception ResNet-v2	84.25%	0.933	0.951	0.716	0.770	0.851	55 hours	164
Inception-v3	84.09%	0.961	0.729	0.973	0.965	0.831	14 hours	48
MobileNet-v2	79.79%	0.880	0.767	0.836	0.824	0.795	6 hours	16

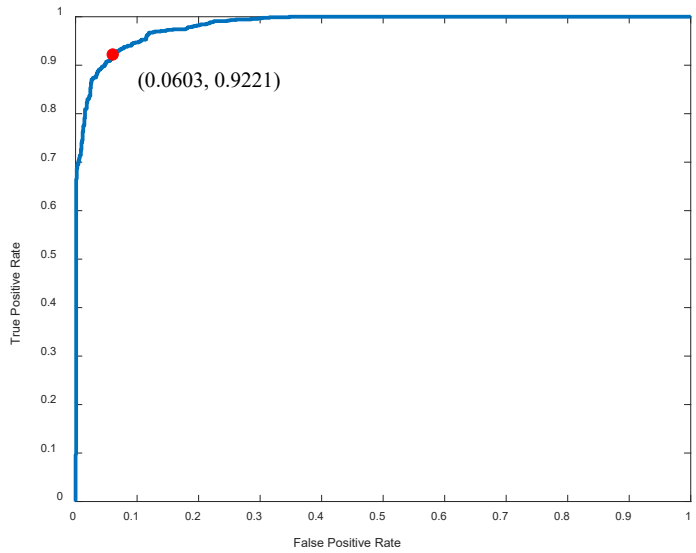
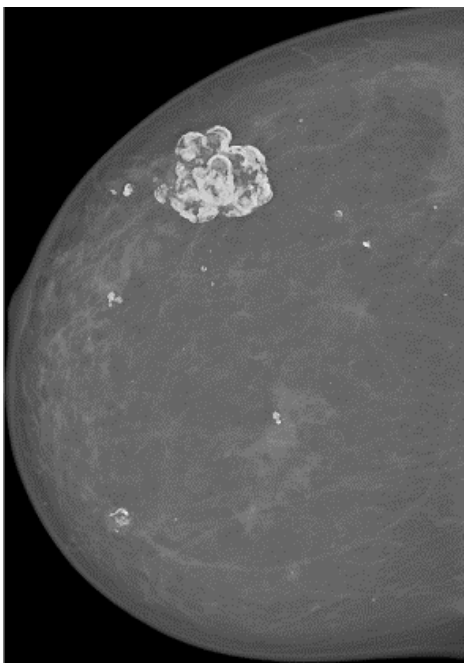
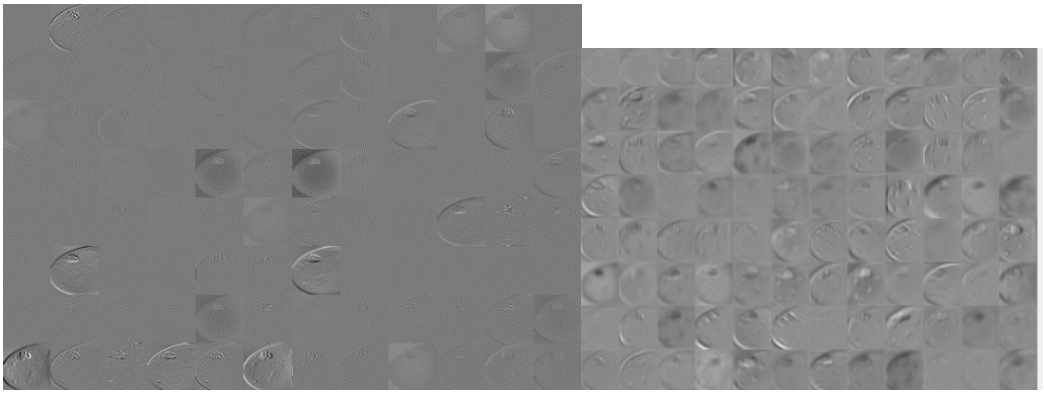


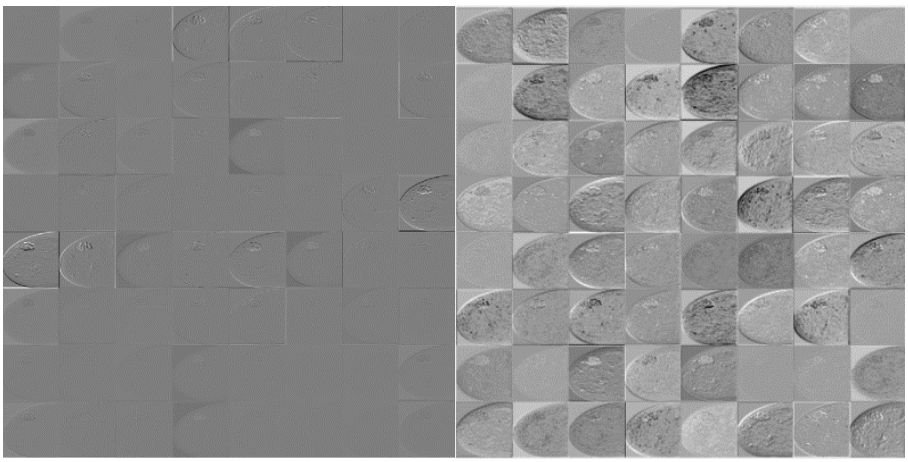
Figure 6.7: The ROC curve for ResNet-18 DCNN architecture.



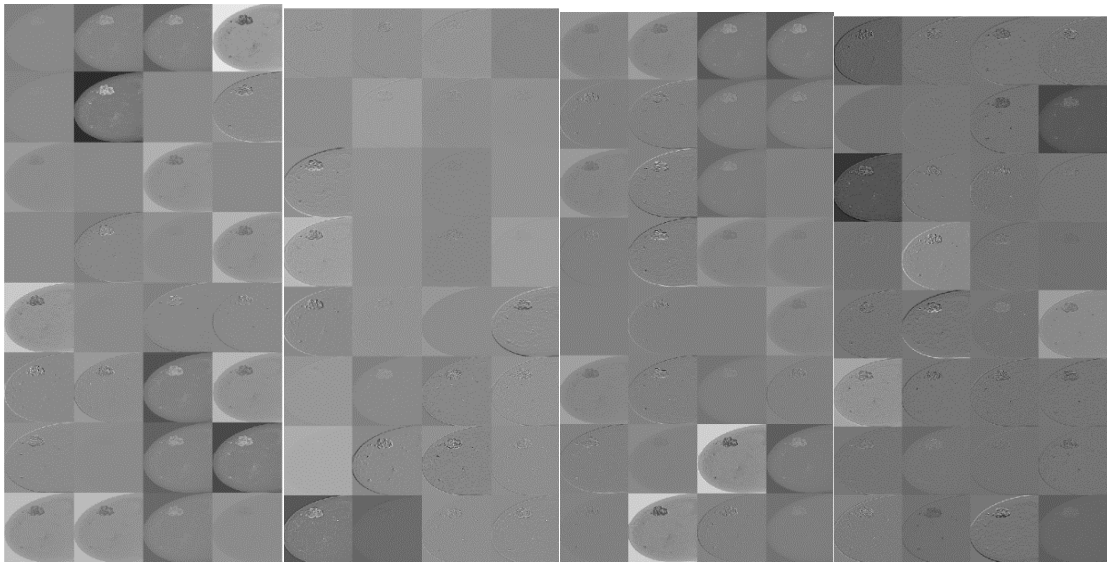
(a)



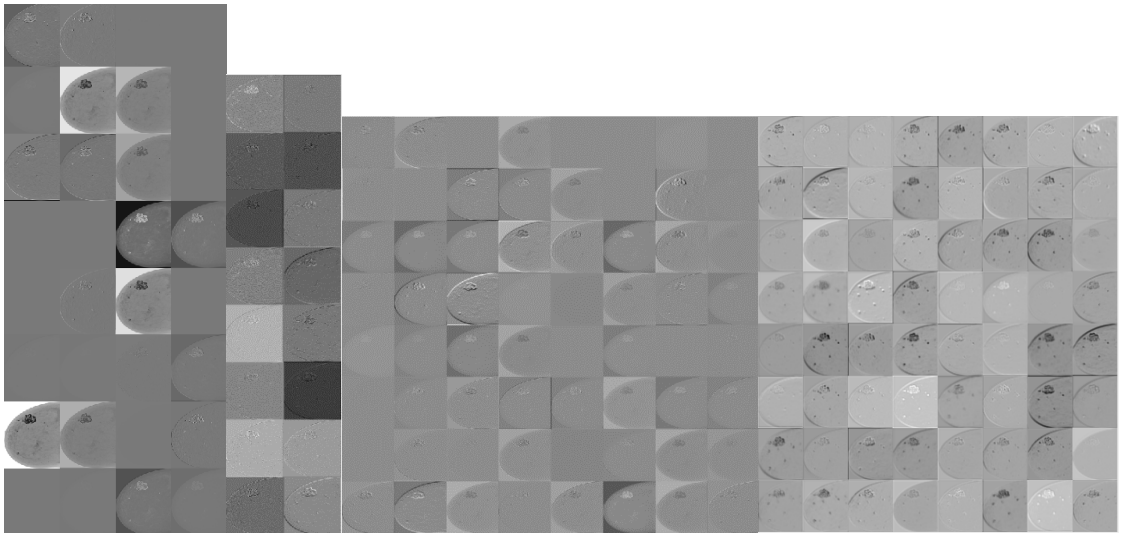
(b) (c)



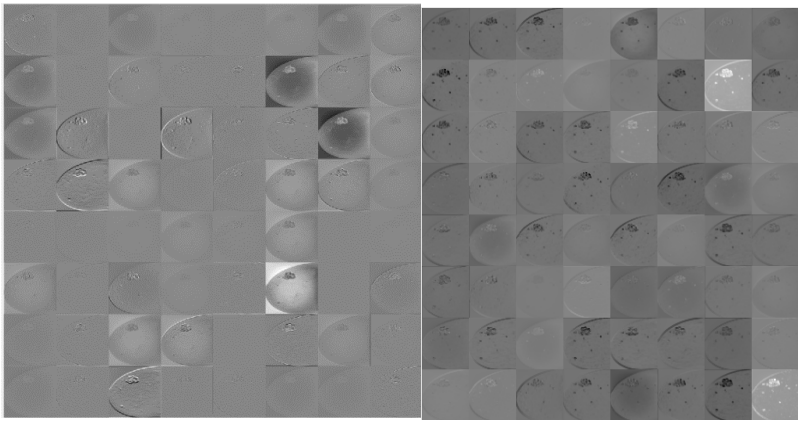
(d) (e)



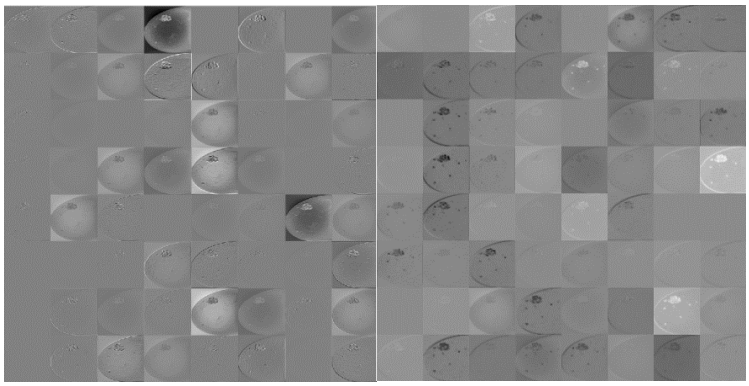
(f) (g) (h) (i)



(j) (k) (l) (m)



(n) (o)



(p) (q)

Figure 6.8: The visualisation of the first and second convolutional layers of the deep features from the 8 fine-tuned DCNN architectures.

(a) Malignant ROI image, (b) the activation features from the first convolutional layer for AlexNet DCNN, (c) the activation features from the second convolutional layer for AlexNet DCNN, (d) the activation features from the first convolutional layer for GoogleNet DCNN, (e) the activation features from the second convolutional layer for GoogleNet DCNN, (f) the activation features from the first convolutional layer for Inception-v3 DCNN, (g) the activation features from the second convolutional layer for Inception-v3 DCNN, (h) the activation features from the first convolutional layer for Inception ResNet-v2 DCNN, (i) the activation features from the second convolutional layer for Inception ResNet-v2 DCNN, (j) the activation features from the first convolutional layer for MobileNet-v2 DCNN, (k) the activation features from the second convolutional layer for MobileNet-v2 DCNN, (l) the activation features from the first convolutional layer for ResNet-18 DCNN, (m) the activation features from the second convolutional layer for ResNet-18 DCNN, (n) the activation features from the first convolutional layer for ResNet-50 DCNN, (o) the activation features from the second convolutional layer for ResNet-50 DCNN, (p) the activation features from the first convolutional layer for ResNet-101 DCNN, and (q) the activation features from the second convolutional layer for ResNet-101 DCNN.

6.5.2 Scenario (2)

In this scenario, the deep features were extracted from each network and went separately through SVM classifiers with different kernel functions to classify normal and abnormal lesions. The deep features extracted from each DCNN network are visualised in Figure 6.8 showing the first and second convolutional layers for the AlexNet, GoogleNet, Inception-v3, Inception ResNet-v2, MobileNet-v2, ResNet-18, ResNet-50, and ResNet-101, respectively. Furthermore, the different scores of the SVM classifiers with different kernels constructed with the eight deep feature sets are illustrated in Table 6.3. Additionally, Figure 6.9 shows a comparison for the classification accuracies of deep features using an SVM classifier with different kernel functions. It was obvious from Figure 6.9 that the classification accuracies had increased and ranged between 91.3% and 98.0% compared to the classification accuracies obtained in scenario (1). Besides, all the other scores obtained from this scenario compared to those in the first scenario increased as well, as illustrated in Table 6.3. The deep features of the ResNet-18, GoogleNet, and ResNet-50 proved to be the highest compared to the other networks. However, there was a small slight difference between the accuracies of these networks. The ResNet-18 deep features attained the highest classification accuracy, using the Medium Gaussian kernel function SVM classifier. The accuracy yielded to 98%. Moreover, the AUC achieved 1.000, however, the sensitivity, specificity, precision, and F1-score attained the same value of 0.981 (98.1%) as shown in Table 6.3. Additionally, the deep features of the GoogleNet ranked the second-highest classification accuracy of 97.8% via the quadratic kernel SVM classifier. The AUC and the sensitivity were 1.000 and 0.978 (97.8%), respectively. Although the rest of the scores were the same achieving 0.979 (97.9%). Furthermore, the accuracy and the AUC of the deep features of the ResNet-50 scored 97.7% and 0.990, respectively. Whereas the sensitivity, specificity, precision, and F1-score achieved almost the same scores with a difference of 0.002%. Although, these scores were accomplished using the cubic kernel SVM classifier. Additionally, as it was noticeable from Table 6.3 that the cubic kernel SVM classifier achieved the highest classification accuracy compared with the other kernels. Consequently, the accuracy in this scenario increased by 5.88%.

Table 6.3: The calculated scores of the different SVM kernel functions for the DCNN deep features.

DCNN	Different Kernels	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
ResNet-18	Linear	97.9% (0.001)	1.000* (0)	0.979 (0.002)	0.98 (0.001)	0.980 (0.001)	0.979 (0.001)
	Quadratic	97.8% (0.001)	1.000* (0)	0.978 (0.001)	0.978 (0.002)	0.978 (0.002)	0.978 (0.001)
	Cubic	97.8% (0.001)	1.000* (0)	0.978 (0.001)	0.980 (0.001)	0.980 (0.001)	0.979 (0.001)
	Medium Gaussian	**98%** (0.001)	**1.000*** (0)	**0.981** (0.001)	**0.981** (0.001)	**0.981** (0.001)	**0.981** (0.001)
	Coarse Gaussian	97.2% (0.002)	1.000* (0)	0.977 (0.001)	0.969 (0.004)	0.968 (0.004)	0.973 (0.003)
GoogleNet	Linear	97.7% (0.001)	1.000* (0)	0.975 (0.001)	0.981 (0.002)	0.981 (0.002)	0.978 (0.001)
	Quadratic	**97.8%** (0.001)	**1.000*** (0)	**0.978** (0.001)	**0.979** (0.001)	**0.979** (0.001)	**0.979** (0.001)
	Cubic	97.7% (0.001)	1.000* (0)	0.976 (0.001)	0.978 (0.002)	0.978 (0.002)	0.977 (0.001)
	Medium Gaussian	97.5% (0.001)	1.000* (0)	0.969 (0.002)	0.981 (0.001)	0.981 (0.001)	0.975 (0.002)
	Coarse Gaussian	97.1% (0.001)	1.000* (0)	0.974 (0.002)	0.970 (0.002)	0.969 (0.002)	0.972 (0.002)
ResNet-101	Linear	97.1% (0.001)	1.000* (0)	0.969 (0.002)	0.974 (0.001)	0.974 (0.001)	0.972 (0.001)
	Quadratic	97.3% (0.001)	1.000* (0)	0.972 (0.001)	0.975 (0.001)	0.975 (0.001)	0.974 (0.001)
	Cubic	**97.5%** (0.001)	**0.991** (0.004)	**0.974** (0.001)	**0.976** (0.001)	**0.976** (0.001)	**0.976** (0.001)
	Medium Gaussian	97.3% (0.001)	0.999 (0.004)	0.972 (0.001)	0.976 (0.001)	0.976 (0.001)	0.974 (0.001)
	Coarse Gaussian	96.2% (0.005)	0.990 (0)	0.962 (0.005)	0.961 (0.006)	0.960 (0.006)	0.962 (0.005)
ResNet-50	Linear	97.3% (0.001)	1.000* (0)	0.973 (0.001)	0.974 (0.001)	0.974 (0.001)	0.974 (0.001)
	Quadratic	97.6% (0.001)	1.000* (0)	0.976 (0.002)	0.976 (0.001)	0.976 (0.001)	0.976 (0.002)
	Cubic	**97.7%** (0.001)	**0.990** (0)	**0.977** (0.002)	**0.978** (0)	**0.978** (0)	**0.978** (0.001)
	Medium Gaussian	97.5% (0.001)	1.000* (0)	0.975 (0.001)	0.976 (0.001)	0.976 (0.001)	0.976 (0.001)
	Coarse Gaussian	96.1% (0.001)	0.990 (0)	0.966 (0)	0.957 (0.001)	0.956 (0.001)	0.961 (0.001)
AlexNet	Linear	96.0% (0.002)	0.990 (0)	0.957 (0.002)	0.963 (0.003)	0.963 (0.003)	0.960 (0.001)
	Quadratic	96.3% (0.002)	0.990 (0)	0.963 (0.003)	0.965 (0.002)	0.964 (0.002)	0.964 (0.001)
	Cubic	**96.4%** (0.002)	**0.990** (0)	**0.962** (0.003)	**0.966** (0.001)	**0.966** (0.001)	**0.964** (0.002)
	Medium Gaussian	96.2% (0.002)	0.990 (0)	0.964 (0.001)	0.962 (0.004)	0.961 (0.004)	0.963 (0.002)
	Coarse Gaussian	95.5% (0.001)	0.990 (0)	0.961 (0.002)	0.950 (0.003)	0.949 (0.003)	0.955 (0.001)
Inception ResNet-v2	Linear	92.0% (0.002)	0.980 (0.001)	0.893 (0.004)	0.947 (0.001)	0.95 (0.001)	0.921 (0.002)
	Quadratic	93.3% (0.004)	0.981 (0.004)	0.921 (0.011)	0.948 (0.002)	0.949 (0.002)	0.933 (0.004)
	Cubic	**94.2%** (0.003)	**0.984** (0.006)	**0.936** (0.004)	**0.948** (0.002)	**0.948** (0.002)	**0.942** (0.003)
	Medium Gaussian	93.2% (0.004)	0.980 (0.001)	0.913 (0.007)	0.950 (0.002)	0.952 (0.002)	0.932 (0.004)
	Coarse Gaussian	91.3% (0.001)	0.980 (0.001)	0.882 (0.009)	0.950 (0.001)	0.954 (0.001)	0.917 (0.005)
Inception-v3	Linear	96.4% (0.001)	0.990 (0)	0.968 (0.002)	0.960 (0.002)	0.959 (0.002)	0.964 (0.001)
	Quadratic	96.7% (0.001)	0.990 (0)	0.969 (0.003)	0.967 (0.001)	0.966 (0.002)	0.968 (0.001)
	Cubic	**96.9%** (0.002)	**0.990** (0)	**0.968** (0.003)	**0.971** (0.001)	**0.971** (0.001)	**0.970** (0.002)
	Medium Gaussian	96.5% (0.001)	0.990 (0)	0.967 (0.001)	0.964 (0.001)	0.963 (0.002)	0.965 (0.001)
	Coarse Gaussian	95% (0.002)	0.990 (0)	0.958 (0.003)	0.944 (0.003)	0.943 (0.003)	0.951 (0.002)
MobileNet-v2	Linear	94.0% (0.002)	0.980 (0.001)	0.937 (0.001)	0.943 (0.003)	0.943 (0.003)	0.940 (0.002)
	Quadratic	94.7% (0.002)	0.990 (0)	0.946 (0.001)	0.948 (0.002)	0.948 (0.002)	0.947 (0.001)
	Cubic	**95.2%** (0.002)	**0.990** (0)	**0.951** (0.002)	**0.954** (0.002)	**0.954** (0.002)	**0.952** (0.002)
	Medium Gaussian	94.8% (0.003)	0.990 (0)	0.946 (0.004)	0.950 (0.002)	0.950 (0.002)	0.948 (0.003)
	Coarse Gaussian	92.9% (0.002)	0.980 (0.001)	0.930 (0.001)	0.930 (0.001)	0.929 (0.002)	0.930 (0.001)
* rounded values							

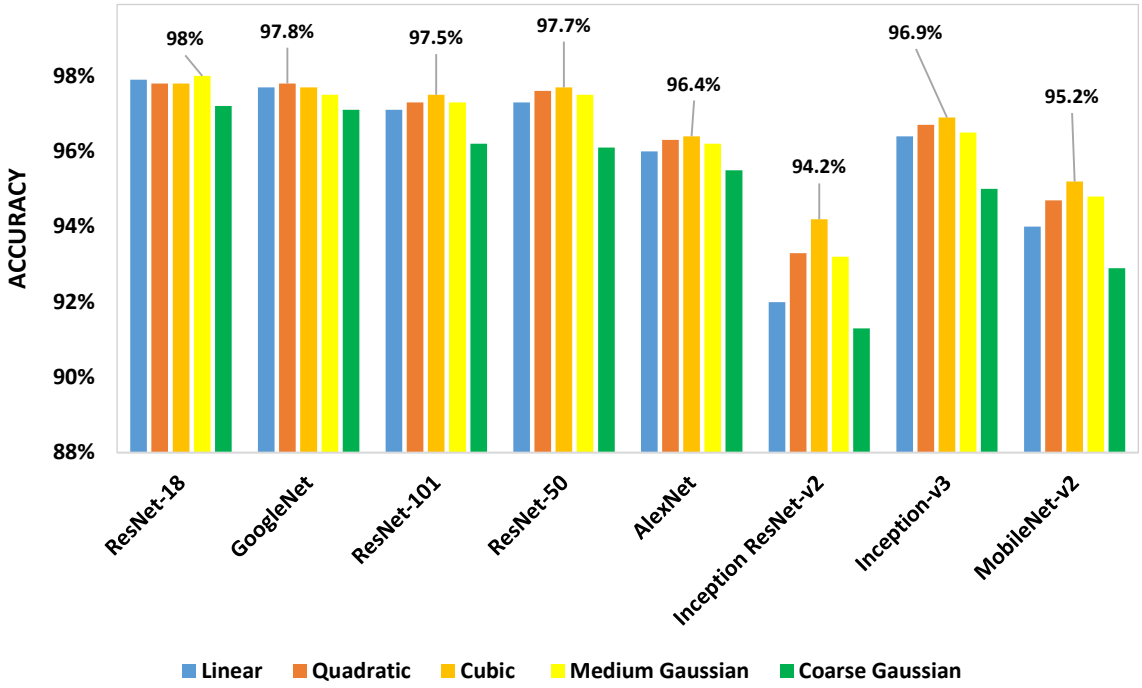


Figure 6.9: The classification accuracies of deep features using SVM classifier with different kernel functions.

6.5.3 Scenario (3)

This scenario was performed to test if combining deep features would improve the performance of the SVM classifiers. Hence, seven feature sets representing a different fusion of deep features were generated. These sets were made-up by arranging the classification accuracies produced in the first scenario in descending order and compose the feature sets by adding the features in a sequential forward strategy. From Table 6.2, since the end-to-end ResNet-18 and GoogleNet DCNNs attained the highest accuracy in the first scenario, therefore, the feature set (1) included these features, giving one feature vector with 1536 features. Then the rest of the sets were generated in the same manner. The features of ResNet-101 were added on the feature set (1) producing one feature vector with 3584 features and named feature set (2). Feature set (3) was 5632 features including the ResNet-50 features added on the feature set (2). Additionally, the AlexNet features were added on the feature set (3) giving one feature vector of 9728 features named feature set (4). Feature set (5) included the combination of feature set (4) and the features of Inception ResNet-v2. The length of the feature set (5) became 11264 features. Furthermore, Inception-v3 features were added on the feature set (5) generating feature set (6) with 13312 features. Finally, all the features were fused to produce feature set (7) containing 14592 features. Table 6.4 tabulates all the scores calculated for the feature sets. Moreover, a comparison between the classification accuracies of different SVM kernels for the different deep feature sets is given in Figure 6.10.

The highest accuracy for all the feature sets was obtained using the linear kernel SVM as shown in Figure 6.10, and they were almost the same but varying by 0.3%. Additionally, all the other scores for all the feature sets were almost the same with a small slight difference, as tabulated in Table 6.4. Therefore, this scenario proved that the fusion of features increased the accuracy compared to the first and second scenarios, thus, the accuracy achieved in this scenario was higher than scenario (2) by 0.8%.

Table 6.4: The different scores calculated for each feature set.

	Different Kernels	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
Feature Set (1)	**Linear**	**98.6%** (0.001)	**1.000*** (0)	**0.986** (0.001)	**0.986** (0.001)	**0.986** (0.001)	**0.986** (0.001)
	Quadratic	98.5% (0.001)	1.000* (0)	0.985 (0.001)	0.985 (0.001)	0.984 (0.001)	0.985 (0.001)
	Cubic	98.4% (0.002)	1.000* (0)	0.985 (0.001)	0.985 (0.001)	0.985 (0.001)	0.985 (0.001)
	Medium Gaussian	98.4% (0.001)	1.000* (0)	0.983 (0.001)	0.985 (0.002)	0.985 (0.001)	0.984 (0.002)
	Coarse Gaussian	98.1% (0.001)	1.000* (0)	0.984 (0.002)	0.979 (0.003)	0.978 (0.003)	0.981 (0.002)
Feature Set (2)	**Linear**	**98.6%** (0.001)	**1.000*** (0)	**0.986** (0.002)	**0.986** (0.001)	**0.986** (0.001)	**0.986** (0.001)
	Quadratic	98.5% (0.001)	1.000* (0)	0.985 (0.002)	0.986 (0.001)	0.986 (0.001)	0.986 (0.001)
	Cubic	98.5% (0.001)	1.000* (0)	0.985 (0.001)	0.985 (0.002)	0.985 (0.001)	0.985 (0.001)
	Medium Gaussian	98.3% (0.002)	1.000* (0)	0.983 (0.001)	0.984 (0.001)	0.984 (0.002)	0.984 (0.002)
	Coarse Gaussian	98.3% (0.001)	1.000* (0)	0.987 (0.002)	0.980 (0.001)	0.979 (0.002)	0.983 (0.001)
Feature Set (3)	**Linear**	**98.5%** (0.001)	**1.000*** (0)	**0.985** (0.002)	**0.985** (0.002)	**0.984** (0.001)	**0.985** (0.001)
	Quadratic	98.5% (0.001)	1.000* (0)	0.985 (0.002)	0.985 (0.002)	0.985 (0.001)	0.985 (0.001)
	Cubic	98.5% (0.001)	1.000* (0)	0.985 (0.003)	0.985 (0.001)	0.985 (0.001)	0.985 (0.001)
	Medium Gaussian	98.5% (0.003)	1.000* (0)	0.985 (0.002)	0.985 (0.002)	0.984 (0.001)	0.985 (0.003)
	Coarse Gaussian	98.3% (0.002)	1.000* (0)	0.990 (0.001)	0.978 (0.003)	0.977 (0.001)	0.984 (0.002)
Feature Set (4)	**Linear**	**98.7%** (0.001)	**1.000*** (0)	**0.990** (0.001)	**0.986** (0.001)	**0.985** (0.001)	**0.988** (0.001)
	Quadratic	98.7% (0.001)	1.000* (0)	0.989 (0.002)	0.986 (0.002)	0.985 (0.001)	0.987 (0.001)
	Cubic	98.6% (0.001)	1.000* (0)	0.987 (0.001)	0.986 (0.002)	0.985 (0.003)	0.986 (0.002)
	Medium Gaussian	98.6% (0.001)	1.000* (0)	0.989 (0.003)	0.984 (0.003)	0.983 (0.001)	0.986 (0.001)
	Coarse Gaussian	98.4% (0.001)	1.000* (0)	0.987 (0.001)	0.981 (0.002)	0.980 (0.001)	0.984 (0.002)
Feature Set (5)	**Linear**	**98.7%** (0.001)	**1.000*** (0)	**0.989** (0.001)	**0.986** (0.001)	**0.985** (0.001)	**0.987** (0.001)
	Quadratic	98.6% (0.001)	1.000* (0)	0.986 (0.002)	0.986 (0.001)	0.986 (0.001)	0.986 (0.002)
	Cubic	98.5% (0.001)	1.000* (0)	0.985 (0.002)	0.986 (0.001)	0.986 (0.001)	0.986 (0.001)
	Medium Gaussian	98.6% (0.001)	1.000* (0)	0.989 (0.001)	0.985 (0.001)	0.984 (0.002)	0.987 (0.003)
	Coarse Gaussian	98.5% (0.001)	1.000* (0)	0.989 (0.001)	0.983 (0.001)	0.982 (0.002)	0.986 (0.001)
Feature Set (6)	**Linear**	**98.8%** (0.001)	**1.000*** (0)	**0.990** (0.001)	**0.986** (0.001)	**0.985** (0.001)	**0.988** (0.002)
	Quadratic	98.6% (0.001)	1.000* (0)	0.986 (0.001)	0.986 (0.001)	0.986 (0.001)	0.986 (0.001)
	Cubic	98.6% (0.001)	1.000* (0)	0.986 (0.001)	0.987 (0.001)	0.987 (0.001)	0.987 (0.001)
	Medium Gaussian	98.6% (0.001)	1.000* (0)	0.989 (0.002)	0.985 (0.001)	0.984 (0.002)	0.987 (0.002)
	Coarse Gaussian	98.5% (0.001)	1.000* (0)	0.989 (0.003)	0.981 (0.001)	0.980 (0.001)	0.985 (0.002)
Feature Set (7)	**Linear**	**98.8%** (0.001)	**1.000*** (0)	**0.991** (0.001)	**0.987** (0.001)	**0.986** (0.002)	**0.989** (0.001)
	Quadratic	98.7% (0.001)	1.000* (0)	0.987 (0.001)	0.988 (0.001)	0.988 (0.001)	0.988 (0.002)
	Cubic	98.6% (0.001)	1.000* (0)	0.986 (0.002)	0.988 (0.001)	0.988 (0.001)	0.987 (0.002)
	Medium Gaussian	98.8% (0.001)	1.000* (0)	0.991 (0.001)	0.987 (0.002)	0.986 (0.001)	0.989 (0.002)
	Coarse Gaussian	98.4% (0.001)	1.000* (0)	0.989 (0.002)	0.981 (0.002)	0.980 (0.003)	0.985 (0.002)

* rounded values

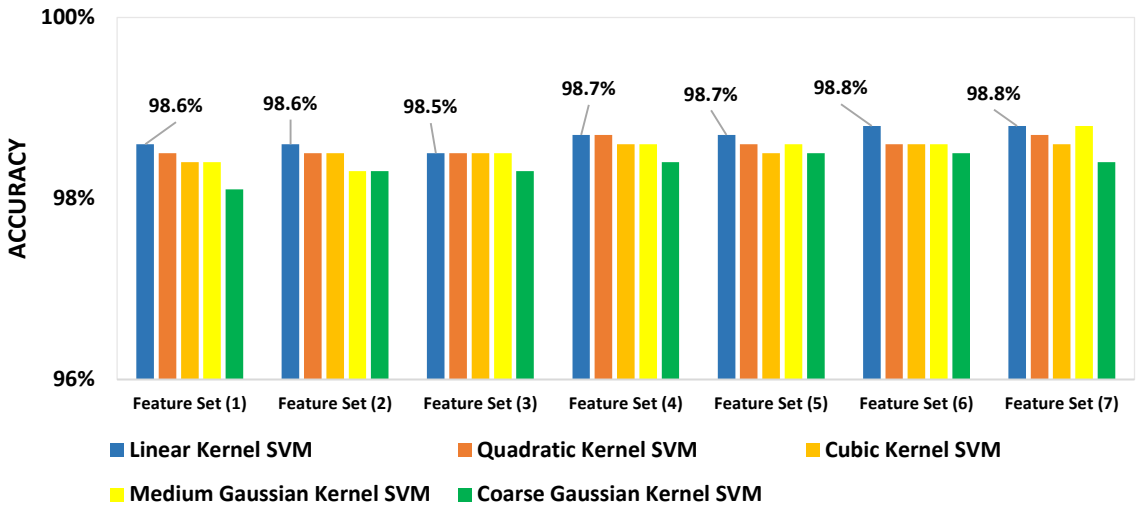


Figure 6.10: The classification accuracies of different SVM kernels for the different DCNN deep feature sets.

Feature Set (1) = ResNet-18 and GoogleNet.
Feature Set (2) = ResNet-18, GoogleNet, and ResNet-101.
Feature Set (3) = ResNet-18, GoogleNet, ResNet-101, and ResNet-50.
Feature Set (4) = ResNet-18, GoogleNet, ResNet-101, ResNet-50, and AlexNet.
Feature Set (5) = ResNet-18, GoogleNet, ResNet-101, ResNet-50, AlexNet, and Inception ResNet-v2.
Feature Set (6) = ResNet-18, GoogleNet, ResNet-101, ResNet-50, AlexNet, Inception ResNet-v2, and Inception-v3.
Feature Set (7) = ResNet-18, GoogleNet, ResNet-101, ResNet-50, AlexNet, Inception ResNet-v2, Inception-v3, and MobileNet-v2.

6.5.4 Scenario (4)

In this scenario, a novel technique was introduced by extracting the decision-levels of each DCNN model constructed in scenario (1) instead of the features. Afterward, these outcomes were classified using individual classifiers such as J48 DT, RF DT, and RT DT. Additionally, the probabilities were classified using MCS classifiers. The MCS classifiers were implemented by three different structures. The first one was the combination of J48 DT, RF DT, and RT DT classifiers. Whereas the second and third structures were the AdaBoosting and bagging ensemble of J48 DT, RF DT, and RT DT classifiers. A comparison between the classification accuracies of individual and MCS used for all DCNN outcomes is shown in Figure 6.11. In addition, Table 6.5 tabulates the classification scores for the outcomes of the DCNN architectures.

For the individual classifiers, the J48 DT classifier achieved the highest accuracy for all DCNN decision levels ranged from 91.56% to 96.74% as shown in Figure 6.11. Additionally, it was clear from Figure 6.11 that the highest and lowest accuracies were for ResNet-18 and Inception ResNet-v2 DCNN decision-levels, respectively. Hence, for the ResNet-18 DCNN decision-levels the AUC yielded to 0.964 (96.4%). The sensitivity and the F1-score achieved 0.961 (96.1%) and 0.968 (96.8%). However, the specificity and the precision reached the same score of 0.975 (97.5%) as shown in Table 6.5.

On the other hand, for the MCS classifiers, the bagging ensemble of J48 DT, RF DT, and RT DT classifiers achieved the highest accuracy for all DCNN outcomes. The accuracies fluctuated from 89.83% to 96.24%. As it was noticeable from Table 6.5, the outcomes of ResNet-18 DCNN reached 96.24% and 0.992 (99.2%) for accuracy and AUC, respectively. However, the outcomes of the Inception-ResNet-v2 achieved 89.83% and 0.951 (95.1%) for accuracy and AUC, respectively. It was obvious that still the classification accuracy achieved in this scenario did not increase compared to that achieved in scenarios (2) and (3). Although the accuracy did not improve; however, the number of features in the feature vector decreased compared to the previous scenarios. Thus, this was efficient as reducing the number of features, reduce the complexity of the classification process.

Table 6.5: The classification scores for the decision levels of the DCNN architectures.

	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
ResNet-18						
J48 DT	**96.74%** (0.001)	**0.964** (0.005)	**0.961** (0.002)	**0.975** (0.002)	**0.975** (0.002)	**0.968** (0.001)
RF DT	94.91% (0.002)	0.986 (0.001)	0.945 (0.002)	0.954 (0.002)	0.955 (0.002)	0.950 (0.002)
RT DT	94.76% (0.002)	0.950 (0.002)	0.941 (0.003)	0.956 (0.002)	0.956 (0.002)	0.949 (0.002)
J48+RF+RT	94.95% (0.002)	0.985 (0.002)	0.946 (0.002)	0.954 (0.002)	0.955 (0.002)	0.950 (0.002)
AdaBoosting	94.91% (0.002)	0.979 (0.003)	0.945 (0.003)	0.954 (0.002)	0.955 (0.003)	0.950 (0.002)
Bagging	96.24% (0.002)	0.992 (0.001)	0.958 (0.002)	0.968 (0.002)	0.969 (0.002)	0.963 (0.002)
GoogleNet						
J48 DT	**96.45%** (0.001)	**0.970** (0.002)	**0.968** (0.005)	**0.963** (0.003)	**0.963** (0.003)	**0.966** (0.002)
RF DT	94.67% (0.002)	0.984 (0.001)	0.943 (0.003)	0.952 (0.002)	0.952 (0.002)	0.948 (0.002)
RT DT	94.65% (0.002)	0.937 (0.003)	0.942 (0.003)	0.952 (0.002)	0.953 (0.002)	0.948 (0.002)
J48+RF+RT	94.66% (0.002)	0.984 (0.001)	0.943 (0.003)	0.952 (0.002)	0.952 (0.002)	0.948 (0.002)
AdaBoosting	94.66% (0.002)	0.960 (0.003)	0.942 (0.002)	0.952 (0.002)	0.952 (0.002)	0.948 (0.002)
Bagging	95.87% (0.001)	0.989 (0.001)	0.957 (0.002)	0.962 (0.001)	0.962 (0.002)	0.960 (0.001)
ResNet-101						
J48 DT	**95.68%** (0.001)	**0.953** (0.002)	**0.940** (0.001)	**0.974** (0.001)	**0.975** (0.001)	**0.957** (0.001)
RF DT	93.25% (0.002)	0.976 (0.002)	0.928 (0.002)	0.938 (0.003)	0.938 (0.002)	0.934 (0.002)
RT DT	93.18% (0.002)	0.926 (0.002)	0.925 (0.002)	0.939 (0.003)	0.940 (0.003)	0.933 (0.002)
J48+RF+RT	93.27% (0.002)	0.976 (0.002)	0.928 (0.002)	0.938 (0.002)	0.939 (0.002)	0.934 (0.002)
AdaBoosting	93.11% (0.002)	0.957 (0.003)	0.927 (0.002)	0.936 (0.002)	0.936 (0.003)	0.932 (0.002)
Bagging	94.78% (0.002)	0.985 (0.002)	0.940 (0.002)	0.956 (0.002)	0.957 (0.003)	0.949 (0.002)
ResNet-50						
J48 DT	**95.81%** (0.001)	**0.958** (0.003)	**0.947** (0.003)	**0.970** (0.003)	**0.971** (0.004)	**0.959** (0.001)
RF DT	93.62% (0.002)	0.978 (0.001)	0.933 (0.003)	0.941 (0.002)	0.942 (0.003)	0.937 (0.002)
RT DT	93.38% (0.002)	0.938 (0.002)	0.926 (0.002)	0.942 (0.002)	0.943 (0.003)	0.935 (0.002)
J48+RF+RT	93.68% (0.002)	0.978 (0.002)	0.933 (0.003)	0.942 (0.002)	0.942 (0.003)	0.938 (0.002)
AdaBoosting	93.72% (0.003)	0.972 (0.002)	0.933 (0.003)	0.942 (0.003)	0.942 (0.003)	0.938 (0.002)
Bagging	95.30% (0.002)	0.985 (0.001)	0.947 (0.003)	0.960 (0.002)	0.961 (0.002)	0.954 (0.002)
AlexNet						
J48 DT	**95.57%** (0.001)	**0.953** (0.003)	**0.944** (0.002)	**0.968** (0.001)	**0.969** (0.001)	**0.957** (0.001)
RF DT	93.09% (0.002)	0.978 (0.001)	0.928 (0.003)	0.935 (0.002)	0.936 (0.002)	0.932 (0.002)
RT DT	93.06% (0.002)	0.931 (0.002)	0.927 (0.003)	0.935 (0.002)	0.936 (0.002)	0.932 (0.002)
J48+RF+RT	93.08% (0.002)	0.978 (0.002)	0.927 (0.003)	0.935 (0.002)	0.936 (0.002)	0.932 (0.002)
AdaBoosting	93.07% (0.002)	0.974 (0.002)	0.927 (0.003)	0.936 (0.002)	0.936 (0.002)	0.932 (0.002)
Bagging	94.59% (0.002)	0.985 (0.001)	0.941 (0.002)	0.951 (0.003)	0.952 (0.003)	0.947 (0.002)

	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
Inception ResNet-v2						
J48 DT	**91.56%** (0.001)	**0.908** (0.001)	**0.884** (0.001)	**0.947** (0.001)	**0.951** (0.001)	**0.917** (0.001)
RF DT	86.62% (0.002)	0.936 (0.002)	0.856 (0.003)	0.877 (0.003)	0.880 (0.003)	0.868 (0.002)
RT DT	86.58% (0.002)	0.864 (0.002)	0.855 (0.002)	0.877 (0.003)	0.880 (0.003)	0.868 (0.002)
J48+RF+RT	86.66% (0.002)	0.937 (0.003)	0.856 (0.003)	0.877 (0.003)	0.881 (0.003)	0.868 (0.002)
AdaBoosting	86.61% (0.002)	0.932 (0.002)	0.856 (0.003)	0.877 (0.003)	0.880 (0.003)	0.868 (0.003)
Bagging	89.83% (0.002)	0.951 (0.002)	0.880 (0.002)	0.917 (0.004)	0.920 (0.004)	0.900 (0.002)
Inception-v3						
J48 DT	**94.70%** (0.001)	**0.944** (0.002)	**0.941** (0.002)	**0.953** (0.003)	**0.953** (0.003)	**0.948** (0.001)
RF DT	91.66% (0.002)	0.970 (0.002)	0.913 (0.003)	0.921 (0.003)	0.922 (0.003)	0.918 (0.002)
RT DT	91.48% (0.002)	0.917 (0.002)	0.909 (0.003)	0.921 (0.003)	0.922 (0.003)	0.916 (0.002)
J48+RF+RT	91.68% (0.002)	0.970 (0.002)	0.913 (0.003)	0.921 (0.003)	0.922 (0.003)	0.918 (0.002)
AdaBoosting	91.66% (0.002)	0.967 (0.003)	0.913 (0.003)	0.921 (0.003)	0.922 (0.003)	0.918 (0.002)
Bagging	93.68% (0.002)	0.980 (0.002)	0.933 (0.003)	0.942 (0.002)	0.942 (0.002)	0.938(0.002)
MobileNet-v2						
J48 DT	**91.69%** (0.002)	**0.915** (0.002)	**0.920** (0.002)	**0.916** (0.003)	**0.915** (0.003)	**0.918** (0.002)
RF DT	87.63% (0.002)	0.948 (0.002)	0.869 (0.003)	0.884 (0.002)	0.885 (0.003)	0.878 (0.002)
RT DT	87.23% (0.002)	0.877 (0.002)	0.860 (0.003)	0.884 (0.003)	0.888 (0.003)	0.874 (0.002)
J48+RF+RT	87.69% (0.002)	0.950 (0.002)	0.871 (0.003)	0.884 (0.003)	0.885 (0.003)	0.878 (0.002)
AdaBoosting	87.69% (0.002)	0.940 (0.002)	0.871 (0.003)	0.883 (0.003)	0.885 (0.003)	0.878 (0.002)
Bagging	90.12% (0.002)	0.962 (0.001)	0.898 (0.003)	0.906(0.003)	0.906 (0.003)	0.902 (0.002)

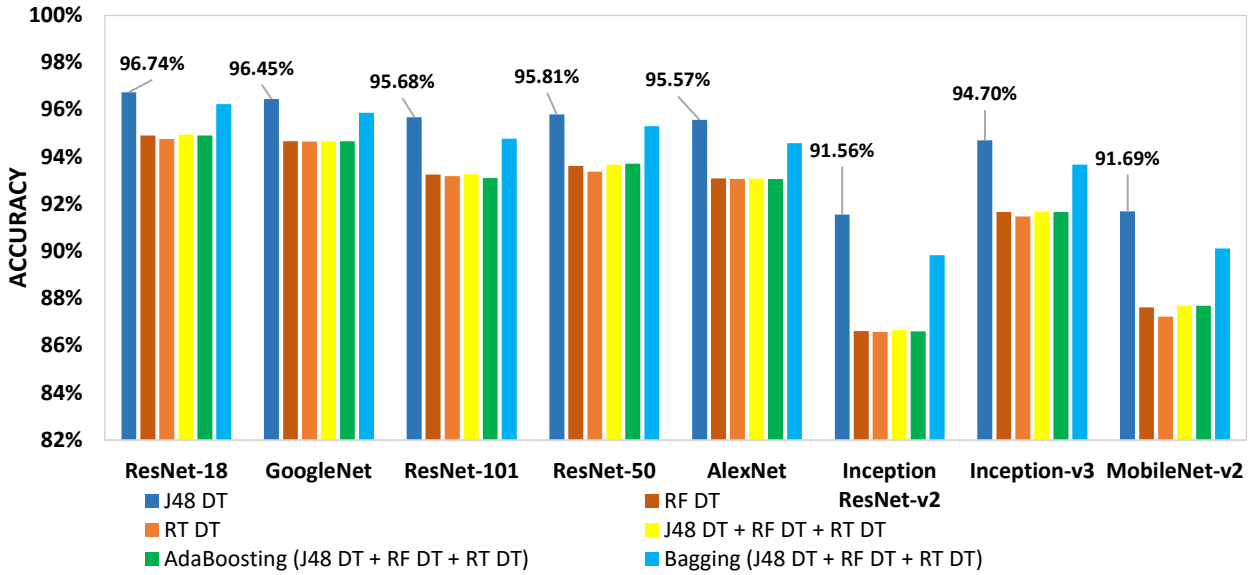


Figure 6.11: A comparison between the classification accuracies of individual and MCS used for all DCNN outcomes.

Table 6.6: The scores calculated for the fusion of the decision levels classified by individual and MCS classifiers.

	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-Score (Std)
J48 DT	**97.69%** (0.002)	**0.989** (0.002)	**0.978** (0.004)	**0.978** (0.003)	**0.977** (0.003)	**0.978** (0.002)
RF DT	97.65% (0.001)	0.999 (0.001)	0.974 (0.002)	0.980 (0.001)	0.980 (0.001)	0.977 (0.001)
RT DT	96.98% (0.002)	0.970 (0.003)	0.966 (0.003)	0.974 (0.002)	0.974 (0.002)	0.971 (0.002)
J48+RF+RT	97.49% (0.002)	0.998 (0)	0.972 (0.003)	0.979 (0.001)	0.979 (0.001)	0.976 (0.002)
AdaBoosting	97.59% (0.002)	0.994 (0.002)	0.973 (0.002)	0.980 (0.002)	0.980 (0.002)	0.977 (0.002)
Bagging	97.66% (0.001)	0.999 (0.001)	0.975 (0.002)	0.980 (0.001)	0.980 (0.001)	0.978 (0.001)

6.5.5 Scenario (5)

Another fusion strategy was performed in this scenario to determine if combining the decision-levels would enhance the performance of the classifiers as when combining the deep features in scenario (3). Therefore, the decision-levels extracted in the scenario (4) were fused to produce one feature vector with only eight features. This feature vector was also fed into the same individual and MCS constructed in the fourth scenario. The scores calculated were illustrated in Table 6.6. The highest and lowest accuracies were 97.69% and 96.98% achieved by the J48 DT and the RT DT, respectively as illustrated in Table 6.6. Furthermore, for the MCS classifiers the bagging ensemble of J48 DT, RF DT, and RT DT classifiers achieved the highest classification accuracy of 97.66%. It was clear that the accuracies achieved in this scenario did not improve compared to scenario (3), however, the feature vector length decreased.

6.5.6 Scenario (6)

In this scenario, a second stage of classification was constructed to improve the classification accuracy and decrease the complexity of the classification process. This was generated by extracting the outcomes of each classifier for the classified decision-levels in scenario (4). The decision levels of each individual and MCS classifiers for all the DCNN models were combined separately forming one feature vector with 8 features. This feature vector was again classified using the same individuals and MCS classifiers. The accuracies in this scenario fluctuated from 91.88% to 97.76%. Figure 6.12 shows the classification accuracies for all the fused decision levels of each classifier classified again using the single and MCS classifiers. As it was clear from this figure that the highest classification accuracy between all the classifiers was the fusion of the decision-levels classified by the J48 DT, therefore, the scores achieved of this classifier were presented in Table 6.7. On the other hand, the least classification accuracy was for the fusion of the decision levels of the RT DT scoring 91.88% using the J48 DT. Additionally, the fusion of the other classifiers achieved a promising result as shown in Figure 6.12.

As it was noticeable, this scenario proved that the second stage of classification improved the classification accuracy results compared to scenarios (4) and (5). Furthermore, when comparing the classification accuracy obtained in scenario (6) with the rest of the scenarios, it was clear that the accuracy decreased by 1.04% from scenario (3). However, the feature vector length decreased, which reduced the complexity of the classification process. A comparison for the classification accuracies achieved in the six scenarios was demonstrated in Figure 6.13.

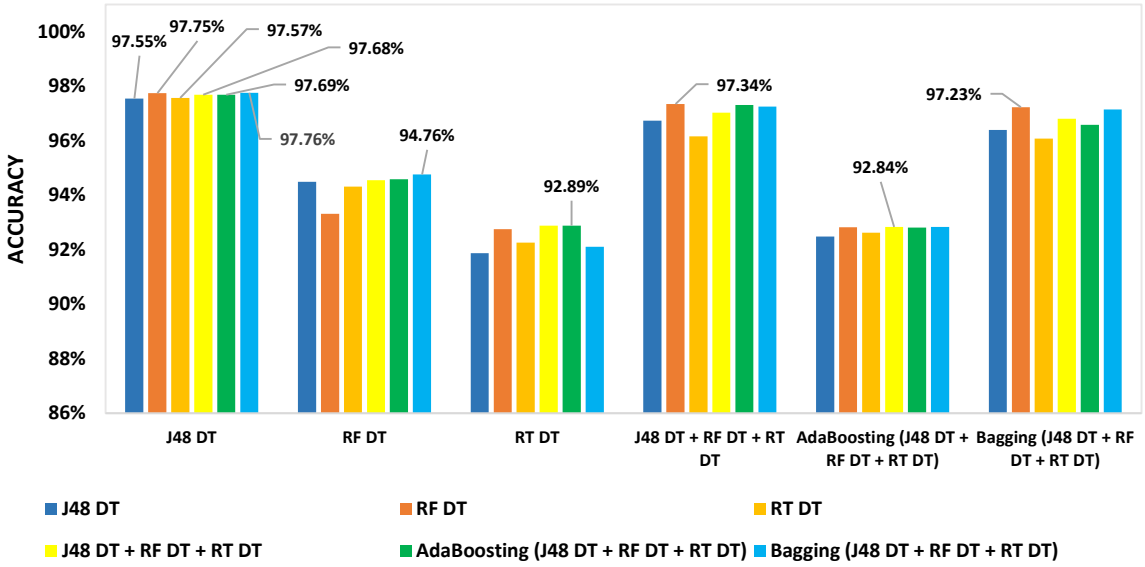


Figure 6.12: The classification accuracies of the second stage classification constructed in scenario (6).

Table 6.7: The scores for the fusion of the J48 DT outcomes classified by the individual and MCS classifiers.

	Accuracy (Std)	AUC (Std)	Sensitivity (Std)	Specificity (Std)	Precision (Std)	F1-score (Std)
J48 DT	97.55% (0.002)	0.990 (0.002)	0.978 (0.002)	0.974 (0.002)	0.973 (0.002)	0.976 (0.002)
RF DT	97.75% (0.001)	0.996 (0.001)	0.982 (0.002)	0.975 (0.002)	0.974 (0.002)	0.978 (0.001)
RT DT	97.57% (0.001)	0.977 (0.002)	0.978 (0.002)	0.976 (0.002)	0.975 (0.002)	0.976 (0.001)
J48+RF+RT	97.68% (0.001)	0.997 (0.001)	0.981 (0.002)	0.974 (0.002)	0.973 (0.002)	0.977 (0.002)
AdaBoosting	97.69% (0.001)	0.990 (0.002)	0.981 (0.002)	0.974 (0.002)	0.973 (0.002)	0.977 (0.001)
Bagging	**97.76%** (0.001)	**0.997** (0.001)	**0.982** (0.002)	**0.974** (0.002)	**0.973** (0.002)	**0.978** (0.001)

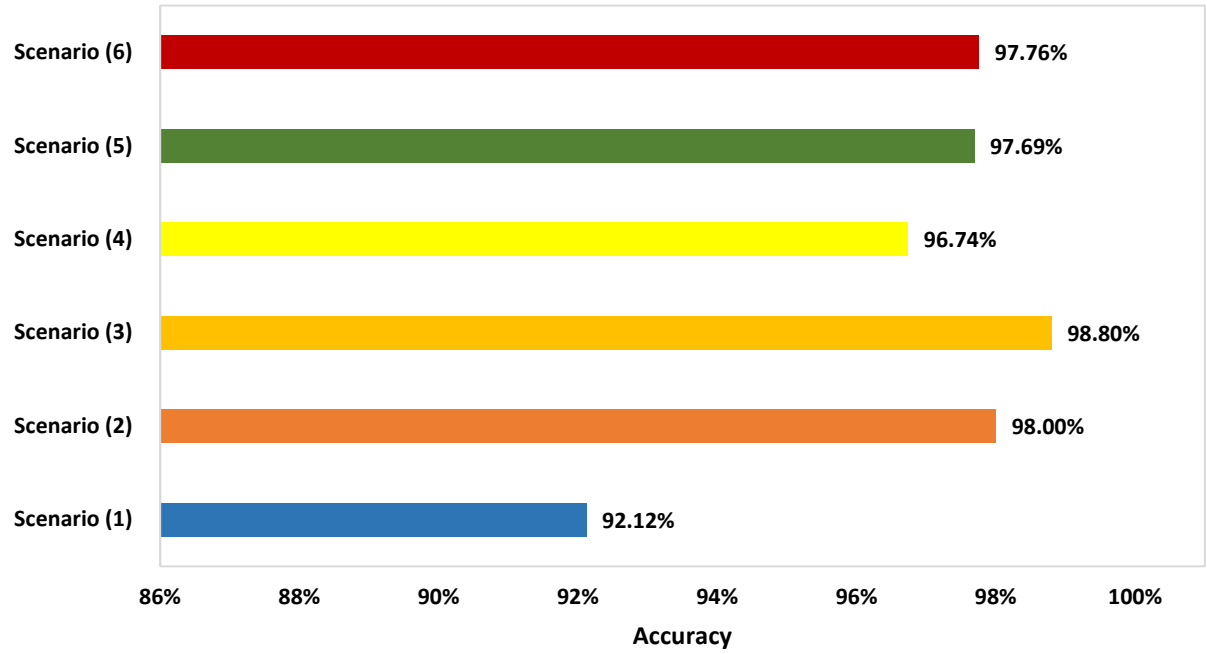


Figure 6.13: A comparison for the accuracies achieved in the six scenarios.

For any medical system to be reliable, it should accomplish a sensitivity, a specificity, and a precision bigger than or equal to 80%, 95%, and 95%, respectively. Therefore, the proposed framework with all scenarios was reliable.

To authenticate the statistical significance of the results obtained in the proposed scenarios, the ANOVA test was performed by the repeated five fold cross-validation procedure. The null hypothesis Ho for all classification was that the mean accuracies achieved in each scenario. For scenario (2), the test was performed on all the classification accuracy results of the different kernels of the SVM classifier to test the statistical significance between them. Tables 6.8 – 6.10 show the ANOVA test details for the deep features of the three highest DCNN architectures results obtained in this scenario, which were ResNet-18, GoogleNet, and ResNet-50, respectively. Moreover, for scenario (3), since there was a slight difference in the classification accuracy values between the different feature sets created in this scenario. Consequently, the ANOVA test was performed on the classification accuracy results for feature set (7) as shown in Table 6.11. For scenario (4), when comparing the classified accuracies for the outcomes of the eight DCNN architectures, it was clear that the decision levels of the ResNet-18 DCNN architecture achieved the highest accuracy as in Table 6.5. Therefore, the ANOVA test was performed on the results achieved by this architecture as illustrated in Table 6.12. Additionally, Table 6.13 presents the ANOVA test details of the results executed in scenario (5). Furthermore, as the fusion of the J48 DT decision-levels classified by the J48 DT achieved the highest classification accuracy in scenario (6). Thus, the ANOVA test was performed on the results achieved on this set as illustrated in Table 6.14. As it was clear from Tables 6.8 – 6.14 that the p-values achieved were lower than α, where $\alpha = 0.05$, accordingly, there was a statistically significant difference between the accuracies of the classifiers constructed.

Table 6.8: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-18 DCNN.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00036	4	9.008×10^{-5}	95.38	< 0.001
Error	0.00004	45	9.44444×10^{-7}		
Total	0.0004	49			

Table 6.9: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of GoogleNet DCNN.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00033	4	8.25×10^{-5}	136.12	< 0.001
Error	0.00003	45	6.06667×10^{-7}		
Total	0.00036	49			

Table 6.10: The ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-50 DCNN.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00163	4	4.1×10^{-4}	1977.48	< 0.001
Error	0.00001	45	0		
Total	0.00164	49			

Table 6.11: The ANOVA test details for the different kernel functions of the SVM classifier for feature set (7) in scenario (3).

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00012	4	2.915×10^{-5}	16.42	< 0.001
Error	0.00008	45	1.77556×10^{-6}		
Total	0.0002	49			

Table 6.12: The ANOVA test details for the ResNet-18 DCNN decision-levels in scenario (4).

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00359	5	0.00072	468.37	< 0.001
Error	0.00008	54	0		
Total	0.00368	59			

Table 6.13: The ANOVA test details for the 8 DCNN decision-levels fusion in scenario (5).

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00037	5	7.34335×10^{-5}	40.17	< 0.001
Error	0.0001	54	1.82815×10^{-6}		
Total	0.00047	59			

Table 6.14: The ANOVA test details for the J48 DT decision-levels fusion in scenario (6).

Source of Variation	SS	df	MS	F	p-Value
Columns	4.00933×10^{-5}	5	8.01867×10^{-6}	11.01	< 0.001
Error	3.9314×10^{-5}	54	7.28037×10^{-7}		
Total	7.94073×10^{-5}	59			

Finally, the proposed framework has been compared with the applicable state-of-the-art frameworks presented in Chapter 3 to prove the efficiency of the proposed method. It was noticeable that the proposed framework with all scenarios has outperformed other systems. Regarding the usage of fine-tuned DCNN architectures, many researchers proposed this method as in Xi et al. [286], Ting et al. [290], Alkhaleefah et al. [298], Al-Antari et al. [301], and Wessels and Van der Haar [303]. The methods proposed by these researchers were evaluated on the publicity datasets. In addition, the classification accuracy achieved was less than the scores obtained in this work. Additionally, the results presented in this work were assessed on a new mammography dataset. Li et al. [292] evaluated the results of end-to-end fine-tuned DCNN on a private dataset collected from the First Hospital of Shanxi Medical University. The accuracy achieved was 94.55%, which also proved to be less than achieved in this framework.

On the other hand, recently, researchers led to different fusion strategies to increase the classification accuracy of a system as in [330]–[332]. The highest accuracies achieved by these researchers ranged from 82.84% to 96.6%. Therefore, when comparing these fusion methods with the one proposed in this framework, it was found that the accuracy attained beaten all other results.

6.6 Summary

In this chapter, a novel decision-levels based framework to distinguish between different breast cancer classes was presented. Addiditionally, a new mammogram dataset namely the "DAR-Breast" was acquired and used to test the performance of the work proposed. This framework was performed by six scenarios. The first scenario composed of constructing eight end-to-end pre-trained fine-tuned DCNN networks of different architectures. Furthermore, scenario (2) created by extracting the deep features of the fine-tuned DCNNs and classify them separately using SVM classifiers with different kernel functions. The goal of this scenario was to improve the classification accuracies of scenario (1). Therefore, the results revealed that the classification accuracies in scenario (2) were higher than scenario (1). Afterward, the classification accuracies achieved in scenario (1) were arranged in descending order. Accordingly, from this arrangement, seven feature sets were generated, forming scenario (3). The classification accuracy achieved in this scenario increased by 0.8% than in scenario (2).

Furthermore, a novel technique to enhance the classification accuracy of the breast cancer classification problem was presented in scenarios (4) – (6). In scenario (4), the decision-levels of the DCNN networks constructed in scenario (1) were extracted separately and considered as features. This was performed to decrease the feature length vector, which proved its efficiency to lessen the complexity of the classification process. These outcomes were fed into individual and MCS classifiers to test the accuracy. The individual classifiers included different types of DT classifiers this was because DT achieved a high accuracy rate in the medical field. The MCS classifiers were constructed using AdaBoosting and bagging ensemble. The accuracies achieved in this scenario did not improve compared to those of scenarios (2) and (3), however, the feature vector decreased. In scenario (5), all the outcomes extracted from scenario (4) were combined forming a new feature vector with only eight features in length. The accuracy achieved in this scenario increased compared to that in scenario (4).

Finally, in scenario (6), another new fusion method was presented by constructing a second stage of classification. The decision-levels of each classifier for the classified outcomes in scenario (4) were extracted. These outcomes were used to construct a second level of classification by fusing separately for each individual and MCS classifiers' outcomes forming one feature vector with eight features. Afterward, the new feature vector was classified using the individual and MCS as well. In this case, the fusion of J48 DT features achieved the highest classification accuracy. All the accuracy results obtained from all scenarios were analysed by the ANOVA test.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Incidence of breast cancer rates is found to be maximal and its survival rates are very less as soon as patients reach stage three of the disease. Moreover, radiologists cannot easily provide accurate evaluation due to the huge number of mammograms generated in widespread screening. To combat these problems, modern techniques and recent advancements in diagnosis are essential to control the progression of tumours and reduce death rates. They are considered as a boon to mankind because they can provide early detection in lesser time and are also non-invasive in nature. Thus, breast cancer can be treated at its earliest stage when compared to conventional methods. Therefore, it is necessary to use new techniques and concentrate on development in the diagnosis of the disease so that treatment of the disease can be made easier and in less possible time. Consequently, CAD systems have been developed to detect the indicators of breast cancer and improve the accuracy of diagnosis.

Currently, ML and specifically the DL methods were used extensively to tackle medical problems. The DL techniques were considered the newest class of ML techniques and were used extensively, as they can overcome the limitations of the classical ML methods with hand-crafted features. DL methods showed their superiority over classical ML methods in most cases.

The training process of DCNNs requires a large number of annotated samples to avoid overfitting to the training dataset, however, there was always an insufficient amount of medical images. In addition, with the increasing amount of data, DCNNs outperform conventional ML techniques. Therefore, the data augmentation technique with different forms was applied to all the datasets used in this work.

In this thesis, we proposed a combination of DL techniques to build efficient frameworks in CAD systems to assist the radiologists in classifying mammography breast cancer tumours. We tested these techniques using three publicly available datasets, MIAS, DDSM, and CBIS-DDSM. We have demonstrated to achieve the state-of-the-art performance using our proposed methodologies. The main contributions of this thesis can be summarised as follows:

In Chapter 4, we proposed two segmentation methodologies i.e., the circular contours and adaptive thresholding to segment the ROI from the original image of the DDSM dataset. The AlexNet DCNN was employed as an end-to-end process, achieving an accuracy for the circular contour method higher than that of the adaptive thresholding method. Additionally, to improve the classification accuracy, the deep features were extracted and classified using the SVM classifier. It was found that the adaptive thresholding method provided better results. On the other hand, for the CBIS-DDSM dataset, the classification accuracy of the end-to-end DCNN increased compared to that achieved using the DDSM samples. Moreover, when classifying the deep features extracted using SVM, the accuracy increased as well. Therefore, the highest accuracy achieved from this framework was 87.2% using the images of the CBIS-DDSM dataset.

In Chapter 5, a framework is proposed to explore different combinations of deep features and select the optimal one, to further improve the classification accuracy. In addition, it has also reduced the computational cost through PCA based dimension reduction of the fuse features. This was performed by constructing four different scenarios and evaluated on two datasets; the CBIS-DDSM and the MIAS. The scenarios were: (1) end-to-end DCNN process, (2) extracting the deep features and classifying them using the SVM classifier with different kernel functions, (3) fusing the deep features in a sequential forward strategy, and classify them sing the SVM classifier with different kernel functions, and (4) reducing the feature dimension and the computational cost. The results of the scenarios indicated that the proposed framework is capable of successfully classifying breast cancer lesions, where the highest accuracies achieved were 97.9% and 97.4% using the fusion of deep features for CBIS-DDSM and MIAS datasets, respectively.

In Chapter 6, novel fusion techniques were used to construct the third framework, by generating six scenarios. The first three scenarios were the same as the ones constructed in the preceding framework in Chapter 5. Although, there was a slight change in scenario (1), using Inception-v3, Inception ResNet-v2, and MobileNet-v2 DCNNs models besides the ones used previously in Chapter 5. On the other hand, scenarios (4) – (6), employed a novel technique to enhance and improve the classification accuracy by decision-level fusion of the outcomes of the DCNNs and classifying them with different trees of DT and their ensembles. This framework was reliable and efficient and it was capable to classify the breast lesions and decrease the complexity. Additionally a new mammogram dataset collected from the Armed Forces Hospital located in Egypt named DAR-Breast was also introduced.

In conclusion, this study is a crucial trial compromising a simple construction, low cost, efficient, and automatic diagnostic tool that can achieve a high accuracy based on multiple DCNNs and various fusion techniques. The DCNN methods are more capable of distinguishing between cancerous and non-cancerous cases than manual interpretation of mammogram images. These frameworks can assist the radiologists to accurately diagnose the breast cancer. It can also reduce the time and effort made by them during the examination process and reduce human misdiagnosis that could occur due to human fatigue. Moreover, the frameworks presented in this thesis could also be potentially adapted and applied for the detection and classification of cancers in any other organs.

7.2 Future Work

The work presented in this thesis was found to successfully obtain some promising results to the automatic interpretation of breast cancer images that might decreases human mistakes in the process of the diagnosis. Moreover, it reduces the complexity and the cost of breast cancer diagnosis. However, there are still many aspects that are worth further investigation to increase the classification accuracy. These are summarised in several future directions as detailed below:

- Deep-learning networks of other kinds, such as RNN, generative adversarial networks (GANs) [348], and clustering should be explored.
- Strategies based on various DCNN architectures, as well as hyperparameter optimization, must be investigated.
- Combining handcrafted features such as statistical features, GLCM features, and Curvelet transform with deep features.
- Extracting the deep features from the handcrafted features may enhance the performance of the classifier.
- Datasets focused on mammography and histopathology are publicly available. In contrast, datasets based on other imaging modalities such as infrared thermal imaging, computed tomography, and digital breast tomosynthesis are not publicly available. Additionally, studies conducted using such imaging modalities go through unpublished datasets. Therefore, datasets for new breast imaging modalities could be collected and acquired for breast cancer diagnosis.

- Segmentation models, such as U-net, have provided cutting-edge segmentation findings in a variety of computer vision datasets [349]. Furthermore, applying this technique involving various imaging modalities may boost breast cancer classification accuracy results.
- Combining the clinical data with the deep features which may boost the performance of the classifier.
- Multi-modality research could be investigated. This coould be performed by combining schemes based on multiple different breast imaging modalities, such as mammography with MRI scans and mammography with PET scans.
- Multi-centre studies could be implemented to construct a powerful framework. The multi-centre studies depends on a dataset obtained from a single scanning centre lead to model complications and make it difficult to be employed on datasets from other locations or using different acquiring device or scanning parameters [350]. Thus, datasets from multi-centre are necessary. The multi-centre study includes combining datasets from several centres acquired from either different locations, scanner, or scanning parameters to increase the size of the dataset used to construct and train the CAD system. Datasets from multiple centres are effectively grouped for detection and repetition [351], [352]. The training set of the data used to construct the model will contain data from the multi-centre, which accommodates the deviation produced due to the variation in scanner or scanning parameters. Additionally, combining data from multi-centre will help recruitment, strengthen the CAD system power, and cope among scanner variance in order to generate better generalisable performances, which disclose joint areas that contribute the classification consistently within each dataset [353]. The multi-centre classification concept was employed in several studies [354]–[358]. Furthermore, a cross-study could be implemented as well. In which a dataset is used for training and testing with another dataset.

Appendix A

State-of-the-Art DCNN Architectures

There are several state-of-the-art DCNN architectures; Figure A.1 shows a comparison between the operating parameters and the estimated accuracies achieved by the most relevant DCNN architectures. In this work five state-of-the-art DCNN architectures were used; AlexNet [23], GoogleNet [24], Inception-v3 [25], the ResNet [26], Inception ResNet-v2 [27], and MobileNet-v2 [28]. These architectures will be discussed in the following sections [359].

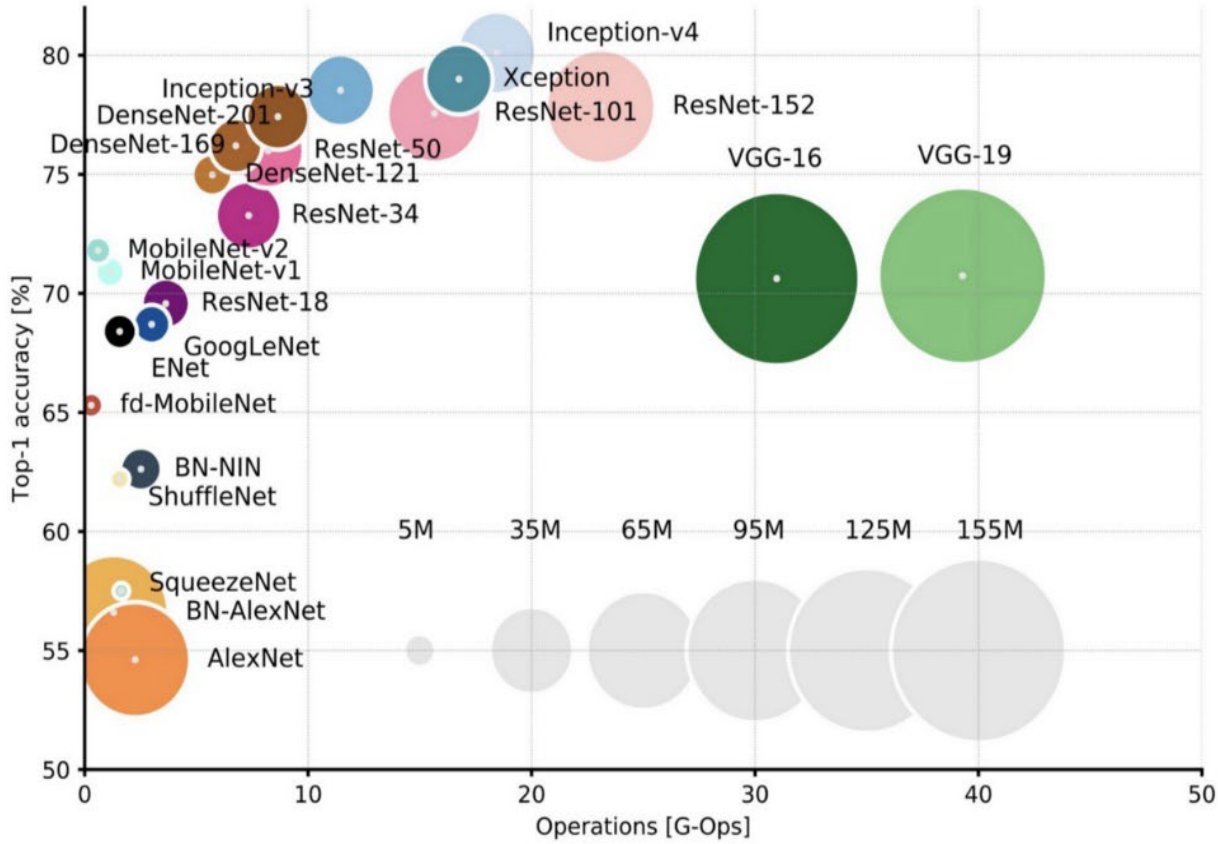


Figure A.1: A comparison between different types of DCNN architectures concerning accuracies. Adopted from [360].

A.1 AlexNet Architecture

The AlexNet architecture [23] achieved significantly improved performance over the other non-deep learning methods for ImageNet large scale visual recognition challenge (ILSVRC) 2012. This success has revived the interest in DCNNs in computer vision. AlexNet has five convolution layers, three pooling layers, and three fully-connected layers with approximately 60 million free parameters [23] as shown in Figure A.2.

A detailed layer description of AlexNet architecture is given in Table A.1. The output layer of the first convolutional layer in the AlexNet architecture is calculated using equation (3.2) in Chapter 3. The output equals 55 × 55 × 96, which demonstrates that the size of the feature map is 55 × 55 in width and height. In addition, the number of feature maps is 96 as clear in Table A.1.

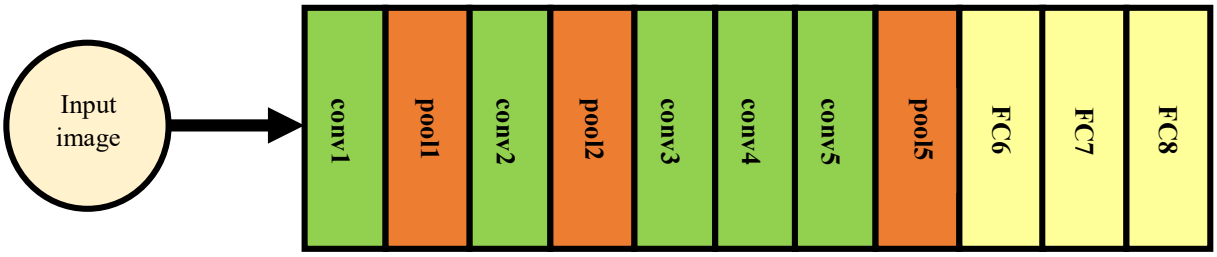


Figure A.2: The AlexNet DCNN architecture. Adapted from [23].

Table A.1: The layers of AlexNet DCNN architecture.

Layer Name	Description		Output Size
Input layer			227 × 227 × 3
conv1	**Filter Size**	11 × 11	55 × 55 × 96
	Stride	4	
	Padding	0	
pool1	**Pooling Size**	3 × 3	27 × 27 × 96
	Stride	2	
conv2	**Filter Size**	5 × 5	27 × 27 × 256
	Stride	1	
pool2	**Pooling Size**	3 × 3	13 × 13 × 256
	Stride	2	
conv3	**Filter Size**	3 × 3	13 × 13 × 384
	Stride	1	
conv4	**Filter Size**	3 × 3	13 × 13 × 384
	Stride	1	
conv5	**Filter Size**	3 × 3	13 × 13 × 256
	Stride	1	
pool5	**Pooling Size**	3 × 3	6 × 6 × 256
	Stride	2	
Fully connected (FC)			4096 × 2

A.2 GoogleNet Architecture

GoogleNet [24] model is significantly more complex and deep than all DCNN architectures. It was introduced by Szegedy C. et al. [24] who proposed a structure that computationally efficient. More importantly, it introduces a new module called “Inception”; therefore, it is named Inception-v1. It concatenates filters of different sizes and dimensions into a single new filter. The main difference between the Inception module and a conventional convolutional layer is that the Inception module allows the extraction of features with different kernel sizes. GoogleNet has two convolution layers, two pooling layers, nine Inception layers, and a fully connected layer before the output. Each Inception layer consists of six convolution layers and one pooling layer as shown in Figure A.3 [24]. The GoogleNet architecture and a detailed layer description for GoogleNet are illustrated in Figure 4.4 and Table A.2, respectively.

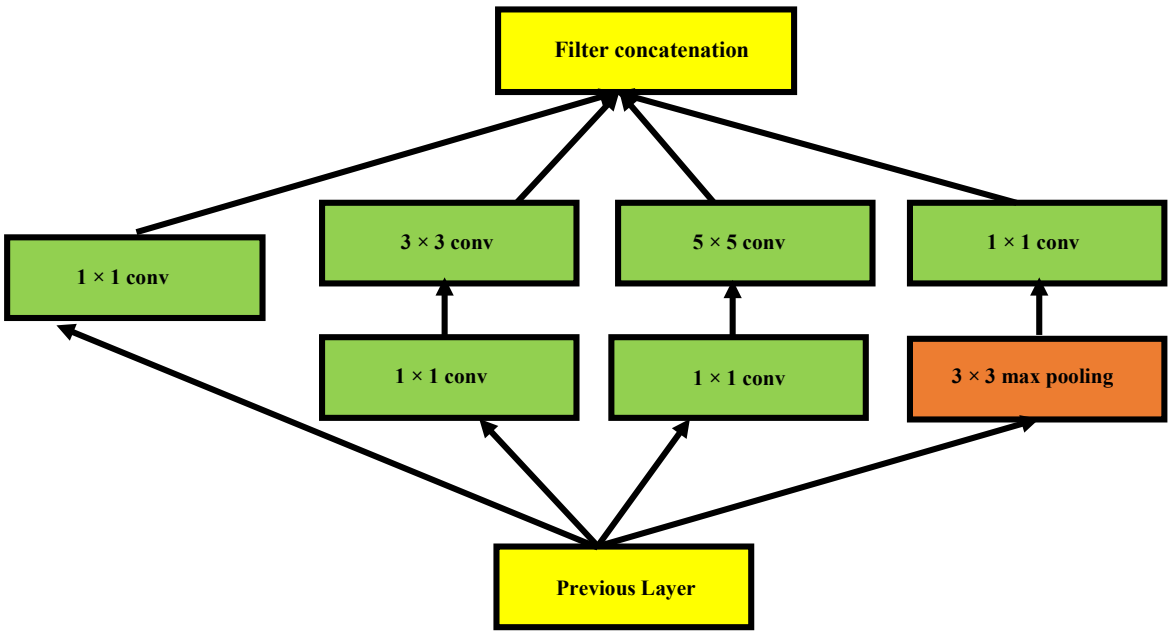


Figure A.3: The Inception module. Adapted from [24].

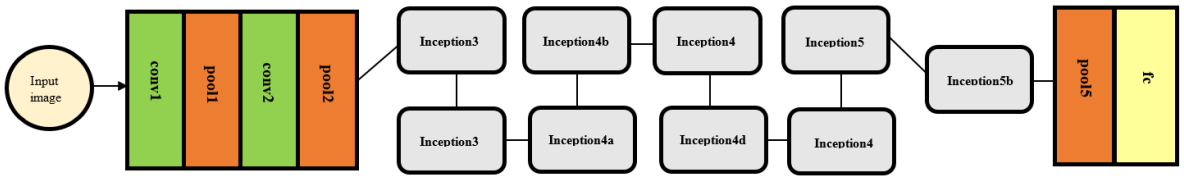


Figure A.4: The GoogleNet DCNN architecture. Adapted from [24].

Table A.2: The layers of GoogleNet DCNN architecture.

Layer Name	Filter Size	Stride	Output Size
Input Layer			224 × 224 × 3
conv1	7 × 7	2	112 × 112 × 64
pool1	3 × 3	2	56 × 56 × 64
conv2	3 × 3	1	56 × 56 × 192
pool2	3 × 3	2	28 × 28 × 192
Inception (3a)	-	-	28 × 28 × 256
Inception (3b)	-	-	28 × 28 × 480
pool3	3 × 3	2	14 × 14 × 480
Inception (4a)	-	-	14 × 14 × 512
Inception (4b)	-	-	14 × 14 × 512
Inception (4c)	-	-	14 × 14 × 512
Inception (4d)	-	-	14 × 14 × 528
Inception (4e)	-	-	14 × 14 × 832
pool4	3 × 3	2	7 × 7 × 832
Inception (5a)	-	-	7 × 7 × 832
Inception (5b)	-	-	7 × 7 × 1024
average pooling	7 × 7	1	1 × 1 × 1024
fully connected (FC)			1024 × 2

A.3 Inception-v3

Inception-v3 [25] is a widely used image recognition model that attains greater than 78.1% accuracy on the ImageNet dataset. It is an extended network of the popular GoogleNet network [24]. It accomplished a good classification performance in several biomedical applications using the transfer-learning technique. Inception-v3 proposed the inception module as well, which concatenates multiple different sized convolutional filters into a new filter. Such design decreases the number of parameters to be trained and thereby reduces the computational complexity [25], [361]. The motivation behind the Inception architecture is that it incorporates different orders of convolutional blocks [25].

Inception-v3 architecture is made-up of symmetric and asymmetric building blocks, including convolutions, average pooling, maximum pooling, con cats, dropouts, and fully connected layers. The first several layers of the Inception-v3 architecture consist of six convolutional layers with kernel sizes of 3 × 3 and an average-pooling layer with a kernel size of 3 × 3, followed by 5 Inception-A, 4 Inception-B, and 2 Inception-C modules as shown in Figure A.5. The Inception modules are illustrated in Figure A.6. The Inception-A module used in Inception-v3 is the same as the Inception module used in GoogleNet architecture, however, the 5 × 5 convolutions are factored to two 3 × 3 convolutions. A detailed layer description for Inception-v3 is given in Table A.3.

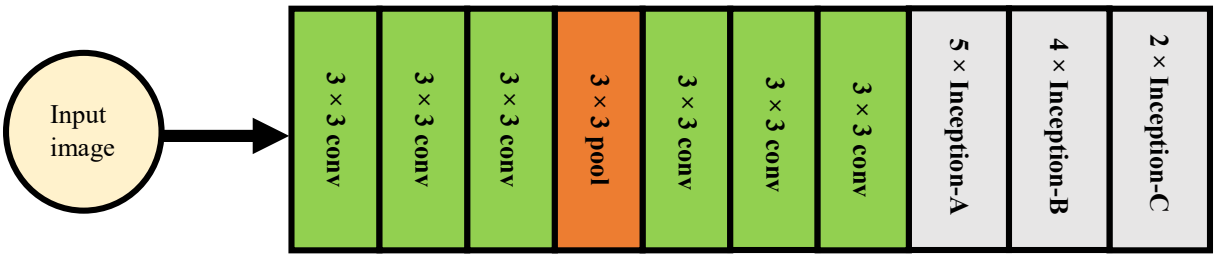
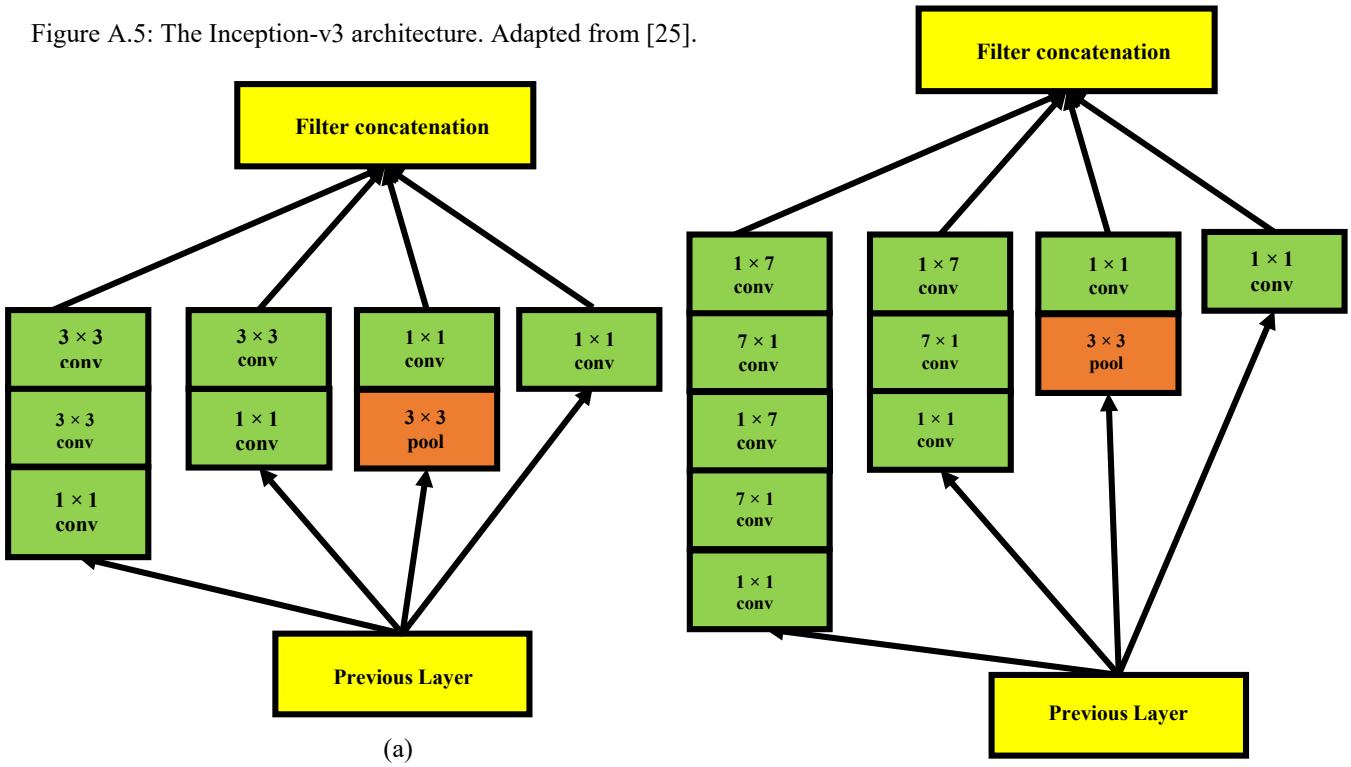


Figure A.5: The Inception-v3 architecture. Adapted from [25].



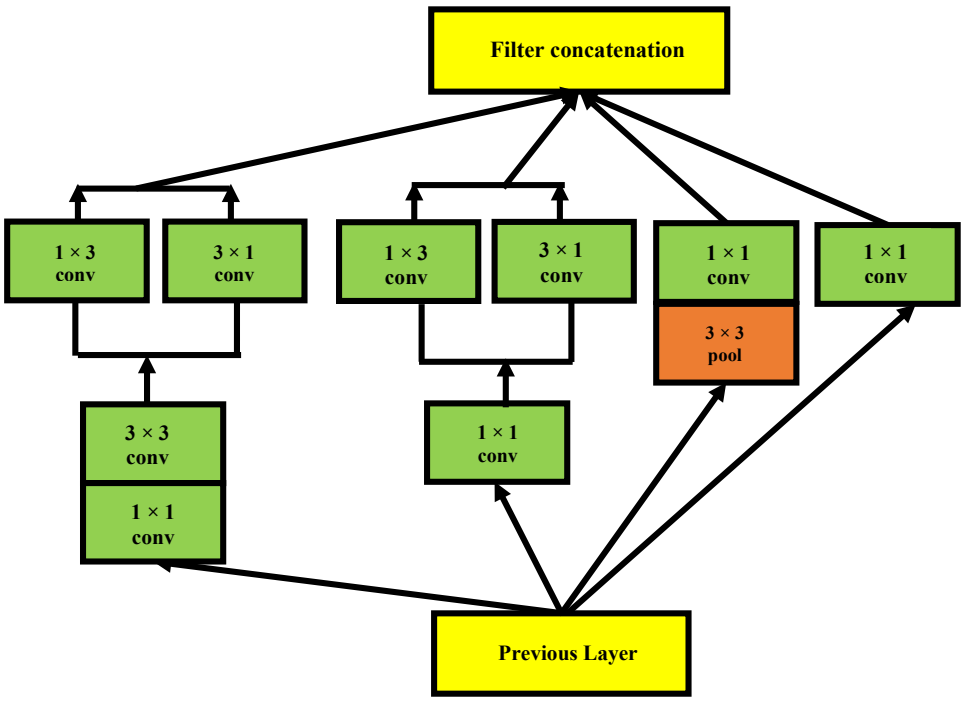


Figure A.6: The Inception modules used in Inception-v3 architecture (a) Inception-A module, (b) Inception-B module, and (c) Inception-C module. Adapted from [25].

Table A.3: The layers of Inception-v3 DCNN architecture.

Layer Name	Filter Size	Output Size
Input Layer		299 × 299 × 3
conv1	3 × 3	149 × 149 × 32
conv2	3 × 3	147 × 147 × 32
conv3	3 × 3	147 × 147 × 64
pool1	3 × 3	73 × 73 × 64
conv4	3 × 3	71 × 71 × 80
conv5	3 × 3	35 × 35 × 192
conv6	3 × 3	35 × 35 × 288
5 × Inception-A	-	17 × 17 × 768
4 × Inception-B	-	8 × 8 × 1280
2 × Inception-C	-	8 × 8 × 2048
fully connected (FC)		2048 × 2

A.4 ResNet Architecture

ResNet won the first place in ILSVRC and COCO 2015 competition in ImageNet Detection, ImageNet localization, Coco detection, and Coco segmentation [26]. The gradients of a deep convolution network have higher chances of fading with the increase of the number of network layers, which hardens the convergence of a network. Batch normalization is commonly used to solve this problem, however, its performance becomes flooded and then worsens quickly at the beginning of the network converge [362]. To overcome this, a deep residual learning algorithm was proposed by He et al. [26] who constructed a network called deep residual network (ResNet) [26]. This method adds shortcuts called residuals between layers of traditional DCNNs to bypass a few convolution layers at a time. It increases the number of deep layers as well to enhance its performance and employs the residual shortcuts to accelerate the convergence of these large numbers of deep layers. A ResNet has several stacks of residual blocks. Each block is made up of several stacked convolution layers. Every single convolution layer takes the output fields of the feature map of the previous layer as its input. The output of every residual block is added to its input through an associate identity mapping path [363]. ResNet has some common architectures such as ResNet-18, ResNet-50, and ResNet-101. A detailed layer description for ResNet-18, ResNet-50, and ResNet-101 architecture is illustrated in Table A.4.

Table A.4: The layers of ResNet DCNN architecture.

Layer Name	Output Size	ResNet-18	ResNet-50	ResNet-101
Input Layer		224 × 224 × 3		
conv1	112 × 112 × 64	Filter size = 7 × 7 Number of filters = 64 Stride = 2 Padding = 3		
pool1	56 × 56 × 64	Pool size = 3 × 3 Stride = 2		
conv2_x	56 × 56 × 64	$\begin{bmatrix} 3\times3, & 64 \\ 3\times3, & 64 \end{bmatrix}\times 2$	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times 3$
conv3_x	28 × 28 × 128	$\begin{bmatrix} 3\times3, & 128 \\ 3\times3, & 128 \end{bmatrix}\times 2$	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times 4$
conv4_x	14 × 14 × 256	$\begin{bmatrix} 3\times3, & 256 \\ 3\times3, & 256 \end{bmatrix}\times 2$	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times 6$	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times 23$
conv5_x	7 × 7 × 512	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix}\times 2$	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times 3$
Average pooling		Pool size = 7 × 7 Stride = 7		
		1 × 1 × 512	1 × 1 × 2048	1 × 1 × 2048
Fully connected (FC)		2 (512 × 2)	2 (2048 × 2)	2 (2048 × 2)

A.5 Inception ResNet-v2

Inception Resnet-v2 [27] is formulated based on a combination of the Inception module and the residual connection. In the Inception-ResNet block, multiple sized convolutional filters are combined with residual connections. The usage of residual connections not only avoids the degradation problem caused by deep structures, but also reduces the training time [27]. The Inception-ResNet-v2 has a computational cost similar to Inception-v4. Figure A.7 shows the basic network architecture of Inception Resnet-v2. Moreover, the block diagrams of the stem, Inception ResNet-A, Reduction-A, Inception ResNet-B, Reduction-B, and Inception ResNet-C are illustrated in Figure A.8, Figure A.9, Figure A.10, Figure A.11, Figure A.12, and Figure A.13, respectively.

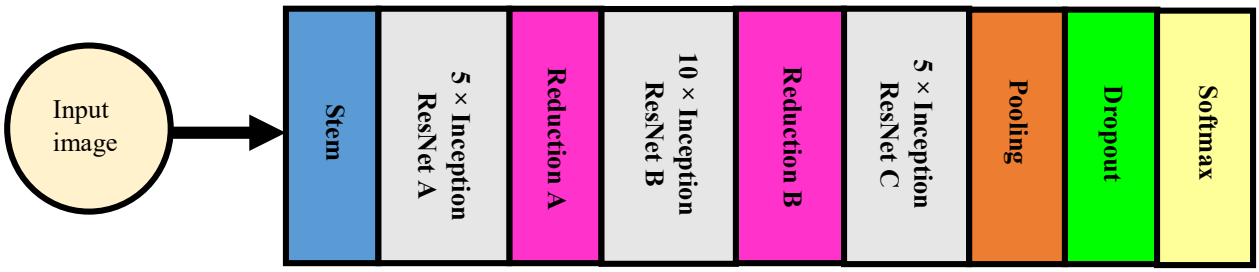


Figure A.7: The Inception ResNet-v2 architecture. Adapted from [27].

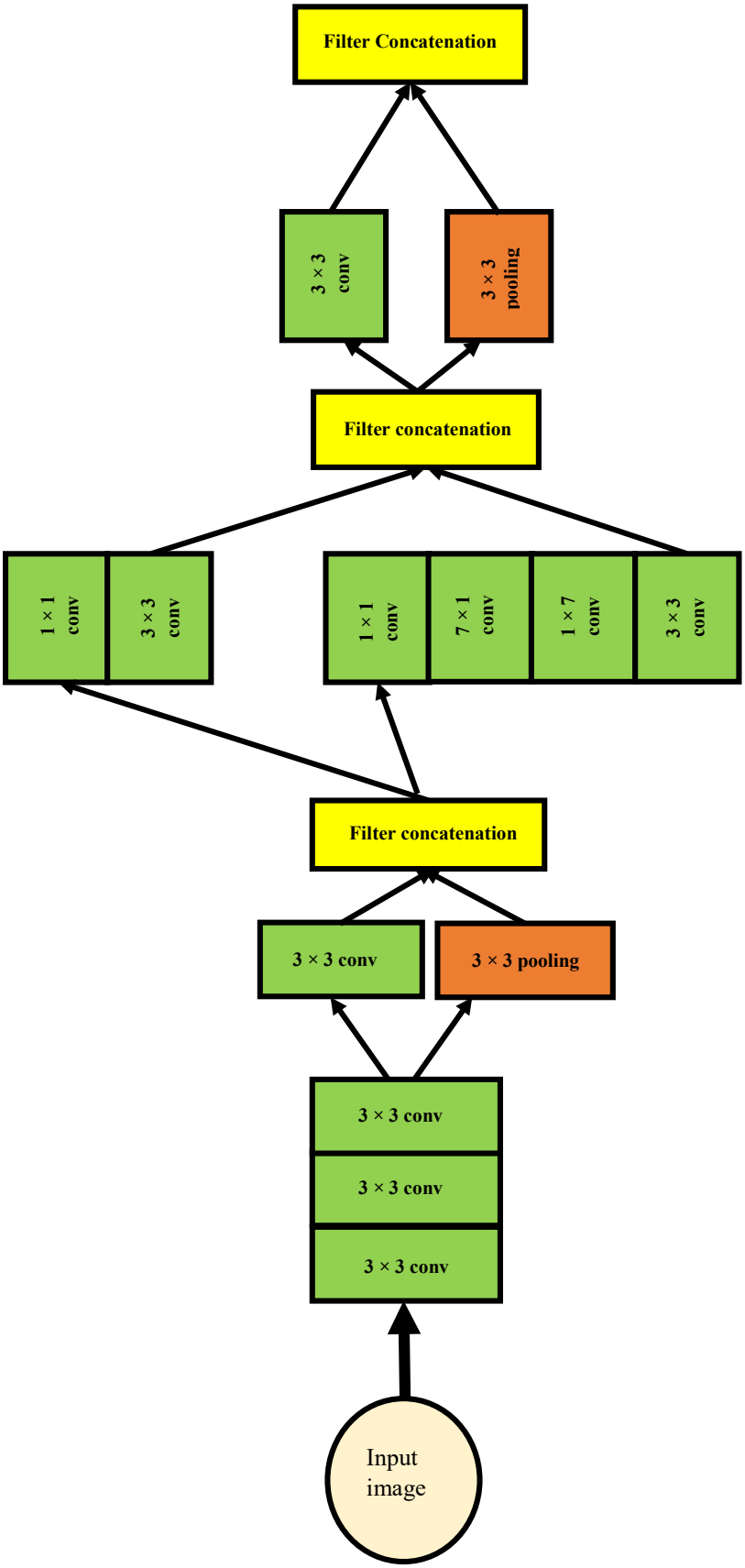


Figure A.8: The scheme for stem module for Inception ResNet-v2 architecture. Adapted from [27].

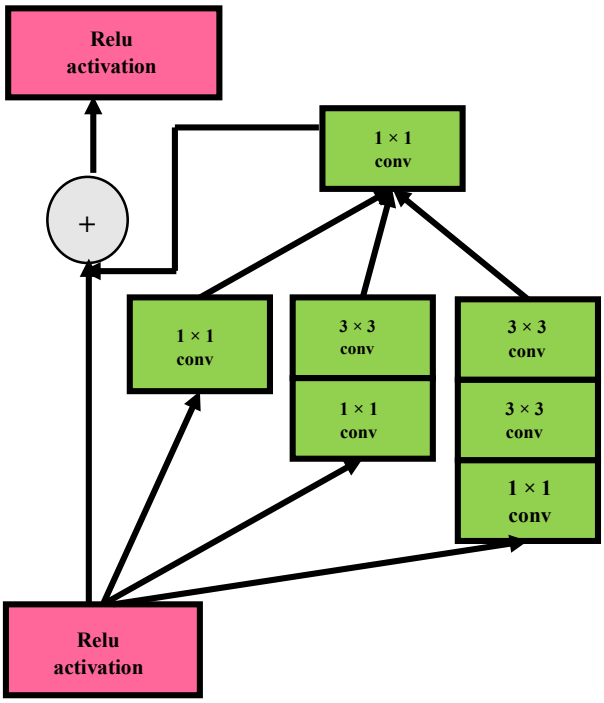


Figure A.9: The schema for the Inception ResNet-A module of Inception ResNet-v2 architecture. Adapted from [27].

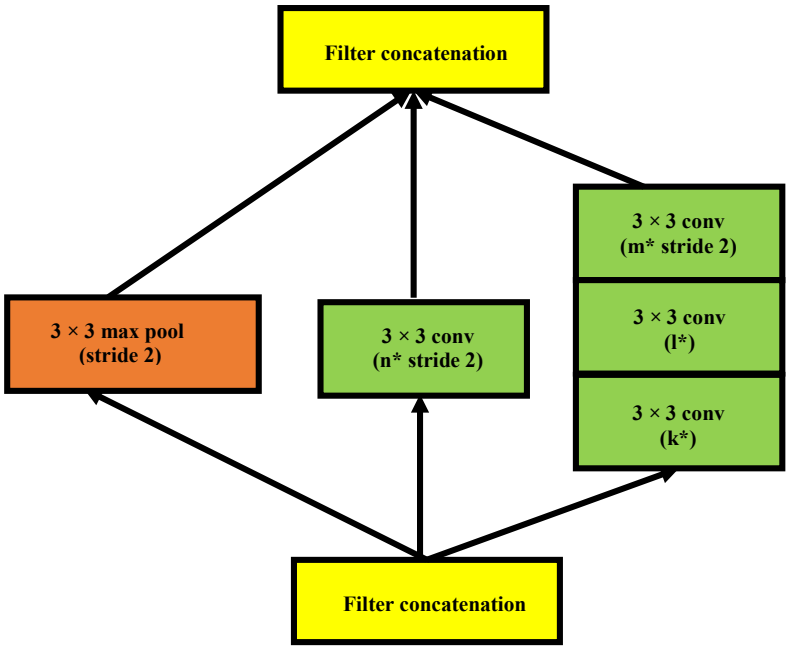


Figure A.10: The schema for the reduction-A module of Inception ResNet-v2 architecture. Adapted from [27].

*k, l, m, n numbers represent the filter bank sizes, which are equals to 256, 256, 384, and 384, respectively.

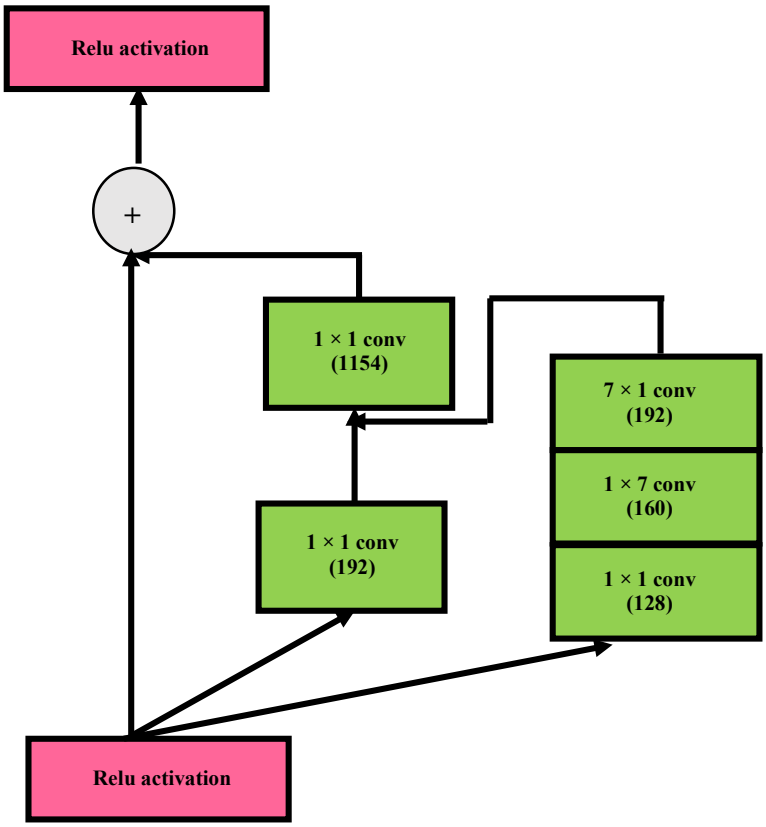


Figure A.11: The schema for the Inception ResNet-B module of Inception ResNet-v2 architecture. Adapted from [27].

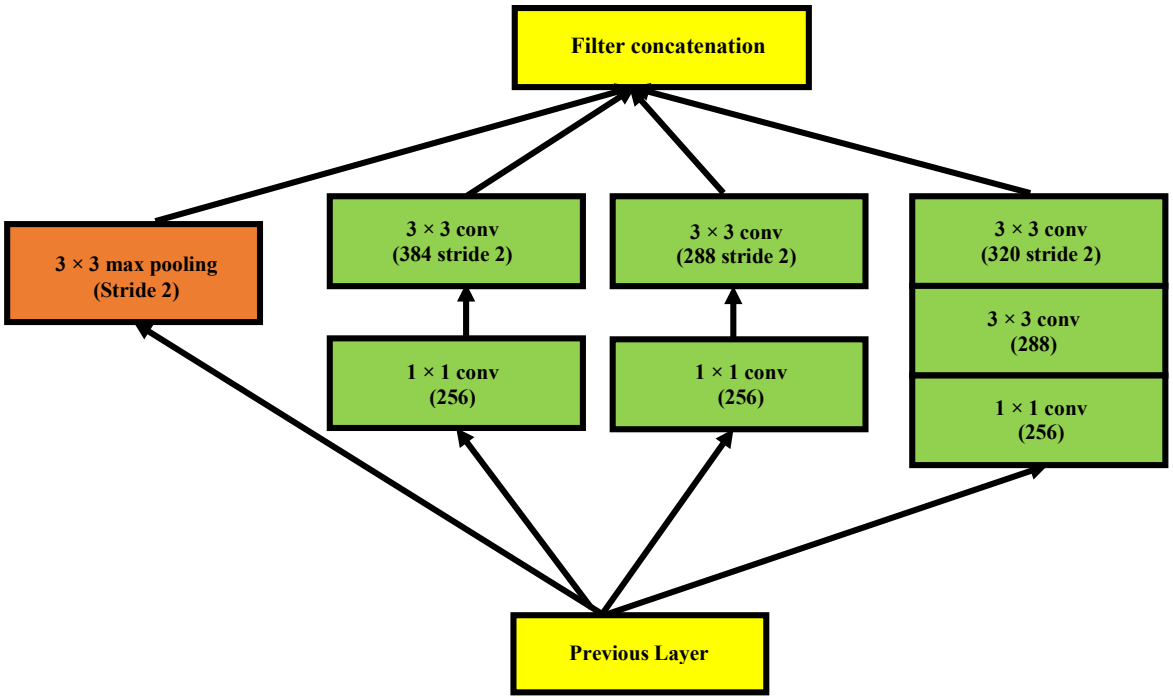


Figure A.12: The schema for the reduction-B module of Inception ResNet-v2 architecture. Adapted from [27].

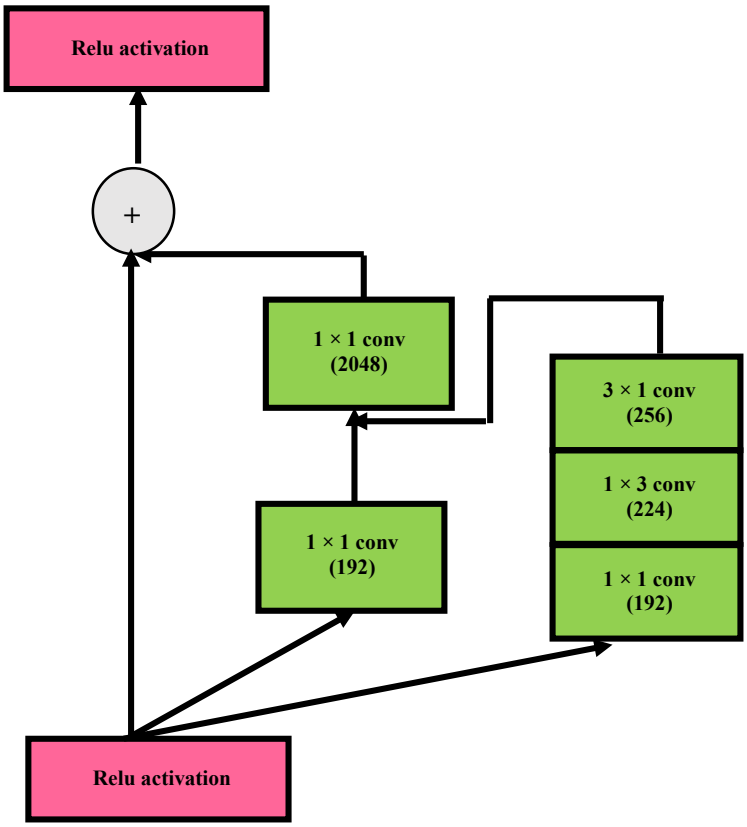


Figure A.13: The schema for the Inception ResNet-C module of Inception ResNet-v2 architecture. Adapted from [27].

A.6 MobileNet-v2

A group of researchers from Google released the MobileNet-v2 [28] architecture, which is optimized for mobile devices. The architecture delivers high accuracy results while keeping the parameters and mathematical operations as low as possible to bring deep neural networks to mobile devices [364]. MobileNet is small, has low-latency, low-power models parameterized to meet the resource constraints of a variety of use cases [365].

MobileNet-v1 is introduced to support classification, detection, embedding, and segmentation. Moreover, the new mobile architecture, MobileNet-v2 is the improved version of MobileNet-v1 [364]. However, MobileNet-v2 [28] architecture improves the state-of-the-art performance of mobile models on multiple tasks and benchmarks as well as across a spectrum of different model sizes. It is a very effective feature extractor for object detection and segmentation. The MobileNet-v2 models are much faster in comparison to MobileNet-v1. It uses two times fewer operations, has higher accuracy, and needs 30 percent less parameters. The architecture of MobileNet-v2 consists of a fully convolution layer with 32

filters, followed by 19 residual bottleneck layers [28]. A detailed layer description of MobileNet-v2 is illustrated in Table A.5.

Table A.5: The layers of MobileNet-v2 DCNN architecture.

Layer Name	Filter Size	Stride	Input Size
Input Layer			224 × 224 × 3
conv	3 × 3 × 3 × 32	2	112 × 112 × 32
conv dw	3 × 3 × 32 dw	1	112 × 112 × 32
conv	1 × 1 × 32 × 64	1	112 × 112 × 32
conv dw	3 × 3 × 64 dw	2	112 × 112 × 64
conv	1 × 1 × 64 × 128	1	56 × 56 × 64
conv dw	3 × 3 × 128 dw	1	56 × 56 × 128
conv	1 × 1 × 128 × 128	1	56 × 56 × 128
conv dw	3 × 3 × 128 dw	2	56 × 56 × 128
conv	1 × 1 × 128 × 256	1	28 × 28 × 128
conv dw	3 × 3 × 256 dw	1	28 × 28 × 256
conv	1 × 1 × 256 × 256	1	28 × 28 × 256
conv dw	3 × 3 × 256 dw	2	28 × 28 × 256
conv	1 × 1 × 256 × 512	1	14 × 14 × 256
5 × conv dw	3 × 3 × 512 dw	1	14 × 14 × 512
5 × conv	1 × 1 × 512 × 512	1	14 × 14 × 512
conv dw	3 × 3 × 512 dw	2	14 × 14 × 512
conv	1 × 1 × 512 × 1024	1	7 × 7 × 512
conv dw	3 × 3 × 1024 dw	2	7 × 7 × 1024
conv	1 × 1 × 1024 × 1024	1	7 × 7 × 1024
Avg. Pool	Pool 7 × 7	1	7 × 7 × 1024
FC	1024 × 1000	1	1 × 1 × 1024
Softmax	classifier	1	1 × 1 × 1000
			Output Size = 1280 × 2

References

[1] J. Ferlay *et al.*, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, 2015, doi: 10.1002/ijc.29210.

[2] A. D. Darwish, A. M. Helal, N. H. Aly El-din, L. L. Solaiman, and A. Amin, "Breast cancer in women aging 35 years old and younger: The Egyptian National Cancer Institute (NCI) experience," *Breast*, vol. 31, pp. 1–8, 2017, doi: 10.1016/j.breast.2016.09.018.

[3] S. N. Sree, E. Y.-K. Ng, R. U. Acharya, and O. Faust, "Breast imaging: A survey," *World J. Clin. Oncol.*, vol. 2, no. 4, pp. 171–178, 2011, doi: 10.5306/wjco.v2.i4.171.

[4] S. Iranmakani *et al.*, "A review of various modalities in breast imaging: Technical aspects and clinical outcomes," *Egypt. J. Radiol. Nucl. Med.*, vol. 51, no. 57, 2020, doi: https://doi.org/10.1186/s43055-020-00175-5.

[5] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms : A survey," *Pattern Recognit.*, vol. 36, no. 12, pp. 2967–2991, 2003, doi: 10.1016/S0031-3203(03)00192-4.

[6] J. Bozek, M. Mustra, K. Delac, and M. Grgic, "A survey of image processing algorithms in digital mammography," in *Recent Advances in Multimedia Signal Processing and Communications*, M. Grgic, K. Delac, and M. Ghanbari, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 631–657.

[7] Q. Li and R. M. Nishikawa, "Computer aided detection and diagnosis in medical imaging: A review of clinical and educational applications," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM '16*, 2016, pp. 1–425, doi: https://doi.org/10.1145/3012430.3012567.

[8] G. Chugh, S. Kumar, and N. Singh, "Survey on machine learning and deep learning applications in breast cancer diagnosis," *Cognit. Comput.*, no. 0123456789, 2021, doi: 10.1007/s12559-020-09813-6.

[9] S. Karamizadeh, S. M. Abdullah, M. Zamani, and A. Kherikhah, "Pattern recognition techniques: Studies on appropriate classifications," in *Advanced Computer and Communication Engineering Technology*, 2015, pp. 791–799.

[10] A. Bohr and K. Memarzadeh, "The rise of artificial intelligence in healthcare applications," *Artif. Intell. Healthc.*, pp. 25–60, 2020, doi: 10.1016/B978-0-12-818438-7.00002-2.

[11] Z. Ahmad, S. Rahim, M. Zubair, and J. Abdul-Ghafar, "Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosoph," *Diagn. Pathol.*, vol. 16, no. 1, pp. 1–16, 2021, doi: 10.1186/s13000-021-01085-4.

[12] D. F. Schneider, "Machine learning and artificial intelligence," in *Health Services Research*, J. B. Dimick and C. C. Lubitz, Eds. Cham: Springer International Publishing, 2020, pp. 155–168.

[13] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," *Cancer Lett.*, vol. 471, pp. 61–71, 2020, doi: 10.1016/j.canlet.2019.12.007.

[14] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement, "Survey on deep learning for radiotherapy," *Comput. Biol. Med.*, vol. 98, no. March, pp. 126–146, 2018, doi: 10.1016/j.compbiomed.2018.05.018.

[15] J. A. Nichols, H. W. Herbert Chan, and M. A. B. Baker, “Machine learning: Applications of artificial intelligence to imaging and diagnosis,” *Biophys. Rev.*, vol. 11, no. 1, pp. 111–118, 2019, doi: 10.1007/s12551-018-0449-9.

[16] A. Kumar, L. Bi, J. Kim, and D. D. Feng, *Machine learning in medical imaging*. Elsevier Inc., 2020.

[17] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding : A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.

[18] M. Inmaculada *et al.*, *Deep Learning in Healthcare*, vol. 171. Springer International Publishing, 2020.

[19] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, “Deep learning in medical image analysis,” in *Deep Learning in Medical Image Analysis : Challenges and Applications*, G. Lee and H. Fujita, Eds. Cham: Springer International Publishing, 2020, pp. 3–21.

[20] L. Saba *et al.*, “The present and future of deep learning in radiology,” *Eur. J. Radiol.*, vol. 114, pp. 14–24, 2019, doi: 10.1016/j.ejrad.2019.02.038.

[21] B. Sahiner *et al.*, “Deep learning in medical imaging and radiation therapy,” *Med. Phys.*, vol. 46, no. 1, pp. e1–e36, 2019, doi: 10.1002/mp.13264.

[22] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, “Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI,” *J. Magn. Reson. Imaging*, vol. 49, no. 4, pp. 939–954, 2019, doi: 10.1002/jmri.26534.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[24] C. Szegedy *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[25] C. Szegedy, V. Vanhoucke, J. Shlens, and Z. Wojna, “Rethinking the Inception architecture for computer vision,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016, doi: 10.1109/CVPR.2016.90.

[27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the impact of residual connections on learning,” in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI’17)*, 2017, pp. 4278–4284.

[28] M. Sandler, M. Zhu, A. Zhmoginov, and C. V Mar, “MobileNetV2: Inverted residuals and linear bottlenecks,” 2018, doi: 10.1109/CVPR.2018.00474.

[29] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[30] A. Khamparia *et al.*, “Diagnosis of breast cancer based on modern mammography using hybrid transfer learning,” *Multidimens. Syst. Signal Process.*, vol. 32, no. 2, pp. 747–765, 2021, doi: 10.1007/s11045-020-00756-7.

[31] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” in *Advances in Neural Information Processing Systems 27 (NIPS’14)*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3320–3328.

[32] Y. Li and H. Chen, “A Survey of Computer-aided Detection of Breast Cancer with Mammography,” *J. Heal. Med. Informatics*, vol. 7, no. 4, 2016, doi: 10.4172/2157-7420.1000238.

[33] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, no. e6201, 2019, doi: 10.7717/peerj.6201.

[34] D. A. Ragab, O. Attallah, M. Sharkas, J. Ren, and S. Marshall, "A Framework for Breast Cancer Classification using Multi-DCNNs," *Comput. Biol. Med.*, vol. 131, p. 104245, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104245.

[35] R. L. Siegel *et al.*, "Colorectal cancer statistics," *CA. Cancer J. Clin.*, vol. 67, no. 3, pp. 177–193, 2017, doi: 10.3322/caac.21395.

[36] "World Health Organization: International agency for research on cancer," 2021. https://gco.iarc.fr/ (accessed Jan. 04, 2021).

[37] S. S. Coughlin and D. U. Ekwueme, "Breast cancer as a global health concern," *Cancer Epidemiol.*, vol. 33, no. 5, pp. 315–318, 2009, doi: 10.1016/j.canep.2009.10.003.

[38] M. Silbermann, "Cancer care in countries and societies in transition: Individualized care in focus," *Cancer Care Ctries. Soc. Transit. Individ. Care Focus*, pp. 1–505, 2016, doi: 10.1007/978-3-319-22912-6.

[39] A. S. Ibrahim, H. M. Khaled, N. N. Mikhail, H. Baraka, and H. Kamel, "Cancer incidence in Egypt: Results of the national population-based cancer registry program," *J. Cancer Epidemiol.*, vol. 2014, 2014, doi: 10.1155/2014/437971.

[40] "Cancer." https://www.cancer.net/cancer-types/breast-cancer (accessed Jan. 25, 2020).

[41] Raquel Cruz da Conceição, "The development of ultra wideband scanning techniques for detection and classification of breast cancer," National University of Ireland Galway, Galway, Ireland, 2010.

[42] "Columbia University Herbert Irving Comprehensive Cancer Center - Breast Cancer," 2020. https://cancer.columbia.edu/breast-cancer (accessed Nov. 20, 2020).

[43] "National breast cancer foundation." http://www.nationalbreastcancer.org/about-breast-cancer (accessed Jan. 25, 2020).

[44] R. J. Lee, L. A. Vallow, S. A. Mclaughlin, K. S. Tzou, S. L. Hines, and J. L. Peterson, "Ductal carcinoma in situ of the breast," *Int. J. Surg. Oncol.*, vol. 2012, 2012, doi: 10.1155/2012/123549.

[45] L. Salvatorelli, L. Puzzo, G. M. Vecchio, R. Caltabiano, V. Virz, and G. Magro, "Ductal carcinoma in Situ of the breast: An update with emphasis on radiological and morphological features as predictive prognostic factors," *Cancers (Basel).*, vol. 12, no. 609, pp. 1–11, 2020.

[46] B. Cutuli *et al.*, "Lobular carcinoma in situ (LCIS) of the breast : is long-term outcome similar to ductal carcinoma in situ (DCIS)? Analysis of 200 cases," *Radiat. Oncol.*, vol. 10, pp. 1–7, 2015, doi: 10.1186/s13014-015-0379-7.

[47] "Web MD: Invasive Ductal Carcinoma (IDC) & Ductal Carcinoma In Situ (DCIS)," 2021. https://www.webmd.com/breast-cancer/ductal-carcinoma-invasive-in-situ#1-2 (accessed Jan. 18, 2021).

[48] "Breast cancer.org: Invasive Lobular Carcinoma (ILC)," 2021. https://www.breastcancer.org/symptoms/types/ilc (accessed Jan. 18, 2021).

[49] "American cancer society." https://www.cancer.org/cancer/breast-cancer/about.html (accessed Dec. 20, 2019).

[50] S. K. M Hamouda, R. H. Abo El Ezz, and M. E. Wahed, "Enhancement accuracy of breast tumor diagnosis in digital mammograms," *J. Biomed. Sci.*, vol. 06, no. 04, pp. 1–8, 2017, doi: 10.4172/2254-609x.100072.

[51] "National breast cancer foundation - Stages of breast cancer," 2019. http://www.nationalbreastcancer.org/breast-cancer-stages (accessed Feb. 27, 2020).

[52] “Mayo Clinic, breast cancer.” http://www.mayoclinic.org/diseases-conditions/breast-cancer/multimedia/stage-of-breast-cancer/sls-20076413?sl=?&slide=5 (accessed Feb. 27, 2020).

[53] S. K. Bandyopadhyay, “Survey on segmentation methods for locating masses in a mammogram image,” *Int. J. Comput. Appl.*, vol. 9, no. 11, pp. 25–28, 2010.

[54] “Johns Hopkins Medicine - Breast Center,” 2021. https://www.hopkinsmedicine.org/breast_center/treatments_services/breast_cancer_screening/digital_mammography/breast_imaging_reporting_data_system.html (accessed Jan. 05, 2021).

[55] M. Sampat, A. Bovik, G. Whitman, and M. Markey, “A model-based framework for the detection of spiculated masses on mammography,” *Med. Phys.*, vol. 35, no. 5, 2008, doi: doi: 10.1118/1.2890080.

[56] K. P. Hermann, “Applications and literature review of the BI-RADS classification,” *Eur Radiol*, vol. 15, pp. 1027–1036, 2005, doi: 10.1007/s00330-004-2593-9.

[57] J. Tang, S. Member, R. M. Rangayyan, J. Xu, and I. El Naqa, “Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

[58] T. Bhangale, U. B. Desai, and U. Shama, “An unsupervised scheme for detection of microcalcifications on mammograms,” in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, 2000, pp. 184–187, doi: 10.1109/ICIP.2000.900925.

[59] “National breast cancer foundation-Breast self-exam,” 2019. https://www.nationalbreastcancer.org/breast-self-exam (accessed Jan. 05, 2021).

[60] “Breast Cancer - Breast self exam,” 2021. https://www.breastcancer.org/symptoms/testing/types/self_exam (accessed Jan. 05, 2021).

[61] “Introduction to breast imaging.” https://undergradimaging.pressbooks.com/chapter/introduction-to-breast-imaging/ (accessed Feb. 27, 2020).

[62] J. Bozek, K. Delac, and M. Grgic, “Computer-aided detection and diagnosis of breast abnormalities in digital mammography,” 2008.

[63] E. P. V. Le, Y. Wang, Y. Huang, S. Hickman, and F. J. Gilbert, “Artificial intelligence in breast imaging,” *Clin. Radiol.*, vol. 74, no. 5, pp. 357–366, 2019, doi: 10.1016/j.crad.2019.02.006.

[64] Y. Yao, “Segmentation of breast cancer in mammograms and detection using magnetic resonance imaging,” Nanyang Technological University, Nanyang, Singapore, 2004.

[65] “CancerQuest.” https://www.cancerquest.org/patients/detection-and-diagnosis/magnetic-resonance-imaging-mri (accessed Jan. 12, 2020).

[66] T. B. Borchartt, A. Conci, R. C. F. Lima, R. Resmini, and A. Sanchez, “Breast thermography from an image processing viewpoint: A survey,” *Signal Processing*, vol. 93, no. 10, pp. 2785–2803, 2013, doi: 10.1016/j.sigpro.2012.08.012.

[67] S. Ekici and H. Jawzal, “Breast cancer diagnosis using thermography and convolutional neural networks,” *Med. Hypotheses*, vol. 137, no. 109542, 2020, doi: 10.1016/j.mehy.2019.109542.

[68] N. Lanisa, N. S. Cheok, and L. K. Wee, “Color morphology and segmentation of the breast thermography image,” 2015, doi: 10.1109/IECBES.2014.7047614.

[69] “Breast tomography.” http://www.deansilvermd.com/breast-thermography-early-breast-cancer-detection/ (accessed Feb. 27, 2020).

[70] “Breast cancer screening using PET scans.” http://breast-cancer.ca/pets-bcanc/ (accessed Feb. 27, 2020).

[71] F. Bénard and E. Turcotte, “Imaging in breast cancer: Single-photon computed tomography and positron-emission tomography,” *Breast Cancer Res.*, vol. 7, no. 4, pp. 153–62, 2005, doi: 10.1186/bcr1201.

[72] “Independent imaging - The Pros and Cons of PET/CT scans.” https://www.independentimaging.com/the-pros-and-cons-of-pet-ct-scans/ (accessed Jan. 12, 2020).

[73] M. Sollini *et al.*, “PET/CT radiomics in breast cancer: Mind the step,” *Methods*, vol. In press, 2020, doi: 10.1016/j.ymeth.2020.01.007.

[74] “Imaginis: Nuclear medicine imaging,” 2020. https://www.imaginis.com/nuclear-medicine/nuclear-medicine-imaging-nm (accessed Feb. 10, 2020).

[75] “Radiology info for patients.” https://www.radiologyinfo.org/en/info.cfm?pg=scintimammo (accessed Feb. 10, 2020).

[76] K. E. Lukong, “Understanding breast cancer – The long and winding road,” *BBA Clin.*, vol. 7, pp. 64–77, 2017, doi: 10.1016/j.bbacli.2017.01.001.

[77] M. Noble, W. Bruening, S. Uhl, and K. Schoelles, “Computer-aided detection mammography for breast cancer screening: Systematic review and meta-analysis,” *Arch. Gynecol. Obstet.*, vol. 279, no. 6, pp. 881–890, 2009, doi: 10.1007/s00404-008-0841-y.

[78] M. Meselhy Eltoukhy, I. Faye, and B. Belhaouari Samir, “A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation,” *Comput. Biol. Med.*, vol. 42, no. 1, pp. 123–128, 2012, doi: 10.1016/j.compbiomed.2011.10.016.

[79] M. L. Giger, H. Chan, and J. Boone, “Anniversary Paper: History and status of CAD and quantitative image analysis : The role of medical physics and AAPM,” *Med. Phys.*, vol. 35, no. 12, pp. 5799–5820, 2008, doi: 10.1118/1.3013555.

[80] M. Kashif, K. R. Malik, S. Jabbar, and J. Chaudhry, *Application of machine learning and image processing for detection of breast cancer*. Elsevier Inc., 2020.

[81] J. Suckling, J. Parker, D. Dance, S. Astley, and E. Al., “The mammographic image analysis society digital mammogram database,” *Exerpta Medica, Int. Congr. Ser.*, vol. 1069, pp. 375–378, 1994.

[82] R. M. and P. K. J. M. Heath, K. Bowyer, D. Kopans, “The digital database for screening mammography,” in *the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., Medical Physics Publishing, 2001.*, 2001, pp. 212–218, doi: ISBN 1-930524-00-5.

[83] “OPTIMAM mammography imaging,” 2019. https://medphys.royalsurrey.nhs.uk/omidb/ (accessed Oct. 29, 2020).

[84] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, “INbreast: Toward a full-field digital mammographic database,” *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, 2012, doi: 10.1016/j.acra.2011.09.014.

[85] “Breast cancer digital repository,” 2020. https://bcdr.eu/information/about (accessed Nov. 03, 2020).

[86] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, “A curated mammography data set for use in computer-aided detection and diagnosis research,” *Sci. Data*, vol. 4, no. 170177, pp. 1–9, 2017, doi: 10.1038/sdata.2017.177.

[87] “Digital Mammography Dream Challenge dataset.” https://www.synapse.org/#!Synapse:syn4224222/wiki/401743 (accessed Nov. 03, 2019).

[88] “The MIAS database of mammograms.” http://peipa.essex.ac.uk/info/mias.html (accessed Feb. 27, 2017).

[89] "Digital database for screening mammography.," 2001. http://www.eng.usf.edu/cvprg/Mammography/Database.html (accessed Nov. 03, 2015).

[90] S. G. Orel, N. Kay, C. Reynolds, and D. C. Sullivan, "BI-RADS categorization as a predictor of malignancy," *Radiology*, vol. 211, no. 3, pp. 845–850, 1999.

[91] "The INbreast dataset," 2010. http://medicalresearch.inescporto.pt/breastresearch/index.php/Get INbreast Database (accessed Nov. 26, 2016).

[92] "The curated breast imaging subset of DDSM." https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM (accessed Nov. 03, 2016).

[93] N. Houssami, C. I. Lee, D. S. M. Buist, and D. Tao, "Artificial intelligence for breast cancer screening: Opportunity or hype?," *Breast*, vol. 36, pp. 31–33, 2017, doi: 10.1016/j.breast.2017.09.003.

[94] P. Taylor, J. Champness, K. Johnston, and H. Potts, "Impact of computer-aided detection prompts on screening mammography," 2005. doi: 10.3310/hta9060.

[95] J. J. Fenton *et al.*, "Effectiveness of computer-aided detection in community mammography practice," *J. Natl. cancer Inst.*, vol. 103, no. 15, pp. 1152–1161, 2011, doi: 10.1093/jnci/djr206.

[96] V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A. J. Frangos, and J. H. Sunshine, "How widely is computer-aided detection used in screening and diagnostic mammography?," *JACR*, vol. 7, no. 10, pp. 802–805, 2010, doi: 10.1016/j.jacr.2010.05.019.

[97] M. Sonka, V. Hlavac, and R. Boyle, "Image pre-processing," in *Image Processing, Analysis and Machine Vision*, Boston, MA: Springer US, 1993, pp. 56–111.

[98] D. A. V. Dyk and X. L. Meng, "The art of data augmentation," *J. Comput. Graph. Stat.*, vol. 10, no. 1, pp. 1–50, 2001, doi: 10.1198/10618600152418584.

[99] A. Namdeo and S. S. Bhadoriya, "A review on image enhancement techniques with its advantages and disadvantages," *Int. J. Sci. Res. Technol.*, vol. 2, no. 5, pp. 171–182, 2016.

[100] S. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Trans. Med. Imaging*, vol. 8, no. 4, pp. 377–386, 1989.

[101] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognit.*, vol. 39, pp. 646–668, 2006, doi: 10.1016/j.patcog.2005.07.006.

[102] K. Wongsritong, K. Kittayaruasiriwat, F. Cheevasuvit, K. Dejhan, and A. Somboonkaew, "Contrast enhancement using multipeak histogram equalization with brightness preserving," in *IEEE Asia-Pacific Conference on Circuits and Systems. Microelectronics and Integrating Systems. Proceedings, APCCAS (Cat. No.98EX242)*, 1998, pp. 455–458, doi: 10.1109/apccas.1998.743808.

[103] H. D. Cheng and X. J. Shi, "A simple and effective histogram equalization approach to image enhancement," *Digit. Signal Process.*, vol. 14, pp. 158–170, 2004, doi: 10.1016/j.dsp.2003.07.002.

[104] J. Zabalza *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, 2015, doi: 10.1109/TGRS.2015.2398468.

[105] S. M.Pizer *et al.*, "Adaptive histogram equalization and its variations," *Comput. vision, Graph. Image Process.*, vol. 39, pp. 355–368, 1987.

[106] T. Qiao *et al.*, "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 119–133, 2017, doi: 10.1109/TGRS.2016.2598065.

[107] E. D. Pisano *et al.*, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *J. Digit. Imaging*, vol. 11, no. 4, pp. 193–200, 1998.

[108] A. and H. S. Sahakyan, "Segmentation of the breast region in digital mammograms and detection of masses," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 2, pp. 102–105, 2012.

[109] A. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 5, 1995.

[110] S. Yu and S. Brown, "Enhancement and identification of microcalcifications in mammogram images using wavelets," in *IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*, 1996, pp. 1166–1171, doi: 10.1109/ICSMC.1996.571251.

[111] H. Inoue, "Data augmentation by pairing samples for images classification," *arXiv Prepr. arXiv1801.02929*, 2018, [Online]. Available: http://arxiv.org/abs/1801.02929.

[112] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential data augmentation techniques for medical imaging classification tasks," in *AMIA 2017, American Medical Informatics Association Annual Symposium*, 2017, pp. 979–984.

[113] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 60, 2019, doi: 10.1186/s40537-019-0197-0.

[114] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?," 2016, doi: 10.1109/DICTA.2016.7797091.

[115] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv Prepr. arXiv1712.04621*, 2017, [Online]. Available: http://arxiv.org/abs/1712.04621.

[116] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation," *Expert Syst. Appl.*, vol. 95, pp. 43–56, 2018, doi: 10.1016/j.eswa.2017.11.028.

[117] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[118] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017, doi: 10.1109/ACCESS.2017.2696121.

[119] D. Kaur and Y. Kaur, "Various image segmentation techniques: A review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 5, pp. 809–814, 2014.

[120] R. Rouhi and M. Jafari, "Classification of benign and malignant breast tumors based on hybrid level set segmentation," *Expert Syst. Appl.*, vol. 46, pp. 45–59, 2016, doi: 10.1016/j.eswa.2015.10.011.

[121] D. Brazakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," *IEEE Trans. Med. Imaging*, vol. 9, no. 3, 1990.

[122] W. Khan, "Image segmentation techniques: A survey," *J. Image Graph.*, vol. 1, no. 4, pp. 166–170, 2013, doi: 10.12720/joig.1.4.166-170.

[123] M. Kaur and P. Goyal, "A review on region based segmentation," *Int. J. Sci. Res.*, vol. 4, no. 4, pp. 3194–3197, 2015.

[124] Y. Cao, X. Hao, X. Zhu, and S. Xia, "An adaptive region growing algorithm for breast masses in mammograms," *Front. Electr. Electron. Eng. China*, vol. 5, no. 2, pp. 128–136, 2010, doi: 10.1007/s11460-010-0017-y.

[125] R. Muthukrishnan and M. Radha, “Edge detection techniques for image segmentation,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 6, 2011, doi: 10.5121/ijcsit.

[126] R. Yogamangalam and B. Karthikeyan, “Segmentation Techniques Comparison in Image Processing,” vol. 5, no. 1, pp. 307–313, 2013.

[127] H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, “Automated breast cancer detection and classification using ultrasound images : A survey,” *Pattern Recognit.*, vol. 43, no. 1, pp. 299–317, 2010, doi: 10.1016/j.patcog.2009.05.012.

[128] A. Meyer and V. Schmid, “Feature selection and extraction,” in *Handbook of pattern recognition and image processing*, 2014, pp. 21–69.

[129] E. Fazeldehkordi and et al., “Chapter 4 Feature extraction,” in *A Machine-Learning Approach to Phishing Detection and Defense*, 2015, pp. 45–54.

[130] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[131] Z. Zhu, E. Albadawy, A. Saha, J. Zhang, M. R. Harowicz, and M. A. Mazurowski, “Deep learning for identifying radiogenomic associations in breast cancer,” *Comput. Biol. Med.*, vol. 109, no. March, pp. 85–90, 2019, doi: 10.1016/j.compbiomed.2019.04.018.

[132] I. Sechopoulos and R. M. Mann, “Stand-alone artificial intelligence - The future of breast cancer screening?,” *Breast*, vol. 49, pp. 254–260, 2020, doi: 10.1016/j.breast.2019.12.014.

[133] E. D. Pisano, “AI shows promise for breast cancer screening,” *Nature Magazine*, pp. 35–36, 2020.

[134] M. Jyotiyana and N. Kesswani, “Deep learning and the future of biomedical image analysis,” in *Deep Learning Techniques for Biomedical and Health Informatics*, S. Dash, B. R. Acharya, M. Mittal, A. Abraham, and A. Kelemen, Eds. Cham: Springer International Publishing, 2020, pp. 329–345.

[135] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.

[136] Y. Hu, S. Luo, L. Han, L. Pan, and T. Zhang, “Deep supervised learning with mixture of neural networks,” *Artif. Intell. Med.*, vol. 102, no. 101764, 2020, doi: 10.1016/j.artmed.2019.101764.

[137] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[138] A. B. Levine, C. Schlosser, J. Grewal, R. Coope, S. J. M. Jones, and S. Yip, “Rise of the machines: Advances in deep learning for cancer diagnosis,” *Trends in Cancer*, vol. 5, no. 3, pp. 157–169, 2019, doi: 10.1016/j.trecan.2019.02.002.

[139] T. Pang, J. Hsiu, D. Wong, W. Lin, and C. Seng, “Deep learning radiomics in breast cancer with different modalities: Overview and future,” *Expert Syst. Appl.*, vol. 158, no. 113501, 2020, doi: 10.1016/j.eswa.2020.113501.

[140] A. Ibrahim *et al.*, “Artificial intelligence in digital breast pathology: Techniques and applications,” *Breast*, vol. 49, pp. 267–273, 2020, doi: 10.1016/j.breast.2019.12.007.

[141] M. Coccia, “Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence,” *Technol. Soc.*, vol. 60, no. 101198, 2020, doi: 10.1016/j.techsoc.2019.101198.

[142] R. Zemouri, N. Zerhouni, and D. Racoceanu, “Deep learning in the biomedical applications: Recent and future status,” *Appl. Sci.*, vol. 9, no. 8, pp. 1–40, 2019, doi: 10.3390/app9081526.

[143] M. Avci, M. Sarıgül, and B. M. Ozyildirim, “Case study: Deep convolutional networks in healthcare,” in *Studies in Computational Intelligence*, vol. 867, Springer International Publishing, 2020, pp. 61–89.

[144] A. Das and M. N. Mohanty, “A review of breast cancer detection using deep learning techniques,” *Adv. Deep Learn. Med. Image Anal.*, no. 0123456789, pp. 69–85, 2022, doi: 10.1201/9781003230540-5.

[145] W. Zhu, L. Xie, J. Han, and X. Guo, “The application of deep learning in cancer prognosis prediction,” *Cancers (Basel).*, vol. 12, no. 3, pp. 1–19, 2020, doi: 10.3390/cancers12030603.

[146] H. P. Chan, L. M. Hadjiiski, and R. K. Samala, “Computer-aided diagnosis in the era of deep learning,” *Med. Phys.*, vol. 47, no. 5, pp. e218–e227, 2020, doi: 10.1002/mp.13764.

[147] M. F. Mridha *et al.*, “A comprehensive survey on deep-learning-based breast cancer diagnosis,” *Cancers (Basel).*, vol. 13, no. 6116, p. S1, 2021, doi: https://doi.org/10.3390/ cancers13236116.

[148] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010, pp. 253–256, doi: 10.1109/ISCAS.2010.5537907.

[149] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolution neural networks,” in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 2560–2567, doi: 10.1109/IJCNN.2016.7727519.

[150] S. Suzuki *et al.*, “Mass Detection Using Deep Convolutional Neural Network for Mammographic Computer-Aided Diagnosis,” in *Proceedings of the SICE Annual Conference 2016 Tsukuba, Japan*, 2016, pp. 1382–1386.

[151] J. Xu, X. Sun, Z. Zhang, G. Zhao, and J. Lin, “Understanding and improving layer normalization,” in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019, pp. 1–11.

[152] A. F. M. Agarap, “Deep learning using rectified linear units (ReLU).” pp. 2–8, 2019, [Online]. Available: https://arxiv.org/abs/1803.08375.

[153] C. E. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation functions: Comparison of trends in practice and research for deep learning.” pp. 1–20, 2018, [Online]. Available: https://arxiv.org/abs/1811.03378.

[154] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255, doi: 10.1109/CVPRW.2009.5206848.

[155] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images,” University of Toronto, 2009.

[156] H. Roth *et al.*, “Improving computer-aided detection using convolutional neural networks and random view aggregation,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016, doi: 10.1109/TMI.2015.2482920.

[157] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv Prepr. arXiv1409.1556*, pp. 1–10, 2014, doi: 10.1016/j.infsof.2008.09.005.

[158] L. Alzubaidi *et al.*, “Towards a better understanding of transfer learning for medical imaging: A case study,” *Appl. Sci.*, vol. 10, no. 13, pp. 1–21, 2020, doi: 10.3390/app10134523.

[159] B. Q. Huynh, H. Li, and M. L. Giger, “Digital mammographic tumor networks from deep convolutional neural classification using transfer learning,” *J. Med. Imaging*, vol. 3, no. 3, pp. 034501–1 to 034501–5, 2016, doi: 10.1117/1.JMI.3.3.034501.

[160] F. Gao, H. Yoon, T. Wu, and X. Chu, “A feature transfer enabled multi-task deep learning model on medical imaging,” *Expert Syst. Appl.*, vol. 143, p. 112957, 2020, doi: 10.1016/j.eswa.2019.112957.

[161] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, “From generic to specific deep representations for visual recognition,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 36–45, doi: 10.1109/CVPRW.2015.7301270.

[162] N. Tajbakhsh *et al.*, “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016, doi: 10.1109/TMI.2016.2535302.

[163] F. Jiang, H. Liu, S. Yu, and Y. Xie, “Breast mass lesion classification in mammograms by transfer learning,” in *Proceedings of the 5th International Conference on Bioinformatics and Computational ICBCB ’17*, 2017, pp. 59–62, doi: http://dx.doi.org/10.1145/3035012.3035022.

[164] R. K. Samala, H. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, “Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets,” *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 686–696, 2019.

[165] H. Chougrad, H. Zouaki, and O. Alheyane, “Multi-label transfer learning for the early diagnosis of breast cancer,” *Neurocomputing*, vol. 392, pp. 168–180, 2020, doi: 10.1016/j.neucom.2019.01.112.

[166] J. M. Keller, M. R. Gray, and J. A. Givens, “A fuzzy k-nearest neighbor algorithm,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, no. 4, pp. 580–585, 1985, doi: 10.1109/TSMC.1985.6313426.

[167] D. J. Lowsky *et al.*, “A k-nearest neighbors survival probability prediction method,” *Stat. Med.*, vol. 32, no. 12, pp. 2062–2069, 2013, doi: 10.1002/sim.5673.

[168] L. M. Bruce and R. R. Adhami, “Classifying mammographic mass shapes using the wavelet transform modulus-maxima method,” *IEEE Trans. Med. Imaging*, vol. 18, no. 12, pp. 1170–1177, 1999, doi: 10.1109/42.819326.

[169] B. Sahiner *et al.*, “Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization,” *IEEE Trans. Med. Imaging*, vol. 20, no. 12, pp. 1275–1284, 2001.

[170] A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman, “Diagnosis of ovarian cancer using decision tree classification of mass spectral data,” *J. Biomed. Biotechnol.*, vol. 2003, no. 5, pp. 308–314, 2003, doi: 10.1155/S1110724303210032.

[171] S. Aruna, S. P. Rajagopalan, and L. V Nandakishore, “An empirical comaprison of supervised learning algorithms in disease detection,” *Int. J. Inf. Technol. Converg. Serv.*, vol. 1, no. 4, pp. 81–92, 2011, doi: 10.5121/ijitcs.2011.1408.

[172] I. Christoyianni, E. Dermatas, and G. Kokkinakis, “Fast detection of masses in computer-aided mammography,” *IEEE Signal Process. Mag.*, vol. 17, no. 1, pp. 54–64, 2000.

[173] B. Sahiner *et al.*, “Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images,” *IEEE Trans. Med. Imaging*, vol. 15, no. 5, pp. 598–610, 1996.

[174] Z. Huo, M. L. Giger, and C. J. Vyborny, “Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis,” *IEEE Trans. Med. Imaging*, vol. 20, no. 12, pp. 1285–1292, 2001.

[175] I. El-naqa *et al.*, “A support vector machine approach for detection of microcalcifications,” *IEEE Trans. Med. Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[176] V. Vapnik, *Statistical learning theory*. New York: John Wiley & Sons Inc., 1998.

[177] Y. Huang and D. Chen, “Support vector machines in sonography application to decision making in the diagnosis of breast cancer,” *J. Clin. Imaging*, vol. 29, no. 3, pp. 179–184, 2005, doi: 10.1016/j.clinimag.2004.08.002.

[178] “Support vector machines.” http://www.dtreg.com/svm.htm (accessed Sep. 15, 2019).

[179] S. R. Gunn, “Support vector machines for classification and regression,” University of Southampton, 1998.

[180] N. Sharma, “Understanding the mathematics behind support vector machines,” 2020. https://heartbeat.fritz.ai/understanding-the-mathematics-behind-support-vector-machines-5e20243d64d5 (accessed Oct. 30, 2020).

[181] D. Martin and W. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation,” *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011, doi: 10.9735/2229-3981.

[182] B. Ramsay, S. Visa, A. Ralescu, and E. van der Knapp, “Confusion matrix-based feature selection,” 2011.

[183] K. Woods and K. W. Bowyer, “Generating ROC curves for artificial neural networks,” *IEEE Trans. Med. Imaging*, vol. 16, no. 3, pp. 329–337, 1997, doi: 10.1109/42.585767.

[184] A. Guillén, I. Rojas, J. González, and H. Pomares, “Improving Performance of Multi-objective Genetic for Function Approximation through island specialisation,” *AI 2006 Adv. Artif. Intell.*, vol. 4304, no. 1, 2006, doi: 10.1007/11941439.

[185] S. Wu, P. Flach, P. Flach, and B. Ac, “A scored AUC metric for classifier evaluation and selection,” 2005.

[186] A. Guillén, I. Rojas, J. González, and H. Pomares, “Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation,” 2006, doi: 10.1007/11941439.

[187] D. Berrar, “Cross-validation,” in *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1, Elsevier, 2018, pp. 542–545.

[188] R. Lowry, *Concepts and applications of inferential statistics*. 2014.

[189] S. Punitha, S. Ravi, M. Anousouya Devi, and J. Vaishnavi, “Breast cancer detection using classification techniques in digital mammography,” *Int. J. Control Theory Appl.*, vol. 9, no. 7, pp. 3123–3134, 2016.

[190] S. Jamal, S. Gardezi, A. Elazab, B. Lei, and T. Wang, “Breast cancer detection and diagnosis using mammographic data: Systematic review,” *J. Med. Interent Res.*, vol. 21, no. 7, 2019, doi: 10.2196/14464.

[191] T. Alsuliman, D. Humaidan, and L. Sliman, “Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality?,” *Curr. Res. Transl. Med.*, vol. In press, 2020, doi: 10.1016/j.retram.2020.01.002.

[192] Z. Rezaei, “A review on image-based approaches for breast cancer detection, segmentation, and classification,” *Expert Syst. Appl.*, vol. 182, no. April 2020, p. 115204, 2021, doi: 10.1016/j.eswa.2021.115204.

[193] N. I. R. Yassin, S. Omran, E. M. F. El Houby, and H. Allam, “Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review,” *Comput. Methods Programs Biomed.*, vol. 156, pp. 25–45, 2018, doi: 10.1016/j.cmpb.2017.12.012.

[194] U. Bottigli *et al.*, “CADe tools for early detection of breast cancer,” 2003, [Online]. Available: http://arxiv.org/abs/physics/0410082.

[195] L. Arbach, L. D. Bennett, J. M. Reinhardt, and G. Fallouh, “Classification of mammographic masses: Comparison between Backpropagation Neural Network (BNN) and human readers,” 2003, doi: 10.1117/12.480864.

[196] C. J. L, V. Ponomaryov, and J. L. S. R, “Detection of microcalcifications in digital mammograms images using wavelet transform,” 2006.

[197] B. Verma, P. McLeod, and A. Klevansky, “Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer,” *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3344–3351, 2010, doi: 10.1016/j.eswa.2009.10.016.

[198] J. Jiang, P. Trundle, and J. Ren, “Medical image analysis with artificial neural networks,” *Comput. Med. Imaging Graph.*, vol. 34, no. 8, pp. 617–631, 2010, doi: 10.1016/j.compmedimag.2010.07.003.

[199] J. Ren, D. Wang, and J. Jiang, “Effective recognition of MCCs in mammograms using an improved neural classifier,” *Eng. Appl. Artif. Intell.*, vol. 24, no. 4, pp. 638–645, 2011, doi: 10.1016/j.engappai.2011.02.011.

[200] J. Ren, “ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging,” *Knowledge-Based Syst.*, vol. 26, pp. 144–153, 2012, doi: 10.1016/j.knosys.2011.07.016.

[201] R. Pereira, M. Zanchetta, and D. Cesar, “Texture extraction : An evaluation of ridgelet , wavelet and co-occurrence based methods applied to mammograms,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11036–11047, 2012, doi: 10.1016/j.eswa.2012.03.020.

[202] C. Dromain, B. Boyer, R. Ferré, S. Canale, S. Delaloge, and C. Balleyguier, “Computed-aided diagnosis (CAD) in the detection of breast cancer,” *Eur. J. Radiol.*, vol. 82, no. 3, pp. 417–423, 2013, doi: 10.1016/j.ejrad.2012.03.005.

[203] A. Vadivel and B. Surendiran, “A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories,” *Comput. Biol. Med.*, vol. 43, no. 4, pp. 259–267, 2013, doi: 10.1016/j.compbiomed.2013.01.004.

[204] Y. Ben Youssef, E. Abdelmoneim, J. Zbitou, and A. Belaguid, “Statistical features and classification of normal and abnormal mammograms,” 2014, doi: 10.1109/ICMCS.2014.6911225.

[205] P. Agrawal, M. Vatsa, and R. Singh, “Saliency based mass detection from screening mammograms,” *Signal Processing*, vol. 99, pp. 29–47, 2014, doi: 10.1016/j.sigpro.2013.12.010.

[206] Z. Wang, G. Yu, Y. Kang, Y. Zhao, and Q. Qu, “Breast tumor detection in digital mammography based on extreme learning machine,” *Neurocomputing*, vol. 128, pp. 175–184, 2014, doi: 10.1016/j.neucom.2013.05.053.

[207] S. Ergin and O. Kilinc, “A new feature extraction framework based on wavelets for breast cancer diagnosis,” *Comput. Biol. Med.*, vol. 51, pp. 171–182, 2014, doi: 10.1016/j.compbiomed.2014.05.008.

[208] X. Li, S. Williams, and M. J. Bottema, “Constructing and applying higher order textons : Estimating breast cancer risk,” *Pattern Recognit.*, vol. 47, no. 3, pp. 1375–1382, 2014, doi: 10.1016/j.patcog.2013.09.033.

[209] Z. Suhail, M. Sarwar, and K. Murtaza, “Automatic detection of abnormalities in mammograms,” *BMC Med. Imaging*, vol. 15, no. 1, pp. 1–8, 2015, doi: 10.1186/s12880-015-0094-8.

[210] K. Vaidehi and T. . Subashini, “Automatic classification and retrieval of mammographic tissue density using texture features,” 2015.

[211] N. Al-najdawi, M. Biltawi, and S. Tedmori, “Mammogram image visual enhancement, mass segmentation and classification,” *Appl. Soft Comput. J.*, vol. 35, pp. 175–185, 2015, doi: 10.1016/j.asoc.2015.06.029.

[212] H. Lee and Y.-P. Phoebe Chen, “Image based computer aided diagnosis system for cancer detection,” *Expert Syst. Appl.*, vol. 42, no. 12, pp. 5356–5365, 2015, doi: 10.1016/j.eswa.2015.02.005.

[213] G. Helena, B. Miranda, and J. Cezar, "Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization," *Comput. Biol. Med.*, vol. 64, pp. 334–346, 2015, doi: 10.1016/j.compbiomed.2014.10.006.

[214] C. C. Jen and S. S. Yu, "Automatic detection of abnormal mammograms in mammographic images," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3048–3055, 2015, doi: 10.1016/j.eswa.2014.11.061.

[215] J. C. Fu, S. K. Lee, S. T. C. Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Comput. Med. Imaging Graph.*, vol. 29, no. 6, pp. 419–429, 2005, doi: 10.1016/j.compmedimag.2005.03.002.

[216] X. Liu, J. Liu, D. Zhou, and J. Tang, "A benign and malignant mass classification algorithm based on an improved level set segmentation and texture feature analysis," 2010, doi: 10.1109/ICBBE.2010.5518284.

[217] M. Sharkas, M. Al-Sharkawy, and D. A. Ragab, "Detection of microcalcifications in mammograms using support vector machine," 2011, doi: 10.1109/EMS.2011.23.

[218] M. Al Sharkawy, M. Sharkas, and D. A. Ragab, "Breast cancer detection using support vector machine technique applied on extracted electromagnetic waves," *Appl. Comput. Electromagn. Soc. J.*, vol. 27, no. 4, pp. 292–301, 2012.

[219] J. Anitha and J. Dinesh Peter, "A wavelet based morphological mass detection and classification in mammograms," in *2012 International Conference on Machine Vision and Image Processing, MVIP 2012*, 2012, pp. 25–28, doi: 10.1109/MVIP.2012.6428752.

[220] D. A. Ragab, M. Sharkas, and M. Al-Sharkawy, "A comparison between support vector machine and artificial neural network for breast cancer detection," in *the 12th International Conference on Signal Processing, Robotics and Automation (ISPRA '13), Cambridge, UK, 2013.*, 2013, pp. 171–176.

[221] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Syst. Appl.*, vol. 41, no. 4 Part 1, pp. 1476–1482, 2014, doi: 10.1016/j.eswa.2013.08.044.

[222] S. Beura, B. Majhi, and R. Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer," *Neurocomputing*, vol. 154, pp. 1–14, 2015, doi: 10.1016/j.neucom.2014.12.032.

[223] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," *Expert Syst. Appl.*, vol. 46, pp. 139–144, 2015, doi: 10.1016/j.eswa.2015.10.015.

[224] D. H. Kim, J. Y. Choi, and Y. M. Ro, "Region based stellate features combined with variable selection using AdaBoost learning in mammographic computer-aided detection," *Comput. Biol. Med.*, vol. 63, pp. 238–250, 2015, doi: 10.1016/j.compbiomed.2014.09.006.

[225] G. Saad, A. Khadour, and Q. Kanafani, "ANN and Adaboost application for automatic detection of microcalcifications in breast cancer," *Egypt. J. Radiol. Nucl. Med.*, vol. 47, no. 4, pp. 1803–1814, 2016, doi: 10.1016/j.ejrnm.2016.08.020.

[226] Y. D. Zhang, S. H. Wang, G. Liu, and J. Yang, "Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform," *Adv. Mech. Eng.*, vol. 8, no. 2, pp. 1–11, 2016, doi: 10.1177/1687814016634243.

[227] M. M. Pawar and S. N. Talbar, "Genetic Fuzzy System (GFS) based wavelet co-occurrence feature selection in mammogram classification for breast cancer diagnosis," *Perspect. Sci.*, vol. 8, pp. 247–250, 2016, doi: 10.1016/j.pisc.2016.04.042.

[228] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 173, pp. 930–941, 2016, doi: 10.1016/j.neucom.2015.08.048.

[229] S. Khan, M. Hussain, H. Aboalsamh, and G. Bebis, “A comparison of different Gabor feature extraction approaches for mass classification in mammography,” *Multimed. Tools Appl.*, vol. 76, no. 1, pp. 33–57, 2017, doi: 10.1007/s11042-015-3017-3.

[230] N. K. Al-Salihy and T. Ibrikci, “Classifying breast cancer by using decision tree algorithms,” in *ACM International Conference Proceeding Series*, 2017, pp. 144–148, doi: 10.1145/3056662.3056716.

[231] O. R. Seryasat and J. Haddadnia, “Evaluation of a new ensemble learning framework for mass classification in mammograms,” *Clin. Breast Cancer*, vol. 18, no. 3, pp. e407–e420, 2018, doi: 10.1016/j.clbc.2017.05.009.

[232] S. B. Yengec Tasdemir, K. Tasdemir, and Z. Aydin, “ROI detection in mammogram images using wavelet-based Haralick and HOG features,” in *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 105–109, doi: 10.1109/icmla.2018.00023.

[233] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, “A support vector machine-based ensemble algorithm for breast cancer diagnosis,” *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 687–699, 2018, doi: 10.1016/j.ejor.2017.12.001.

[234] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. F. Nagi, “Automated breast cancer diagnosis based on machine learning algorithms,” *J. Healthc. Eng.*, vol. 2019, no. Article ID 4253641, pp. 1–11, 2019, doi: https://doi.org/10.1155/2019/4253641 Research.

[235] D. A. Ragab, M. Sharkas, and O. Attallah, “Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers,” *Diagnostics*, vol. 9, no. 4, p. 165, Oct. 2019, doi: 10.3390/DIAGNOSTICS9040165.

[236] A. S. Assiri, S. Nazir, and S. A. Velastin, “Breast tumor classification using an ensemble machine learning method,” *J. Imaging*, vol. 6, no. 39, 2020, doi: 10.3390/jimaging6060039.

[237] S. A. Mohammed, S. Darrab, S. A. Noaman, and G. Saake, “Analysis of breast cancer detection using different machine learning techniques,” in *Data Mining and Big Data*, 2020, pp. 108–117.

[238] M. Mangukiya, A. Vaghani, and M. Savani, “Breast cancer detection with machine learning,” *Int. J. Reserach Appl. Sci. Eng. Technol.*, vol. 10, no. II, 2022, doi: 10.22214/ijraset.2022.40204.

[239] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, “Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review,” *Expert Syst. Appl.*, vol. 167, p. 114161, 2021, doi: 10.1016/j.eswa.2020.114161.

[240] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, “Convolutional neural networks for breast cancer detection in mammography: A survey,” *Comput. Biol. Med.*, vol. 131, p. 104248, 2021, doi: 10.1016/j.compbiomed.2021.104248.

[241] J. Bai, R. Posner, T. Wang, C. Yang, and S. Nabavi, “Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review,” *Med. Image Anal.*, vol. 71, p. 102049, 2021, doi: 10.1016/j.media.2021.102049.

[242] X. Yu, Q. Zhou, S. Wang, and Y.-D. Zhang, “A systematic survey of deep learning in breast cancer,” *Int. J. Intell. Syst.*, no. July 2021, pp. 152–216, 2021, doi: 10.1002/int.22622.

[243] G. Murtaza *et al.*, “Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges,” *Artif. Intell. Rev.*, 2019, doi: 10.1007/s10462-019-09716-5.

[244] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, “Deep learning for breast cancer diagnosis from mammograms—A comparative study,” *J. Imaging*, vol. 5, no. 3, p. 37, 2019, doi: 10.3390/jimaging5030037.

[245] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, and Y. Xie, "A technical review of convolutional neural network-based mammographic breast cancer diagnosis," *Hindawi Comput. Math. Methods Med.*, vol. 2019, no. Article ID 6509357, pp. 1–16, 2019, doi: 10.1155/2019/6509357.

[246] O. Emad, I. A. Yassine, and A. S. Fahmy, "Automatic localization of the left ventricle in cardiac MRI images using deep learning," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 683–686, doi: 10.1109/EMBC.2015.7318454.

[247] G. Piantadosi, M. Sansone, and C. Sansone, "Breast segmentation in MRI via U-Net deep convolutional neural networks," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3917–3922, doi: 10.1109/ICPR.2018.8545327.

[248] M. Benjelloun, M. El Adoui, M. A. Larhmam, and S. A. Mahmoudi, "Automated breast tumor segmentation in DCE-MRI using deep learning," in *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*, 2018, pp. 1–6, doi: 10.1109/CloudTech.2018.8713352.

[249] P. Herent *et al.*, "Detection and characterization of MRI breast lesions using deep learning," *Diagn. Interv. Imaging*, vol. 100, no. 4, pp. 219–225, 2019, doi: 10.1016/j.diii.2019.02.008.

[250] H. M. Whitney, H. Li, P. Liu, and M. L. Giger, "Comparison of breast MRI tumor classification using radiomics, transfer learning from deep convolutional neural networks, and fusion methods," *Proc. IEEE*, vol. 108, no. 1, pp. 163–177, 2020.

[251] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "AggNet: Deep Learning from Crowds for Mitosis Detection in Breast Cancer Histology Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1–1, 2016, doi: 10.1109/TMI.2016.2528120.

[252] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2016, doi: 10.1109/TBME.2015.2496264.

[253] M. Zejmo, M. Kowal, J. Korbicz, and R. Monczak, "Classification of breast cancer cytological specimen using convolutional neural network," *J. Phys. Conf. Ser. 783 012060*, 2017, doi: 10.1088/1742-6596/783/1/012060 Classification.

[254] A. Sarmiento, C. Eloy, and P. Aguiar, "Automatic classification of tissue malignancy for breast carcinoma diagnosis," *Comput. Biol. Med.*, vol. 96, pp. 41–51, 2018, doi: 10.1016/j.compbiomed.2018.03.003.

[255] B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks," *Pattern Recognit.*, vol. 84, pp. 345–356, 2018, doi: 10.1016/j.patcog.2018.07.022.

[256] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning," *IEEE Access*, vol. 7, pp. 21400–21408, 2019, doi: 10.1109/ACCESS.2019.2898044.

[257] S. U. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, 2019, doi: 10.1016/j.patrec.2019.03.022.

[258] Y. Jiang, L. C. Id, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PLoS One*, vol. 14, no. 3, p. e0214587, 2019.

[259] B. Akdemir, "HIC-net : A deep convolutional neural network model for classification of histopathological breast images," *Comput. Electr. Eng.*, vol. 76, pp. 299–310, 2019, doi: 10.1016/j.compeleceng.2019.04.012.

[260] N. Brancati, G. D. E. Pietro, M. Frucci, and D. Riccio, “A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images,” *IEEE Access*, vol. 7, pp. 44709–44720, 2019, doi: 10.1109/ACCESS.2019.2908724.

[261] Y. Wang, B. Lei, S. Member, A. Elazab, X. Gong, and T. Wang, “Breast cancer image classification via multi- network features and dual-Network orthogonal low-rank learning,” *IEEE Access*, vol. 8, pp. 27779–27792, 2020.

[262] M. Toğaçar, K. Baran, B. Ergen, and Z. Cömert, “BreastNet : A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer,” *Physica A*, vol. 545, p. 123592, 2020, doi: 10.1016/j.physa.2019.123592.

[263] S. Saxena and M. Gyanchandani, “Machine learning methods for computer-aided breast cancer diagnosis using histopathology: A narrative review,” *J. Med. Imaging Radiat. Sci.*, vol. 51, no. 1, pp. 182–193, 2020, doi: 10.1016/j.jmir.2019.11.001.

[264] Z. Gao, Z. Lu, J. Wang, S. Ying, and J. Shi, “A convolutional neural network and graph convolutional network based framework for classification of breast histopathological images,” *IEEE J. Biomed. Heal. Informatics*, vol. 2194, no. c, pp. 1–11, 2022, doi: 10.1109/JBHI.2022.3153671.

[265] C. H. Chen *et al.*, “Computer-aided diagnosis of endobronchial ultrasound images using convolutional neural network,” *Comput. Methods Programs Biomed.*, vol. 177, pp. 175–182, 2019, doi: 10.1016/j.cmpb.2019.05.020.

[266] W. Liao *et al.*, “Automatic identification of breast ultrasound image based on supervised block-based region segmentation algorithm and features combination migration deep learning model,” *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 4, pp. 984–993, 2020.

[267] A. Feizi, “A gated convolutional neural network for classification of breast lesions in ultrasound images,” *Soft Comput.*, vol. 9, 2022, doi: 10.1007/s00500-022-07024-9.

[268] W. K. Moon, Y. W. Lee, H. H. Ke, S. H. Lee, C. S. Huang, and R. F. Chang, “Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks,” *Comput. Methods Programs Biomed.*, vol. 190, no. 105361, pp. 1–12, 2020, doi: 10.1016/j.cmpb.2020.105361.

[269] Y. Pi *et al.*, “Automated diagnosis of multi-plane breast ultrasonography images using deep neural networks,” *Neurocomputing*, vol. 403, pp. 371–382, 2020, doi: 10.1016/j.neucom.2020.04.123.

[270] Y. Wang, E. J. Choi, Y. Choi, H. Zhang, G. Y. Jin, and S.-B. Ko, “Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning,” *Ultrasound Med. Biol.*, vol. 46, no. 5, pp. 1119–1132, 2020, doi: https://doi.org/10.1016/j.ultrasmedbio.2020.01.001.

[271] R. Irfan, A. A. Almazroi, H. T. Rauf, R. Damaševičius, E. A. Nasr, and A. E. Abdelgawad, “Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion,” *Diagnostics*, vol. 11, no. 7, pp. 1–20, 2021, doi: 10.3390/diagnostics11071212.

[272] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. Guevara Lopez, “Representation learning for mammography mass lesion classification with convolutional neural networks,” *Comput. Methods Programs Biomed.*, vol. 127, pp. 248–257, 2016, doi: 10.1016/j.cmpb.2015.12.014.

[273] I. Wichakam, P. Vateekul, K. Learning, and C. N. Convnets, “Combining deep convolutional networks and SVMs for mass detection on digital mammograms,” in *2016 8th International Conference on Knowledge and Smart Technology (KST)*, 2016, pp. 239–244, doi: 10.1109/KST.2016.7440527.

[274] A. Jain and D. Levy, “Breast mass classification using deep convolutional neural networks,” in *30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.*, 2016, pp. 1–6.

[275] D. Yi and J. Dunnmon, “Optimizing and visualizing deep learning for benign/malignant classification in breast tumors,” 2016.

[276] J. M.M., Z. Q., H. I.U., B. S., and J. A., “Three-class mammogram classification based on descriptive CNN features,” *Hindawi BioMed Res. Int.*, vol. 2017, no. Article ID 3640901, 2017, doi: 10.1155/2017/3640901.

[277] G. Carneiro, J. Nascimento, and A. P. Bradley, “Automated analysis of unregistered multi-view mammograms with deep learning,” *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2355–2365, 2017, doi: 10.1109/TMI.2017.2751523.

[278] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, “Deep multi-instance networks with sparse label assignment for whole mammogram classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10435 LNCS, pp. 603–611, 2017, doi: 10.1007/978-3-319-66179-7_69.

[279] X. Zhang *et al.*, “Whole mammogram image classification with convolutional neural networks,” in *2017 IEEE 11th International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 700–704, doi: 10.1109/BIBM.2017.8217738.

[280] G. Carneiro and A. P. Bradley, “Fully automated classification of mammograms using deep residual neural networks,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 310–314, doi: 10.1109/ISBI.2017.7950526.

[281] L. Shen, “End-to-end training for whole image breast cancer diagnosis using an all convolutional design,” *arXiv:1708.09427*, pp. 1–12, 2017, [Online]. Available: http://arxiv.org/abs/1708.09427.

[282] X. Zhang *et al.*, “Classification of mammographic masses by deep learning,” in *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2017, pp. 793–796, doi: 10.23919/SICE.2017.8105545.

[283] S. Duraisamy and S. Emperumal, “Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier,” *IET Comput. Vis.*, vol. 11, no. 8, pp. 656–662, 2017, doi: 10.1049/iet-cvi.2016.0425.

[284] R. Platania, S. Shams, S. Yang, J. Zhang, K. Lee, and S.-J. Park, “Automated breast cancer diagnosis using deep learning and region of interest detection (BC-DROID),” in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics - ACM-BCB ’17*, 2017, pp. 536–543, doi: 10.1145/3107411.3107484.

[285] A. H. Ahmed and M. A. M. Salem, “Mammogram-based cancer detection using deep convolutional neural networks,” in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, 2018, pp. 694–699, doi: 10.1109/ICCES.2018.8639224.

[286] P. Xi, C. Shu, and R. Goubran, “Abnormality detection in mammography using deep convolutional neural networks,” in *2018 13th IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, Italy*, 2018, pp. 1–6, doi: 10.1109/MeMeA.2018.8438639.

[287] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, “Detecting and classifying lesions in mammograms with deep learning,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, 2018, doi: 10.1038/s41598-018-22437-z.

[288] M. A. Al-antari, M. A. Al-masni, M. Choi, S. Han, and T. Kim, “A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification,” *Int. J. Med. Inform.*, vol. 117, no. April, pp. 44–54, 2018, doi: 10.1016/j.ijmedinf.2018.06.003.

[289] H. Chougrad, H. Zouaki, and O. Alheyane, “Deep convolutional neural networks for breast cancer screening,” *Comput. Methods Programs Biomed.*, vol. 157, pp. 19–30, 2018, doi: 10.1016/j.cmpb.2018.01.011.

[290] F. F. Ting, Y. J. Tan, and K. S. Sim, “Convolutional neural network improvement for breast cancer classification,” *Expert Syst. Appl.*, vol. 120, pp. 103–115, 2019, doi: 10.1016/j.eswa.2018.11.008.

[291] H. Mechria, M. S. Gouider, and K. Hassine, “Breast cancer detection using deep convolutional neural network,” in *Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART2019)*, 2019, pp. 655–660, doi: 10.5220/0007386206550660.

[292] H. Li, S. Zhuang, D. ao Li, J. Zhao, and Y. Ma, “Benign and malignant classification of mammogram images based on deep learning,” *Biomed. Signal Process. Control*, vol. 51, pp. 347–354, 2019, doi: 10.1016/j.bspc.2019.02.017.

[293] K. Mendel, H. Li, D. Sheth, and M. Giger, “Transfer learning from convolutional neural networks for computer-aided diagnosis: A comparison of digital breast tomosynthesis and full-field digital mammography,” *Acad. Radiol.*, vol. 26, no. 6, pp. 735–743, 2019, doi: 10.1016/j.acra.2018.06.019.

[294] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, “Automatic mass detection in mammograms using deep convolutional neural networks,” *J. Med. Imaging*, vol. 6, no. 03, p. 1, 2019, doi: 10.1117/1.jmi.6.3.031409.

[295] S. Shamy and J. Dheeba, “A research on detection and classification of breast cancer using k-means GMM & CNN algorithms,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6S, pp. 501–505, 2019, doi: 10.35940/ijeat.F1102.0886S19.

[296] W. Lotter *et al.*, “Robust breast cancer detection in mammography and digital breast tomosynthesis using annotation-efficient deep learning approach,” *arXiv Prepr. arXiv1912.11027*, pp. 1–16, 2019.

[297] S. M. Mckinney *et al.*, “International evaluation of an AI system for breast cancer screening,” *Nature*, vol. 577, pp. 89–94, 2020, doi: 10.1038/s41586-019-1799-6.

[298] M. Alkhaleefah, P. Kumar Chittem, V. P. Achhannagari, S. C. Ma, and Y. L. Chang, “The influence of image augmentation on breast lesion classification using transfer learning,” in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, 2020, pp. 1–5, doi: 10.1109/AISP48273.2020.9073516.

[299] R. Agarwal, O. Díaz, M. H. Yap, X. Lladó, and R. Martí, “Deep learning for mass detection in full field digital mammograms,” *Comput. Biol. Med.*, vol. 121, p. 103774, 2020, doi: 10.1016/j.compbiomed.2020.103774.

[300] N. Wu *et al.*, “Deep neural networks improve radiologists’ performance in breast cancer screening,” *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1184–1194, 2020, doi: 10.1109/TMI.2019.2945514.

[301] M. A. Al-antari and T. Kim, “Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms,” *Comput. Methods Programs Biomed.*, vol. 196, p. 105584, 2020, doi: 10.1016/j.cmpb.2020.105584.

[302] D. Arefan, A. A. Mohamed, W. A. Berg, M. L. Zuley, J. H. Sumkin, and S. Wu, “Deep learning modeling using normal mammograms for predicting breast cancer risk,” *Med. Phys.*, vol. 47, no. 1, pp. 110–118, 2020, doi: 10.1002/mp.13886.

[303] S. Wessels and D. van der Haar, “Applying deep learning for the detection of abnormalities in mammograms,” in *Information Science and Applications.*, vol. 621, K. H. (eds) Kim K., Ed. Singapore: Springer Singapore, 2020, pp. 201–210.

[304] S. A. Agnes, J. Anitha, S. I. A. Pandian, and J. D. Peter, "Classification of mammogram images using multiscale all convolutional neural network (MA-CNN)," *J. Med. Syst.*, vol. 44, no. 30, pp. 29–30, 2020, doi: 10.1007/s10916-019-1494-z.

[305] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, and H. Chen, "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique," *IEEE Access*, vol. 9, pp. 71194–71209, 2021, doi: 10.1109/ACCESS.2021.3079204.

[306] G. Hamed, M. Marey, S. Amin, and M. F. Tolba, "Automated breast cancer detection and classification in full field digital mammograms using two full and cropped detection paths approach," *IEEE Access*, vol. 9, pp. 116898–116913, 2021, doi: 10.1109/ACCESS.2021.3105924.

[307] S. J. Malebary and A. Hashmi, "Automated breast mass classification system using deep learning and ensemble learning in digital mammogram," *IEEE Access*, vol. 9, pp. 55312–55328, 2021, doi: 10.1109/ACCESS.2021.3071297.

[308] U. Mangai, S. Samanta, S. Das, and P. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Tech. Rev.*, vol. 27, no. 4, p. 293, 2010, doi: 10.4103/0256-4602.64604.

[309] J. Dheeba, N. A. Singh, and S. T. Selvi, "Computer-aided detection of breast cancer on mammograms : A swarm intelligence optimized wavelet neural network approach," *J. Biomed. Inform.*, vol. 49, pp. 45–52, 2014, doi: 10.1016/j.jbi.2014.01.010.

[310] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection from mammograms using deep learning and random forest," 2015, doi: 10.1109/DICTA.2015.7371234.

[311] N. Dhungel, G. Carneiro, and A. P. Bradley, "The automated learning of deep features for breast mass classification from mammograms," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2016*, 2016, pp. 106–114.

[312] Q. Abbas, "DeepCAD: A computer-aided diagnosis system for mammographic masses using deep invariant features," *Computers*, vol. 5, no. 4, p. 28, 2016, doi: 10.3390/computers5040028.

[313] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, 2016, doi: 10.1016/j.neucom.2016.02.060.

[314] T. Kooi *et al.*, "Large scale deep learning for computer aided detection of mammographic lesions," *Med. Image Anal.*, vol. 35, pp. 303–312, 2017, doi: 10.1016/j.media.2016.07.007.

[315] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale CNN and curriculum learning strategy for mammogram classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2017, pp. 169–177.

[316] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Med. Image Anal.*, vol. 37, pp. 114–128, 2017, doi: 10.1016/j.media.2017.01.009.

[317] Y. J. Tan, K. S. Sim, and F. F. Ting, "Breast cancer detection using convolutional neural networks for mammogram imaging system," in *2017 2nd International Conference on Robotics, Automation and Sciences (ICORAS)*, 2017, pp. 1–5, doi: 10.1109/ICORAS.2017.8308076.

[318] W. Sun, T. L. B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Comput. Med. Imaging Graph.*, vol. 57, pp. 4–9, 2017, doi: 10.1016/j.compmedimag.2016.07.004.

[319] S. Karthik, R. Srinivasa Perumal, and P. V. S. S. . Chandra Mouli, "Breast cancer classification using deep neural networks," in *Knowledge Computing and Its Applications*, no. February, S. M. A. and U. K. W. (eds.), Ed. Springer Nature Singapore, 2018, pp. 227–241.

[320] J. Lamy, B. Sekar, G. Guezennec, J. Bouaud, and B. Séroussi, “Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach,” *Artif. Intell. Med.*, vol. 94, pp. 42–53, 2019, doi: 10.1016/j.artmed.2019.01.001.

[321] V. Bevilacqua, A. Brunetti, A. Guerriero, G. F. Trotta, M. Telegrafo, and M. Moschetta, “A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images,” *Cogn. Syst. Res.*, vol. 53, pp. 3–19, 2019, doi: 10.1016/j.cogsys.2018.04.011.

[322] P. Wang *et al.*, “Cross-task extreme learning machine for breast cancer image classification with deep convolutional features,” *Biomed. Signal Process. Control*, vol. 57, 2020, doi: 10.1016/j.bspc.2019.101789.

[323] A. H. Osman and H. M. A. Aljahdali, “An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model,” *IEEE Access*, vol. 8, pp. 39165–39174, 2020, doi: 10.1109/ACCESS.2020.2976149.

[324] G. Hamed, M. A. E.-R. Marey, S. E.-S. Amin, and M. F. Tolba, “Deep learning in breast cancer detection and classification,” in *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, 2020, pp. 322–333.

[325] M. He, D. Lin, Z. Gao, and J. Fan, “Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis,” *IEEE Access*, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/ACCESS.2020.2993536.

[326] R. Lavanya and N. Nagarajan, “Information fusion in CAD systems for breast cancer diagnosis using mammography and ultrasound imaging: A survey,” *J. Artif. Intell.*, vol. 7, no. 3, pp. 113–122, 2014, doi: 10:3923/jai.2014.113.122.

[327] E. E. Kim *et al.*, “High-resolution breast cancer screening with multi-view deep convolutional neural networks,” *Comput. Methods Programs Biomed.*, vol. 9006, no. 2, pp. 89–107, 2017, doi: 10.1007/978-3-319-10593-2_13.

[328] N. Dhungel, G. Carneiro, and A. P. Bradley, “Combining deep learning and convolutional neural networks for medical image computing,” in *Deep Learning and Convolutional Neural Networks for Medical Image Computing, Advances in Computer Vision and Pattern Recognition*, L. Lu et al. (eds.), Ed. Springer International Publishing Switzerland 2017, 2017, pp. 225–240.

[329] Z. Wang *et al.*, “Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features,” *IEEE Access*, vol. 7, pp. 1–1, 2019, doi: 10.1109/access.2019.2892795.

[330] H. N. Khan, A. R. Shahid, A. H. Dar, and H. Alquhayz, “Multi-view feature fusion based four views model for mammogram classification using convolutional neural network,” *IEEE Access*, vol. 7, pp. 165724–165733, 2019, doi: 10.1109/ACCESS.2019.2953318.

[331] R. Song, T. Li, and Y. A. N. Wang, “Mammographic classification based on XGBoost and DCNN with multi features,” *IEEE Access*, vol. 8, pp. 75011–75021, 2020, doi: 10.1109/ACCESS.2020.2986546.

[332] H. Zhang *et al.*, “DE-Ada*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions,” *Inf. Sci. (Ny).*, vol. 539, pp. 461–486, 2020, doi: 10.1016/j.ins.2020.05.080.

[333] X. Yu, W. Pang, Q. Xu, and M. Liang, “Mammographic image classification with deep fusion learning,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020-71431-x.

[334] R. Arora, P. K. Rai, and B. Raman, “Deep feature–based automatic classification of mammograms,” *Med. Biol. Eng. Comput.*, vol. 58, no. 6, pp. 1199–1211, 2020, doi: 10.1007/s11517-020-02150-8.

[335] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages," *Appl. Sci.*, vol. 12, no. 7, p. 3273, 2022, doi: 10.3390/app12073273.

[336] S. Zahoor, U. Shoaib, and I. U. Lali, "Breast cancer mammograms classification using deep neural network and entropy-controlled whale optimization algorithm," *Diagnostics*, vol. 12, 2022, doi: https://doi.org/10.3390/diagnostics12020557.

[337] E. M. Alkabawi, A. R. Hilal, and O. A. Basir, "Computer-aided Classification of multi-types of dementia via convolutional neural networks," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2017, pp. 45–50, doi: 10.1109/MeMeA.2017.7985847.

[338] A. Brunetti, L. Carnimeo, G. F. Trotta, and V. Bevilacqua, "Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: A survey based on medical images," *Neurocomputing*, vol. 335, pp. 274–298, 2019, doi: 10.1016/j.neucom.2018.06.080.

[339] G. S. Tandel *et al.*, "A review on a deep learning perspective in brain cancer classification," *Cancers (Basel).*, vol. 11, no. 1, 2019, doi: 10.3390/cancers11010111.

[340] L. I. Smith, "A tutorial on Principal Components Analysis," *Cornell University, USA*. 2002.

[341] F. Mohanty, S. Rup, B. Dash, B. Majhi, and M. N. S. Swamy, "Mammogram classification using contourlet features with forest optimization-based feature selection approach," *Multimed. Tools Appl.*, pp. 1–30, 2018, doi: 10.1007/s11042-018-5804-0.

[342] P. U. Hepsağ, S. A. Özel, and A. Yazici, "Using deep learning for mammography classification," in *2nd International Conference on Computer Science and Engineering, UBMK 2017*, 2017, pp. 418–423, doi: 10.1109/UBMK.2017.8093429.

[343] "Mammomat inspiration with PRIME technology." https://static.healthcare.siemens.com/siemens_hwem-hwem_ssxa_websites-context-root/wcm/idc/groups/public/@global/@imaging/@mammo/documents/download/mda1/nzez/~edisp/mammography_mammomat_inspiration_prime_mammography_screening_machine_product_brochure-feb- (accessed Nov. 03, 2019).

[344] J. Nagi, S. A. Kareem, F. Nagi, and S. K. Ahmed, "Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms," in *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2010, pp. 87–92.

[345] O. Attallah *et al.*, "Using multiple classifiers for predicting the risk of endovascular aortic aneurysm repair re-intervention through hybrid feature selection," *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, vol. 231, no. 11, pp. 1048–1063, 2017, doi: 10.1177/0954411917731592.

[346] "Weka." https://www.cs.waikato.ac.nz/ml/weka/ (accessed Sep. 15, 2018).

[347] D. Posada and T. R. Buckley, "Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests," *Syst. Biol.*, vol. 53, no. 5, pp. 793–808, 2004, doi: 10.1080/10635150490522304.

[348] M. Halicek *et al.*, "Conditional generative adversarial network for synthesizing hyperspectral images of breast cancer cells from digitized histology," *Med. Imaging Digit. Pathol. Int. Soc. Opt. Photonics*, vol. 11320, pp. 139–148, 2020, doi: doi:10.1117/12.2549994.

[349] M. U. Dalmiş *et al.*, "Using deep learning to segment breast and fibroglandular tissue in MRI volumes," *Med. Phys.*, vol. 44, no. 2, pp. 533–546, 2017, doi: 10.1002/mp.12079.

[350] H. Chen *et al.*, "Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity-A multi-center study," *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 64, pp. 1–9, 2016, doi: 10.1016/j.pnpbp.2015.06.014.

[351] D. Tomasi and N. D. Volkow, “Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder,” *Biol. Psychiatry*, vol. 75, no. 5, pp. 443–450, 2012, doi: 10.1038/jid.2014.371.

[352] R. Mokni, N. Gargouri, A. Damak, D. Sellami, W. Feki, and Z. Mnif, “An automatic Computer-Aided Diagnosis system based on the Multimodal fusion of Breast Cancer (MF-CAD),” *Biomed. Signal Process. Control*, vol. 69, no. October 2020, 2021, doi: 10.1016/j.bspc.2021.102914.

[353] J. Suckling *et al.*, “Power calculations for multicenter imaging studies controlled by the false discovery rate,” *Hum. Brain Mapp.*, vol. 31, no. 8, pp. 1183–1195, 2010, doi: 10.1002/hbm.20927.

[354] D. L. Rimm *et al.*, “An international multicenter study to evaluate reproducibility of automated scoring for assessment of Ki67 in breast cancer,” *Mod. Pathol.*, vol. 32, no. 1, pp. 59–69, 2019, doi: 10.1038/s41379-018-0109-4.

[355] J. Park *et al.*, “PROMISE CLIP project: A retrospective, multicenter study for prostate cancer that integrates clinical, imaging and pathology data,” *Appl. Sci.*, vol. 9, no. 15, pp. 1–7, 2019, doi: 10.3390/app9152982.

[356] G. Hamerla *et al.*, “Comparison of machine learning classifiers for differentiation of grade 1 from higher gradings in meningioma: A multicenter radiomics study,” *Magn. Reson. Imaging*, vol. 63, no. August, pp. 244–249, 2019, doi: 10.1016/j.mri.2019.08.011.

[357] L.-H. Hu *et al.*, “Machine learning predicts per-vessel early coronary revascularization after fast myocardial perfusion SPECT: results from multicentre REFINE SPECT registry,” *Eur. Hear. J. - Cardiovasc. Imaging*, pp. 1–11, 2019, doi: 10.1093/ehjci/jez177.

[358] T. S. Dong *et al.*, “Machine learning-based development and validation of a scoring system for screening high-risk esophageal varices,” *Clin. Gastroenterol. Hepatol.*, vol. 17, no. 9, pp. 1894-1901.e1, Aug. 2019, doi: 10.1016/j.cgh.2019.01.025.

[359] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” *Artif. Intell. Rev.*, vol. 56, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[360] E. Culurciello, “Towards data science: Neural network architectures,” 2017. https://towardsdatascience.com/neural-network-architectures-156e5bad51ba (accessed Oct. 30, 2020).

[361] L. D. Nguyen, D. Lin, and Z. Lin, “Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation,” 2018, doi: 10.1109/ISCAS.2018.8351550.

[362] H. Lei *et al.*, “A deeply supervised residual network for HEp-2 cell classification via cross modal transfer learning,” *Pattern Recognit.*, vol. 79, pp. 290–302, 2018, doi: https://doi.org/10.1016/j.patcog.2018.02.006.

[363] Z. Li, S. Yu, W. Qin, L. Liu, K. Li, and T. Xiao, “Comparison of transferred deep neural networks in ultrasonic breast masses discrimination,” *Biomed Res. Int.*, vol. 2018, no. Article ID 4605191, pp. 1–9, 2018, doi: 10.1155/2018/4605191.

[364] “Why Google’s MobileNet-V2 is a revolutionary next gen on-device computer vision network.” https://analyticsindiamag.com/why-googles-mobilenetv2-is-a-revolutionary-next-gen-on-device-computer-vision-network/ (accessed Dec. 24, 2019).

[365] A. G. Howard *et al.*, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv:1704.04861*, 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.