



UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: Development of Deep Learning-based Model for Breast Cancer Detection and Identification of Molecular Subtypes using Mammography Images

Trabajo de integración curricular presentado como requisito para la
obtención del título de Ingeniera Biomédica

Autor/a:

Brenda Dayana Guanulema Barba

Tutor/a:

PhD - Diego Alfonso Almeida Galárraga

Co-tutor/a:

PhD - Manuel Eugenio Morocho Cayamcela

Urcuquí, Enero del 2025

Autoría

Yo, **BRENDA DAYANA GUANULEMA BARBA**, con cédula de identidad 1726640285, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autor/a del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Enero del 2025.

Brenda Dayana Guanulema Barba

CI: 1726640285

Autorización de publicación

Yo, **BRENDA DAYANA GUANULEMA BARBA**, con cédula de identidad 1726640285, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación

Urcuquí, Enero del 2025.

Brenda Dayana Guanulema Barba

CI: 1726640285

Dedication

With deep gratitude and love, I dedicate this work to those who have been my pillars of strength and inspiration throughout this journey.

To my beloved mom, Eulidia, my unconditional support and my strength. Your love, sacrifice, and guidance have been the pillar that has sustained me through every step of this journey.

To my best friend and soul sister Yadi, for always encouraging me and standing by my side.

To Anne, Lya, Cata, and Eimy—you are my greatest inspiration.

And finally, to my faithful companion, Kirei.

Brenda Dayana Guanulema Barba

Acknowledgment

I am deeply grateful to my parents Eulidia and Jorge, for their immense sacrifice and unconditional support. To my siblings, Jorge, Jonathan, and my sister Yadira, thank you for your guidance, advice, and for always being there. Your encouragement has been a constant source of strength, and your little sister will forever be grateful

A special thanks to my chivi Lya, for always brightening my days with your delightful smile.

My girls, Pame and Dey thank you for being by my side since day one of this journey. We've shared so good memories together that I will always cherish with great affection. Also, thanks to Mabe, Diego, Emi, and Ali, you brought that spark of joy my life needed. The meals, parties, trips, and long nights of gossip with you all have been and will always be some of the best moments of my life.

To my roommate, May, thank you for listening to me, for your advice, and for motivating me when I needed it the most.

I am immensely grateful to my dear guide during this long trip called '*biomedical engineering*', to my favorite professor and thesis advisor, Diego Almeida, for his patience and dedication. Thank you for believing in me and for allowing me to be a part of your projects.

Thank you to my thesis co-advisor, Eugenio Morocho, for sharing your knowledge and guiding me through this process. I have learned so much from you, and I am deeply appreciative. To my alma mater, Yachay Tech University, thank you for being my second home and the place where curiosity turned into knowledge and challenges became opportunities for growth.

Brenda

Resumen

La detección temprana y la clasificación precisa de subtipos moleculares del cáncer de mama son importantes para proporcionar un tratamiento adecuado. Este proyecto se centra en el desarrollo de un modelo de aprendizaje profundo para la detección de esta enfermedad y la identificación de sus subtipos moleculares a través de imágenes de mamografía. Para este estudio, se emplearon dos conjuntos de datos que incluyen mamografías de pacientes sin cáncer de mama y pacientes con esta enfermedad. La primera base de datos incluye mamografías de 1416 pacientes, mientras que la segunda contiene mamografías de 1775 pacientes, de los cuales 749 están divididos en cuatro subtipos moleculares: Luminal A, Luminal B, HER2-enriquecido y Triple Negativo. Para los cuales, se realizó la segmentación de las lesiones mamarias. El modelo de detección de cáncer de mama basado en AlexNet, alcanzó una precisión de 98.60%. Para la identificación y clasificación de subtipos moleculares, se implementó un modelo de aprendizaje por transferencia utilizando MobileNet, logrando una precisión general del 73.95% y un área bajo la curva (AUC) del 89%. Los resultados indicaron que el subtipo HER2-enriquecido fue el mejor caracterizado, con un AUC del 96%, seguido del subtipo Triple negativo y Luminal A, ambos con un AUC del 89% y finalmente el subtipo Luminal B con un 82%. En conclusión, estos hallazgos proporcionan un modelo potencial basado en imágenes médicas para la detección del carcinoma de mama y la predicción de sus subtipos moleculares.

Palabras Clave:

Detección de cáncer de mama, subtipos moleculares, mamografía, aprendizaje profundo, aprendizaje por transferencia.

Abstract

Early detection and accurate classification of molecular subtypes of breast cancer are essential to provide appropriate treatment. This project focuses on the development of a Deep Learning model for the detection of this disease and the identification of its molecular subtypes through mammography images. In this study, two datasets were used, including mammograms of patients without breast cancer and patients with breast cancer. The first database includes mammograms of 1416 patients, while the second contains mammograms of 1775 patients, of which 749 are divided into four molecular subtypes: Luminal A, Luminal B, HER2-enriched and Triple Negative. For which, segmentation of breast lesions was performed. The AlexNet-based breast cancer detection model achieved an accuracy of 98.60%. For the identification and classification of molecular subtypes, a transfer learning model was implemented using MobileNet, achieving an overall accuracy of 73.95% and Area Under Curve (AUC) of 89%. The results indicated that the HER2-enriched subtype was the best characterized, with an AUC of 96%, followed by the Triple-negative and Luminal A, both with an AUC of 89%, and finally, Luminal B with 82%. In conclusion, these findings provide a potential model based on medical imaging for the detection of breast carcinoma and the prediction of its molecular subtypes.

Keywords:

Breast cancer detection, molecular subtypes, deep learning, mammography, transfer learning.

Contents

Dedication	iii
Acknowledgment	iv
Resumen	v
Abstract	vi
Contents	vii
List of Tables	x
List of Figures	xi
List of Abbreviations	xiii
1 Introduction	1
1.1 Background	1
1.2 Problem statement	2
1.3 Objectives	2
1.3.1 General Objective	2
1.3.2 Specific Objectives	2
1.4 Document Organization	3
2 Theoretical Framework	4
2.1 Breast Cancer	4
2.1.1 Risk factors	4
2.1.2 Screening Detection Methods	5
2.1.3 Molecular subtypes	7

2.1.4	Imaging features of Molecular Subtypes	9
2.2	Artificial Intelligence	12
2.2.1	Machine Learning	12
2.2.2	Deep Learning	13
2.2.3	Convolutional Neural Networks	14
3	State of the Art	19
3.1	Deep learning-based model for detection of breast cancer	19
3.2	Deep learning-based models for identification and classification of molecular subtypes	22
4	Methodology	25
4.1	Dataset description	26
4.1.1	CMMD	26
4.1.2	KAU-BCMD	27
4.1.3	CBIS-DDSM	28
4.2	Detection of Breast Cancer	28
4.2.1	Data Preparation	28
4.2.2	Data Pre-processing	28
4.2.3	Proposed model based on AlexNet	30
4.3	Identification and classification of breast cancer molecular subtypes	31
4.3.1	Data Preparation	31
4.3.2	Data Pre-processing	32
4.3.3	Transfer Learning Models	34
4.4	Experimental Setup	35
4.4.1	Software and Hardware	35
4.4.2	Experiments for Breast Cancer Detection	35
4.4.3	Experiments for Molecular Subtypes Classification	36
4.5	Model Performance Metrics	40
4.5.1	Evaluation metrics for Image Classification	40
4.5.2	Evaluation metrics for Image Segmentation	41

5	Results and Discussion	43
5.1	Detection of Breast Cancer	43
5.1.1	Experiment 1: Comparison between the previous model based on AlexNet and fine-tuned proposed model.	43
5.2	Identification and classification of Molecular Subtypes with Transfer Learning	46
5.2.1	Experiment 1: Comparison of TL with ResNet50, VGG16, Xception, InceptionV3 models using BROI images.	46
5.2.2	Experiment 2: Comparison of Unet and Attention Unet in breast tumor segmentation.	47
5.2.3	Experiment 3: Evaluation of TL models for identification of molecular subtypes using breast tumors ROI images	48
5.2.4	Experiment 4: Selection and evaluation of best TL model for classification of breast cancer molecular subtypes	50
5.2.5	Overall Analysis and Comparison	55
6	Conclusion and Future Works	58
6.1	Conclusion	58
6.2	Limitations	59
6.3	Future works	59
6.4	Scientific Dissemination	60
	Bibliography	61
	Appendices	71
.1	Appendix 1. Algorithms for Data Pre-processing	72
.2	Appendix 2. Complementary results of segmentation models (plots)	73
.2.1	U-net model	73
.2.2	Attention U-net model	74

List of Tables

2.1	Clinicopathological characteristics of molecular subtypes of breast cancer.	8
2.2	Imaging features of molecular subtypes of breast cancer.	10
3.1	Related works on deep learning models for breast cancer detection that use mammographies.	21
3.2	Related works on deep learning models for classification and prediction of molecular subtypes.	24
4.1	Summary of the datasets used, with their specifications.	26
5.1	Comparison of proposed fine-tuned AlexNet model with related works	46
5.2	Model performance of U-net and Attention U-net	47
5.3	Performance comparison of different TL in molecular subtypes classification using breast tumors ROI.	49
5.4	Transfer learning model based on MobileNet performance in multi-class classification.	54
5.5	Comparison of our proposed model and state-art related works	55

List of Figures

2.1	Imaging features of Luminal A subtype.	10
2.2	Imaging features of Luminal B subtype.	11
2.3	Imaging features of HER2 subtype	11
2.4	Imaging features of Triple Negative subtype	11
2.5	Biological neural network and their relationship with artificial neural network.	13
2.6	Structure of an Artificial Neural Network (ANN)	14
2.7	General structure of a typical convolutional neural network (CNN) for image classification.	16
4.1	Flowchart of the proposed methodology.	25
4.2	CC and MLO views of the left and right breasts of sample mammogram of CMMD dataset.	26
4.3	Distribution of molecular subtypes per class in CMMD2 dataset.	27
4.4	Block diagram for data pre-processing.	29
4.5	Breast ROI cropping and CLAHE application for mammogram image processing.	30
4.6	AlexNet model architecture proposed for breast cancer detection.	31
4.7	Data split into Train, Validation and Test for each molecular subtype.	31
4.8	Attention U-net model architecture.	33
4.9	Breast tumor ROI extraction process.	34
4.10	Block diagram of model segmentation selection process to obtained predicted mask with location of breast tumor.	37

4.11	Samples training set of breast tumors images corresponding to HER2-enriched, Luminal A, Luminal B, and Triple negative subtypes cropped using ROI extraction.	38
4.12	Flow-chart of evaluation of different CNNs models in classification of molecular subtypes using ROI breast tumors images.	38
4.13	Flowchart of the Transfer Learning Process with best model for the Classification of Molecular Subtypes.	39
5.1	Accuracy and Loss curves of previous pre-trained AlexNet model.	44
5.2	Accuracy and Loss curves of fine-tuning AlexNet model.	44
5.3	Confusion matrix of proposed Alexnet model applying fine-tuning.	45
5.4	Comparison of Transfer Learning models performance (ResNet50, Xception, VGG16, InceptionV3) using BROI images.	47
5.5	Comparison of sample segmentation results of Unet and Att-Unet models .	48
5.6	Bar chart-based performance comparative analysis of different CNNs in molecular subtypes classification.	49
5.7	Architecture proposed of transfer learning model based on MobileNet.	51
5.8	Accuracy and Loss curves of Fine-tuned MobileNet model.	51
5.9	Confusion matrix of Fine-tuned MobileNet model.	53
5.10	AUC-ROC curves of proposed Fine-Tuned MobileNet model	53
1	Accuracy, Loss (Top), Jaccard index (Bottom) curves of U-net model.	74
2	Accuracy, Loss (Top), and Jaccard index (Bottom) curves of Att-Unet model.	74

List of Abbreviations

AI, Artificial intelligence

AUC, Area under the curve

AUG, Augmentation

BC, Breast cancer

BI-RADS, Breast Imaging Reporting and Data System

BROI, Breast region of interest

CAD, Computer-aided diagnosis

CBIS-DDSM, Curated Breast Imaging Subset of DDSM **CC**, Craniocaudal

ChT, Chemotherapy

CLAHE, Contrast Limited Adaptive Histogram Equalization

CMMD, The Chinese Mammography Database

CNN, Convolutional neural network

DBT, Digital breast tomosynthesis

DDSM, Digital Database for Screening Mammography

DICOM, Digital Imaging and Communications in Medicine

DL, Deep learning

ER, Estrogen receptor

HER2, Human epidermal growth factor receptor 2

HRT, Hormonal replacement therapy

IHC, Immunohistochemistry or immunohistochemical

IoU, Intersection over Union

LumA, Luminal A

LumB, Luminal B

MG, Mammography

ML, Machine learning

MLO, Mediolateral oblique

MRI, Magnetic resonance imaging

PR, Progesterone receptor

ReLU, Rectified Linear Unit

ROC, Receiver operating characteristic

ROI, Region of interest

SGD, Stochastic Gradient Descent

TCIA, The Cancer Imaging Archive

TL, Transfer learning

TN, Triple negative

US, Ultrasound

Chapter 1

Introduction

1.1 Background

Breast cancer (BC) is a common type of cancer in women, with a higher incidence in patients worldwide as skin cancer [1]. According to Surveillance, Epidemiology, and End Results (SEER) Program's 2024 estimates, nearly 310,720 women will be diagnosed with breast cancer, and 42,250 will die caused by this disease [2]. BC can also occurs in males, though at a much lower rate of about 0.5% to 1% of all breast cancer cases [3]. Consequently, early detection of breast cancer is crucial for improving patient outcomes and survival rates. However, conventional breast cancer detection methods—such as mammography, ultrasound, and magnetic resonance imaging (MRI)—have limitations. These methods can be subjective, prone to inter-reader variability, and time-consuming to analyze.

Deep learning (DL), a branch of artificial intelligence, has become as a potent tool for medical imaging analysis, especially used for detection of diseases such as breast cancer. Deep learning algorithms, particularly convolutional neural networks (CNNs) have the capacity to learn complex patterns from large datasets of images, that radiologists might overlook. Studies have demonstrated that deep learning models significantly outperforms traditional methods in breast cancer detection, with higher accuracy, fewer rate of false positives, and overall better results [4].

This groundbreaking technology holds great promise for early breast cancer detection, contributing to lower mortality rates, and more effective treatment planning.

1.2 Problem statement

Breast cancer remains a critical health issue globally, contributing to a substantial number of deaths each year. Furthermore, the World Health Organization (WHO) reported several new cases of breast cancer in females of all ages in Ecuador in 2020, of 3563 women [5]. The accurate diagnosis and identification of breast cancer molecular subtypes are crucial for determining the most appropriate therapy and providing valuable information for predicting prognosis in order to improve survival rate. However, identifying them today requires invasive and time-consuming procedures such as biopsies. Therefore, it is crucial to seek for more accurate, reliable, non-invasive and automated approaches to classify breast cancer molecular subtypes. The development of a deep learning-based approach using mammograms could offer a promising avenue to enhance breast cancer detection, diagnosis and treatment outcomes.

1.3 Objectives

1.3.1 General Objective

Develop a deep learning model for the detection of breast cancer and identification of its molecular subtype based on mammography images.

1.3.2 Specific Objectives

- Design and implement a deep neural network to detect the breast cancer in mammography images.
- Develop and train a deep learning model capable of identifying and classifying molecular subtypes from mammography images.
- Determine the relationship between mammography features and the breast cancer molecular subtypes.
- Evaluate the performance of proposed model using standard classification metrics, such as accuracy, precision, recall , F1-score.

- Compare the proposed model with existing models of breast cancer detection and molecular subtype classification.

1.4 Document Organization

This research work is structured into 6 chapters. These are detailed below:

- Chapter 1: Introduction. This chapter presents the motivation and general context of the research topic, highlight the importance of early detection of breast cancer, and the objectives of the study are established.
- Chapter 2: Theoretical framework. This chapter covers an overview of key concepts in the field of breast cancer detection and molecular subtypes. The theoretical bases of deep learning and convolutional neural networks.
- Chapter 3: State of art. In this section reviews current and past research work in the development of models based on deep learning for breast cancer detection and classification of their molecular subtypes.
- Chapter 4: Methodology. This section provides a detailed procedure used to design and develop our proposed model, including data collection, training process, the experimental setup and the evaluation metrics used to measure model performance.
- Chapter 5: Results and Discussion. In this chapter, our study's results are presented and analyzed. It also includes the implications of the findings, highlighting our model's performance and potential areas for future improvement
- Chapter 6: Conclusion and Future Work. This chapter summarizes the main findings of the research, discuss the limitations of the model and proposes possible improvements for future research.

Chapter 2

Theoretical Framework

2.1 Breast Cancer

Breast cancer is a major health concern for women globally, accounting for 12.5% of all new cases each year and ranking as the leading cause of cancer death among women [6]. This disease consists of the abnormal growth of breast cells forming tumors that can be benign and malignant. Benign tumors are non-cancerous tumors that typically grow slowly and stay confined, while malignant tumors can invade surrounding tissues and spread throughout the body [7]. It usually originates in the ducts (ductal carcinoma) or lobules (lobular carcinoma) of the breast, which are the glands that produce milk. It can present as a lump in the breast, changes in breast shape, fluid discharge from the nipple or red skin patches [8].

2.1.1 Risk factors

Breast cancer is a multi-factorial disease influenced by a complex interaction of biological, genetic and environmental factors, some of them are described below:

- Age, the likelihood of developing increases with advancing age, with most cases diagnosed in women over the age of 50 [9]. This rise can be attributed to long-term exposure to carcinogens and a decrease in the body's efficiency in repairing cellular damage.
- Genetics plays a significant role in the development of breast cancer, with family history accounting for about 25% of cases [10]. Inherited mutations in breast cancer-

related genes, such as BRCA1 and BRCA2, two anti-oncogenes located on chromosomes 17q21 and 13q12, respectively [11]. The mutation and abnormal amplification of these tumor suppressor genes play crucial roles in the tumor initiation and progression process.

- Reproductive and hormonal factors significantly influence breast cancer risk, particularly through prolonged exposure to endogenous estrogen. This hormone, which stimulates the growth of epithelial duct cells in the mammary glands, is a key factor in breast cancer development. The risk is modulated by several reproductive events: early menarche and late menopause extend the duration of estrogen exposure, increasing the likelihood of estrogen receptor-positive (ER+) breast cancer. Additionally, nulliparity or delayed childbearing until an older age can elevate the risk. Hormone replacement therapy (HRT) during menopause involves a high exposure of exogenous estrogen that produces a hormonal disorder increasing the breast cancer risk [9, 12].
- Lifestyle factors, certain habits and behaviors can increase the risk of developing BC. Obesity and overweight, particularly after menopause when the ovaries stop producing estrogens, the adipose tissue produces estrogen, promoting the development of hormone-sensitive breast tumors [13]. In addition, tobacco and alcohol consumption increases estrogen levels and in the case of tobacco, it contains mutagens that damage the DNA of breast cells [11].

2.1.2 Screening Detection Methods

Breast cancer detection is crucial for early diagnosis and successful treatment. Different imaging modalities, including mammography (MG), magnetic resonance imaging (MRI), ultrasound (US), among others that are utilized to identify and characterize breast tumors. Each method will be described below in detail, focusing on their effectiveness, limitations, and relevance to the breast cancer molecular subtypes.

Mammography

Mammography (MG), a low-dose X-ray technique, is a commonly used method for breast cancer detection and recommended for women aged 40 and older. This procedure involves

taking X-ray images in two views, craniocaudal (CC) and mediolateral oblique (MLO) to identify abnormalities within the breast tissue. during mammogram interpretation, radiologists look for lesions, including small lumps and clusters of calcification, which are often undetectable by manual examination [14]. Despite its widespread used, mammography is not a definitive diagnostic method. When suspicious areas are detected, a biopsy is required for a more precise histopathology and molecular analysis [11]. According to its role in subtype differentiation, MG provides an important initial indicators related to imaging features align with specific molecular profiling. For example, MG is able to recognize microcalcifications and subtle abnormalities, which frequently presents in HER2-enriched subtypes, as well another morphological features [15]. With a specificity of 75% and sensitivity of 67,8% [16], its lower sensitivity that can lead false negatives cases, especially in younger women with dense breast that can obscure tumors, necessitating additional screening examinations [17], such as US or MRI [18]. Despite these limitations, this technique is particularly useful in screening for less aggressive, early-stage cancer, contributing to a 30% to 50% reduction in mortality rates [19, 20].

Magnetic resonance imaging

Magnetic Resonance Imaging (MRI) apply strong magnetic fields and radiofrequency signals to produce highly detailed images of soft tissues, making it particularly useful for detecting small tumors in young women at high risk of breast cancer and detection of multifocal and multicentric lesions [21]. Unlike mammography and ultrasound, MRI offers the highest sensitivity for detecting occult cancer [22], with a sensitivity rate of 94.4% but a lower specificity of 26.4% [16]. In the case of its specificity, lower values indicates that lead to increase false positives, but it can vary based on whether the assesment focuses on morphological features or enhancement patters [23]. Furthermore, MRI is instrumental in forming personalized therapeutic strategies for aggressive subtypes, such as HER2-enriched and triple-negative tumors, which often exhibit distinct MRI features, including concentric tumor reduction and rim enhancement [24, 25].

Ultrasound

Also known as ultrasonography (US) or sonography, is a cost-effective and widely accessible screening tool that uses high frequency sound waves to view tissues and organs inside the body, allowing the detection of anomalous tissues as breast tumors, distinguishing between cysts and solid masses [16]. Additionally, US provides high-resolution images of the shape, orientation, internal structure, and borders of breast lesions without exposing patients to ionizing radiation. Nonetheless, a notable limitation of ultrasound is its relatively low specificity of 34%, which can result in many unnecessary recalls and biopsies [26]. In addition, US can measure the echogenicity of tissue, estimate vascularity on Doppler, and evaluating tumor stiffness via elastography, that are highly associated predicting molecular subtypes [27]. However, the sensitivity of US in distinguishing between different breast cancer subtypes remains uncertain, although some studies indicate that it may effectively predict triple negative subtype [28]. Despite these limitations, US remains a good tool for evaluating breast tissue and assessing lymph node involvement, complementing other imaging modalities in comprehensive breast cancer diagnosis and management [29].

2.1.3 Molecular subtypes

Breast cancer is a heterogeneous disease with various subtypes characterized by distinct gene expression patterns and are associated with different clinical features, tumor behaviors and treatment outcomes. In 2000, Perou et al. [30] proposed a classification of molecular subtypes based on gene profiles using microarrays. Initially, the subtypes were divided into four groups: Luminal-like tumors, Her2-enriched tumors, basal-like and normal-like tumors [31]. Subsequently, this classification was expanded to include more specific subtypes through the use of protein biomarkers present in different breast carcinomas and provide information about biological process that drive cancer growth and progression. The immunohistochemistry (IHC) technique allowed the identification and classification of these intrinsic subtypes based on the expression of their hormonal receptors, which are divided into Luminal A, Luminal B, triple negative and Her2-enriched [32]. The clinicopathological characteristics of each of these subtypes are detailed in Table 2.1.

Table 2.1: Clinicopathological characteristics of molecular subtypes of breast cancer.

Characteristics	Molecular subtypes			
	Luminal A	Luminal B	HER2-enriched	Triple Negative
IHC Markers	ER/PR+, Her2-	ER/PR+, Her2-	ER/PR-, Her2+	ER/PR-, Her2-
Frequency	50%	15%	20%	15%
Ki-67	<14%	≥14%	High	High
Prognosis	Good	Middle	Bad	Bad
Therapy	HT	ChT, HT	HT, ChT, Herceptin	ChT, experimental

Luminal A

This subtype occurs most frequently in breast cancer, accounting for 50% of cases. It is characterized by the expression of hormone receptors (ER/PR+) and an absence of expression of Human epidermal growth factor receptor 2 (Her2-). Compared to the other subtypes, this is the one with the best prognosis, responding well to hormonal therapy (HT), which consists in blocking hormonal activity and preventing the progression of cancer cells. In addition, it indicates a low Ki-67 index of less than 14% [33].

Luminal B

It constitutes about 15% of breast cancer cases. Like Luminal A, it presents (ER/PR) positive and low or no presence of the Her2 receptor, which influences its treatment and prognosis, which is intermediate, but less favorable than Luminal A [34]. The suggested treatments in this case are chemotherapy (ChT) and hormonal therapy [35]. This subtype has a higher cell proliferation with Ki-67 greater than 14%.

HER2-enriched

Human epidermal growth factor receptor 2 (Her2) positive or enriched, representing approximately 15% of mammary carcinomas, is characterized by an overexpression of the HER2 receptor and the absence of (ER/PR). This subtype presents a poor prognosis and

a high Ki-67 index, indicating increased cell proliferation that facilitates metastasis [36]. Patients with this subtype respond better to chemotherapy and the use of drugs such as Herceptin, and in certain cases hormone therapy [24].

Triple negative

The triple negative or basal-like subtype, which comprises 15% of breast cancer cases, has a worse prognosis and induces metastasis in the brain and lungs [37], due to the lack of hormone receptors (ER/PR) and (Her2-) expression limiting therapeutic options, so chemotherapy is recommended but experimental studies for new treatments are still underway [38]. In addition, it is often associated with BRCA1 mutations and has a high Ki-67 index with a rapid cell proliferation [39].

2.1.4 Imaging features of Molecular Subtypes

As previously mentioned, each of the subtypes presents not only specific gene expression patterns and biomarkers, but also distinct morphological characteristics of the tumors that can be identified in medical images. In particular, Wu and Ma, 2017 [40] demonstrates how molecular subtypes are associated with specific radiological features that are observed in mammograms and MRI. In the case of mammograms, some of the variables that are considered are the shape of the breast lesion, the shape of the margin, breast density, the presence of calcifications and the presence of structural distortions.

In ultrasound images (US), there are some particular features associated with molecular subtypes were also found, such as mass appearance, mass margins, mass boundary, echogenicity, which can be classified as homogeneous, heterogeneous or complex [41]. In addition, if the tumor present posterior acoustic characteristics, such as shadowing, posterior enhancement [42, 43].

On the other hand, in images obtained by MRI, the differentiating indicators include tumor size and morphology and texture such as rim (edges) enhancement [44–46], and where the tumors are located or clustered for example if there are multi-centered [47].

The imaging features of each of the different breast cancer detection methods for each molecular subtype are summarized in Table 2.2.

Table 2.2: Imaging features of molecular subtypes of breast cancer.

Imaging Modality	Molecular subtypes			
	Luminal A	Luminal B	HER2-enriched	Triple Negative
MG features	Irregular spiculated mass.	Microlobulated, spiculated margin, architectural distortion.	Mass with associated fine-pleomorphic calcifications.	Circumscribed or oval mass.
MRI features	Irregular mass with spiculated margin.	Irregular mass with spiculated margin	Multi-focal or multi-center enhancement	Round or oval mass, rim enhancement.
US features	Irregular hypo-echoic mass with posterior acoustic shadowing.	Irregular mass with angular margin.	Hypoechoic with irregular margins.	Circumscribed margins and posterior enhancement.
Image sample	Figure 2.1	Figure 2.2	Figure 2.3	Figure 2.4

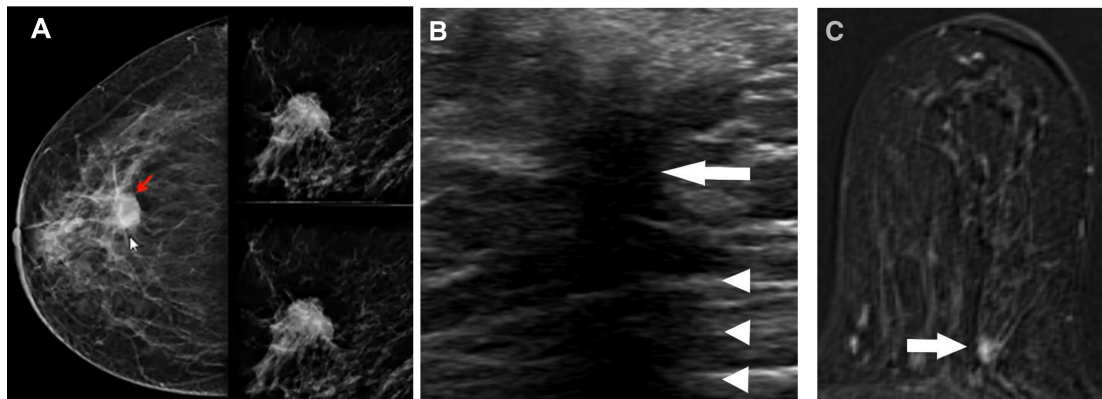


Figure 2.1: Imaging features of Luminal A subtype. A) Mammography shows a mass with spiculated margin, B) US shows hypo-echoic mass with posterior acoustic shadow, and C) MRI shows an irregular enhancing with spiculated mass.

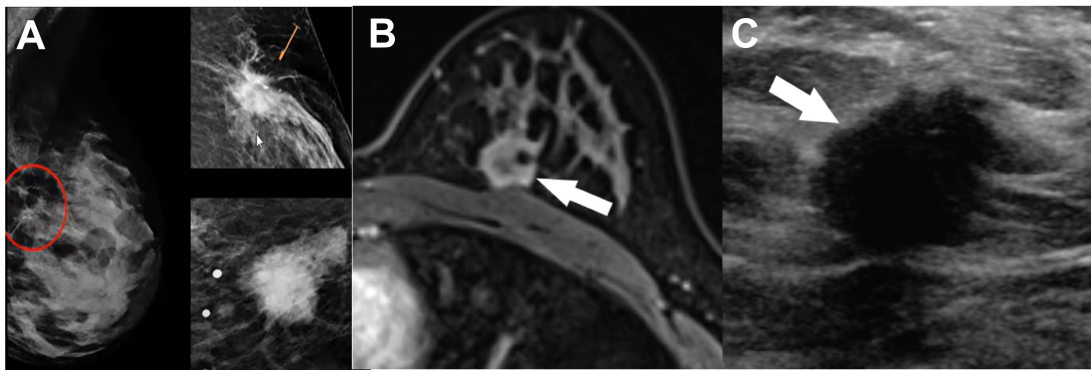


Figure 2.2: Imaging features of Luminal B subtype: A) At mammogram showed a microlobulated spiculated mass with indistinct margins and architectural distortion, B) on MRI there is an enhancing irregular mass with spiculated margins , and C) on US a hypoechoic with irregular mass is observed.

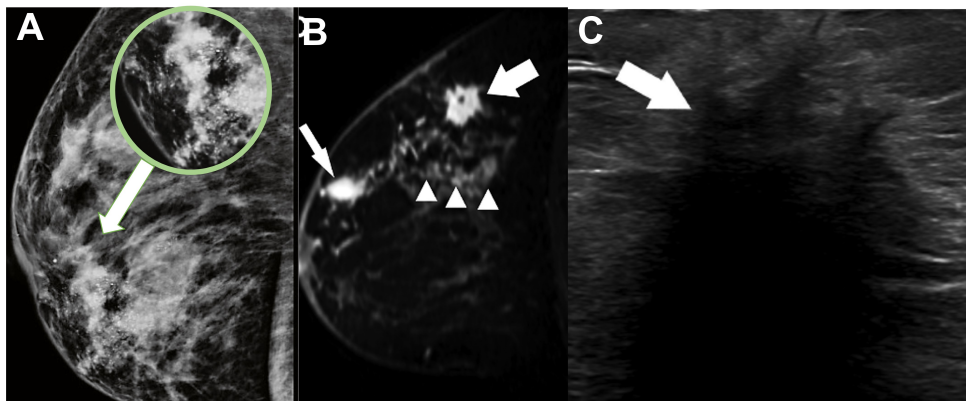


Figure 2.3: Imaging features of HER2 subtype. A) The mammography shows a mass with fine linear calcifications, B) MRI shows a non-mass multi-focal enhancement, and C) on US, the mass is hypoechoic with irregular margins.

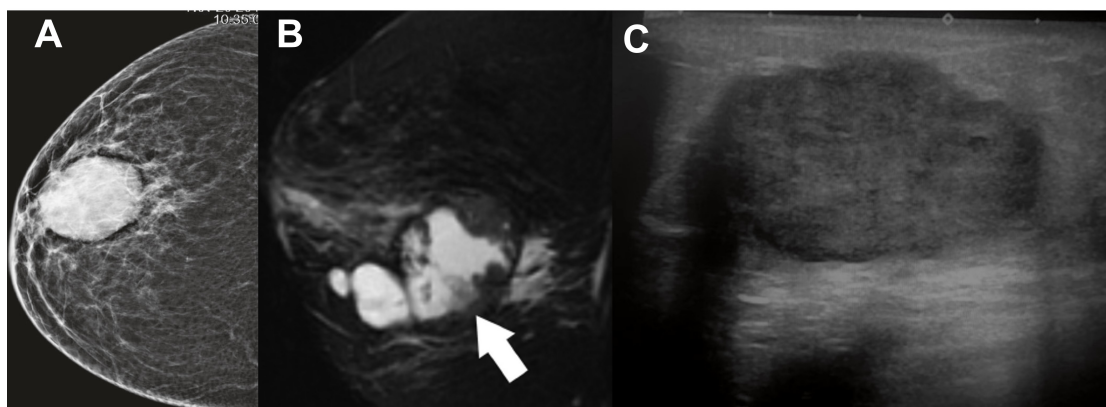


Figure 2.4: Imaging features of Triple Negative subtype. A) circumscribed oval mass is on MG, B) a round mass with rim enhancement is observed on MRI, and C) on US reveals hypoechoic and round mass.

2.2 Artificial Intelligence

Artificial intelligence (AI) is a field of computer science that consist of the development of “intelligent agents” that mimics the human intelligence in order to perform tasks such as learning, decision-making, recognizing patters among others cognitive abilities within a machine framework and to perform them is necessary the execution of a series of AI algorithms. The term of AI was coined with Alan Turing in the 1950s in this book *Computers and Intelligence*, who proposed the “*Turing Test*” as a way to measure of a machine’s intelligence [48]. The development of Artificial Intelligence (AI) in the last years has a great impact and revolutionize the world, and has brought great advances in the health care field. AI encompasses machine learning and deep learning.

2.2.1 Machine Learning

Machine Learning (ML) is a branch of artificial intelligence, that focuses on training algorithms and models that can analyze data, make predictions and decisions using this data. These algorithms can enhance their efficiency in a particular task as they encounter increased large data. ML is commonly divided in four types, that are explained below:

- Supervised Learning, training data contains optimal outcomes (also known as inductive learning). Learning is tracked in this method. Algorithms are trained using labeled examples, such as an input where the desired output is known.
- Unsupervised Learning, There are not the desired outputs in the training results. The goal is to find hidden patterns or intrinsic structures in the input data, where the data does not have historical label. Unsupervised learning models examine data without human supervision, in contrast to supervised learning, which depends on labelled data with specified target values.
- Semi-supervised Learning, is when a few desired outputs are included in the training data. This type of learning combine labeled data (including explicit target values) and unlabeled data (no targets). Semi-supervised learning finds a balance by using the labelled instances that are already available and looking for patterns in the unlabeled data. This makes it helpful in situations when getting enough labelled data is difficult.

- Reinforcement Learning, where rewards are given after a sequence of actions. In a given case, it is a matter of taking appropriate steps to maximize compensation. It is the most ambitious method of learning in AI and used in the robotics field. It consists of three components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do) [49].

2.2.2 Deep Learning

It is a subset of ML, deep learning (DL) employs multi-layer deep neural networks and a large-scale data to learn automatically. The DL emerged thanks to advances in artificial neural networks (ANNs), which were introduced by Frank Rosenblatt [50], who in 1958 introduced the term of “*perceptron*”, a simple learning model that could learn new skills through a process of trial and error. The development of ANN was inspired by the biological functioning of the human brain [51]. The brain is composed by neurons which are responsible for receiving sensory information (signals) from the external world through dendrites, then processes the signals in the cell body, and transmits the information as output through the axon [52]. As illustrated in Figure 2.5.

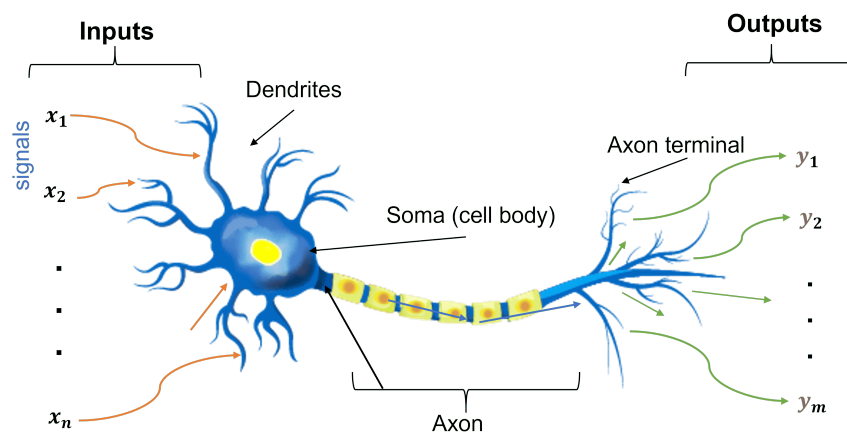


Figure 2.5: Biological neural network and their relationship with artificial neural network.

Similarly, a neural network receives multiple inputs, process them, applies a weighted sum followed by a non-linear function (activation function), and generate outputs. Structurally, an ANN consists of three layers of interconnected nodes or neurons, as illustrated in Figure 2.6 and detailed below:

- The input layer is analogous to independent variables in a mathematical model, serving as the initial data points that the network processes. It receives and process input information from external sources, transforming it into a format that the network can utilize.
- The hidden layers, are intermediate layers between the input and output layers in the neural network. These layers are able to extract complex features from the input data, transforming it into a more abstract, intricate and meaningful representation. They are the primary drivers of the network's learning capability, which enables to adapt and enhance performance over time.
- The output layers, on the other hand, is the final layer, and it is responsible for producing the prediction or results for a given input. These layer takes the output from the hidden layers and generates a prediction or classification based on learned patterns in the data.

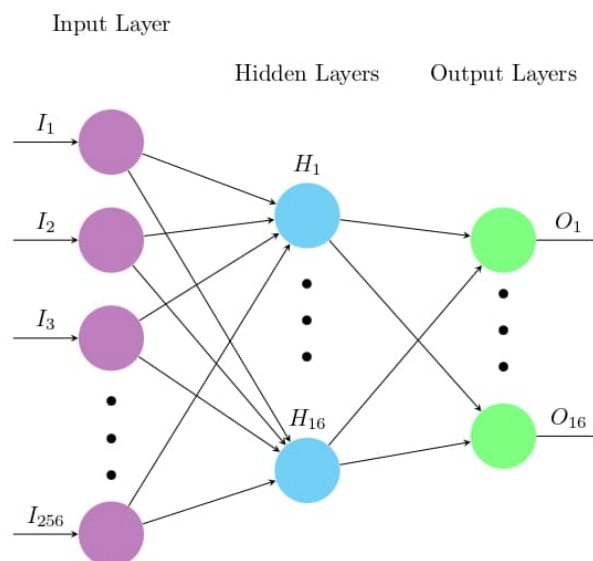


Figure 2.6: Structure of an Artificial Neural Network (ANN)

2.2.3 Convolutional Neural Networks

A convolutional neural network (CNN), also known as ConvNet, is a type of deep learning that specialized in processing and analyzing visual data, especially images. They are

feedforward networks, which means that information flows in one direction, from input to output. Inspired by the human visual cortex, CNNs process this data in a grid-like arrangement in order to extract important features and recognize patterns automatically [53]. They exhibit translation-invariant characteristics, making them effective for tasks like image classification, object detection, and segmentation [54].

Structure of CNN

The architecture of a CNN consists of several building blocks. The key components and how they are organized to form the network are described below:

- Convolution layer, is the first layer and most essential component of a CNN. This layer contains filters also known as kernels that are applied to the input images to extract features such as edges, textures, and specific patterns. A kernel is a matrix of weights that slides over the input image, performing a convolution operation, which is mathematically defined in the equation (2.1) [55]. The input data is combined with the learned weights to compute a new feature map, and the resulting convolved values are processed through a non-linear activation function [56].

$$\text{conv}_{x,y} = \sum_i w_i x_i \quad (2.1)$$

Where:

w_i are the kernel weights associated with the position i ,

x_i are the values of the spatially corresponding the input image,

$\text{conv}_{x,y}$, is the result represents the output of the convolution operation at the spatial coordinates (x, y) .

- Pooling layer (Downsampling), reduces the spatial dimensions (width and height) of the feature maps while retaining important information that helps the network become more robust to variations in the input and reduce the computation complexity. Besides, the pooling layer is responsible for combining semantically similar features into one.
- Fully connected layer, also known as a dense layer, is responsible for making the final

predictions or classifications based on the features extracted by the previous layers applying a series of weights and biases.

Activation functions follow the convolution and fully-connected layers in CNNs to make the network be able to learn non-linearity. ReLU (Rectified Linear Unit), TanH and sigmoid function are some examples of activation functions. These activation functions also allow the network to extract complex relationships between input data and learned features so that better predictions or classifications can be made [57].

During the training of a CNN, the weights inside the kernels are adjusted through the process of backpropagation and optimization algorithms in order to select appropriate hyperparameters [58]. Some commonly used optimization techniques include Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam). These algorithms dynamically update the learning rate using gradient momentum, facilitating efficient training.

The structure of a typical CNN is shown in the Figure 2.7, from the image input to the output layer.

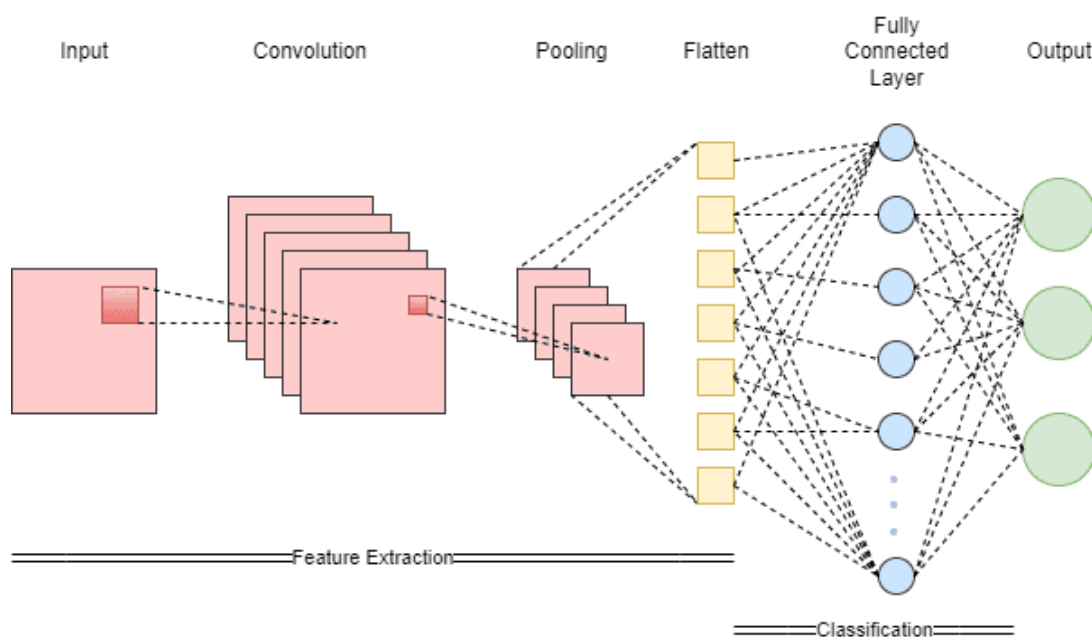


Figure 2.7: General structure of a typical convolutional neural network (CNN) for image classification.

CNNs for Image Classification

In medical image classification, CNNs are widely used to identify patterns that indicate the presence of anomalies. Some of the most widely used architectures in this field include:

- AlexNet, considered the pioneering CNN, consists of eight layers: five convolutional layers and three fully connected layers. The first layer has 96 kernel of size 11 x 11 with a stride of 4. It was the first to introduce the ReLU activation function and dropout regularization to prevent overfitting. This architecture significantly contributed to the advancement of deep learning in computer vision by winning the 2012 ImageNet competition [59]. Input size: 227x227x3.
- ResNet50, is characterized by its 50-layer depth and the use of residual connections to extract complex features from images, solving the vanishing gradient problem. ResNet50 set a new standard in deep learning by enabling very deep networks to converge [60]. Input size: 224x224.
- MobileNet, uses depthwise separable convolutions, which split the convolution operation into two steps: depthwise convolution and pointwise convolution, that significantly reduces the number of parameters, and making it useful for working with transfer learning [61]. Input size: 224x224.
- VGG CNNs, are known for their simplicity, utilizing a uniform architecture based on 3x3 kernels and follow two basic principles: maintain a constant number of filters for a given output feature map size and, double the number of filters when the feature map size is halved by convolutional layers with stride 2, which keeps the time complexity per layer constant [62]. There are two versions: VGG16 and VGG19, named according to the number of layers they contain. VGG16 consists of 13 convolutional layers and 3 fully connected layers, while VGG19 has additional 3 convolutional layers. Input size: 224x224.
- InceptionV3, is a groundbreaking deep learning model designed by Google researchers based on GoogLeNet [63]. This is used to optimize the depth and width of neural network without increasing the number of parameters. This model consists 48 layers

that employs Inception modules, which utilize multiple parallel branches to capture features and introduces factorized 7x7 convolutions, reducing computational complexity while maintaining efficiency. Input size: 299x299.

- Xception, is an improved version of the previous architecture, using 36 depthwise separable convolutional layers instead of Inception modules to enhance efficiency while using fewer parameters. It also offers an innovative approach, combined with linear bottleneck layers, and has 71 layers. Input size: 299x299.

Chapter 3

State of the Art

Nowadays, the development of Deep Learning models has revolutionized the diagnosis, detection and prediction of complex diseases such as breast cancer. This state of the art reviews the main recent research and developments in the application of innovative models in the detection of this disease, as well as the classification of its molecular subtypes using mammography images.

3.1 Deep learning-based model for detection of breast cancer

Deep learning (DL) models for breast cancer detection, described in Table 3.1, mainly use convolutional neural networks (CNNs), which focus on classification and segmentation of medical images. Firstly, Sheen et al. [4] propose a CNN-based model that achieves an AUC of 91%. This model employs the “end-to-end” training technique, which allows it to learn directly from the raw data with minimal reliance on detailed annotations of breast lesion regions of interest (ROIs) after the initial phase, thus facilitating their efficient classification. The model was fine-tuned from the VGG16 and ResNet50 architectures, using the public dataset as INbreast and CBIS-DDSM. The combination of both CNNs can enhance the model’s ability to generalize patterns. They also applied data augmentation to improve their AUC of 88% to 91%. Similarly, Li et al. [64] performed ROI extraction of malignant, benign, and normal cases using digital mammography and digital breast tomosynthesis (DBT) images (3D-mammograms), for subsequent multimodal classification.

To do so, the researchers used VGG16 architecture with transfer learning for both imaging modalities. The study found that DBT-based model had significant performance than using digital mammography, with accuracy rates of 91% for malignant, 95.4% for benign, and 98.4% for normal tissues. On the other hand, the custom model developed by Mahmood et al. [65] offers promising results for breast tumors classification, with an accuracy of 98% , high sensitivity of 99% and AUC of 99%. This model consists of three Conv layers, followed by a max-pooling and a two-category softmax classifier, corresponding to benign and malignant tumors. The dataset used consists of mammographic images from MIAS and private datasets, which was improved using pre-processing techniques such as contrast-limited adaptive histogram equalization (CLAHE), noise reduction with median, Gaussian, and bilateral filters, and segmentation using Otsu threshold.

Another proposal, based on a different CNN, was developed by Omonigho et al., [66] utilizing modified Alexnet reducing to five Conv layers and softmax activation for binary classification (benign vs. malignant tumors). Also, the use of data augmentation effectively increased the datasets size from 322 to 2576 images using transformations such as flipping, rotation and enhancement using Gaussian filters to improve model performance, achieving an accuracy of 95.70% for classifying benign and malignant lesions. Likewise, Ragab et al., [59] employed pre-trained AlexNet architecture with the ML technique, support vector machines (SVM) for the development of a computer-aided detection (CAD) system, this implementation improved the accuracy to 87.2% and AUC of 94%. In addition, CLAHE was used as a mammography image processing method, and tumor ROIs were manually extracted and segmented using threshold technique. In contrast to previous studies, this model has a lower accuracy, which is less than 90%, the authors suggest is due to the quality of dataset. In the case of segmentation of abnormal lesions and their subsequent automated classification, using CC and MLO views of mammogram separately. Salama and Aly [67] propose an implementation of U-net segmentation model together with CNNs such as InceptionV3, DenseNet121, ResNet50, VGG16 and MobileNetV2. However, InceptionV3 demonstrated the best results with an accuracy of 98.87%. Overall, the literature highlights the importance of applying pre-processing and data augmentation techniques to improve the generalization of the model to detect and classify patterns. In addition, the different CNNs commonly used in cancer detection are presented.

Table 3.1: Related works on deep learning models for breast cancer detection that use mammographies.

Perform task	Architecture	Dataset	Size(n)	Class	Performance metrics	Ref.
Classification	VGG16,	INBreast, CBIS-	155,	Benign,	Acc, AUC	[4]
	ResNet50	DDSM	2478	malignant		
Classification	VGG16	Private	927	Benign, malignant, normal	Acc, AUC	[64]
Classification	Custom model	MIAS, Private	322, 580	Benign, malignant	Acc, AUC	[65]
Classification	AlexNet	MIAS	322	Benign, malignant	Acc	[66]
Classification	AlexNet, SVM	DDSM, CBIS- DDSM	2620	Mass, calcifications	Acc, AUC	[59]
Segmentation, Classification	U-net, InceptionV3	DDSM, MIAS, CBIS-DDSM	564, 322, 330	Benign, malignant	IoU, DC, Acc, AUC	[67]

3.2 Deep learning-based models for identification and classification of molecular subtypes

Some previous studies on the classification of molecular subtypes using mammograms are described in Table 3.2.

Firstly, a TL model based on the ResNet-18 architecture for binary classification of luminal (Luminal A and Luminal B) and non-luminal (HER2-enriched and triple negative) subtypes proposed by Bhandary et al., [68]. This model uses full mammography images and employs the 5-fold cross-validation technique, highlighting the use of hyperparameters such as a batch size of 16, a learning rate of 1×10^{-5} for the Adam optimizer, and a weight decay of 5×10^{-3} . This approach achieved an AUC score of 66.88% and F1-score of 66.93%. Some limitations that authors mentioned about their results, is lack of detailed information about morphology, size of breast lesions and class imbalance, especially with predominance of Luminal subtype.

On the other hand, Mota et., [69] explain how to classify molecular subtypes when there is an imbalance of data, which can cause issues in classification since the class with a higher number of images will have a preference. To address this, they compare methodologies such as oversampling (increasing data to balance), undersampling (reducing data from the class with more images), and data augmentation by generating images with geometric transformations. Besides, the authors proposed two types of classification, including binary (one subtype against all others) and multiclass (evaluating all subtypes simultaneously), the former achieving an average accuracy of 79.02% and an AUC of 64.69%, while the multi-class classification achieved an average AUC of 60.62% with oversampling and data augmentation. The model in this study is based on pre-trained Resnet-101 CNN. The dataset OPTIMAM in this study contained information about tumor locations, which allowed for the extraction of the ROI (Region of Interest).

The incorporation of multi-modal models, which use two different types of data for classification, is very promising. An innovative approach proposed by Zhang et al., [70], who developed an inter-modal DL model (MDL-IIA) that used ResNet50 as base model with intra- and inter-modality attention modules, in order to determine patterns in both ultrasound and mammogram images. For this, it was necessary to crop the breast lesions.

The MDL-IIA achieved a high accuracy of 88.5% and a Matthews correlation coefficient (MCC) of 88.7%. This novel model compared to previous researches show a high performance in the prediction of molecular subtypes, due to its large dataset, multimodal integration, and modification of CNN with attention models.

Finally, DL techniques can also be used in the determination of hormone receptors such as HER2, ER and PR that define molecular subtypes. Ueda et al., [71], used cross-validation for their classification model, which basen on four architectures: VGG16, InceptionV3, ResNet52 amd Dense121. To determine which models perform better, the author used an ensemble method with the best models for each receptor. For each biomarker, an AUC 67% was obtained for ER, 61% for PR and 75% for HER2. Using for each one these ensemble Inception + ResNet for ER , VGG16 + Inception for PgR and HER2 biomarkers.

Therefore, the evidence found related works suggest that development of deep learning models to predict and classify molecular subtypes using mammograms are promising. Nevertheless, there are some limitations in current studies that need to be addressed. For example, limited databases and data imbalance. In addition, complete information on clinical data of breast lesions should be included.

Table 3.2: Related works on deep learning models for classification and prediction of molecular subtypes.

Model	Dataset	Size (n)	Pre-processing	Class	Performance metrics	Ref.
ResNet18	CMMD	1498	AUG	Luminal, Non-luminal	AUC, MLMC	[68]
ResNet-101	OPTIMAM	1397	Crop ROI tu- mor, AUG	LumA, LumB1, LumB2, HER2, TN	Acc, AUC	[69]
MDL-IIA	Private	3360	AUG	LumA, LumB, HER2, TN	Acc, AUC, MCC	[70]
VGG16, ceptionV3, ResNet52, DenseNet1211	In- Private	1448	Bounding box crop	ER,PR, HER2	AUC	[71]

Chapter 4

Methodology

This section details the methodology used to develop a DL model for detection of BC and the classification of its molecular subtypes from mammography images. The process consists of several steps: data collection and pre-processing, segmentation, model training and evaluation. Figure 4.1 presents the overall flowchart of the methodology proposed.

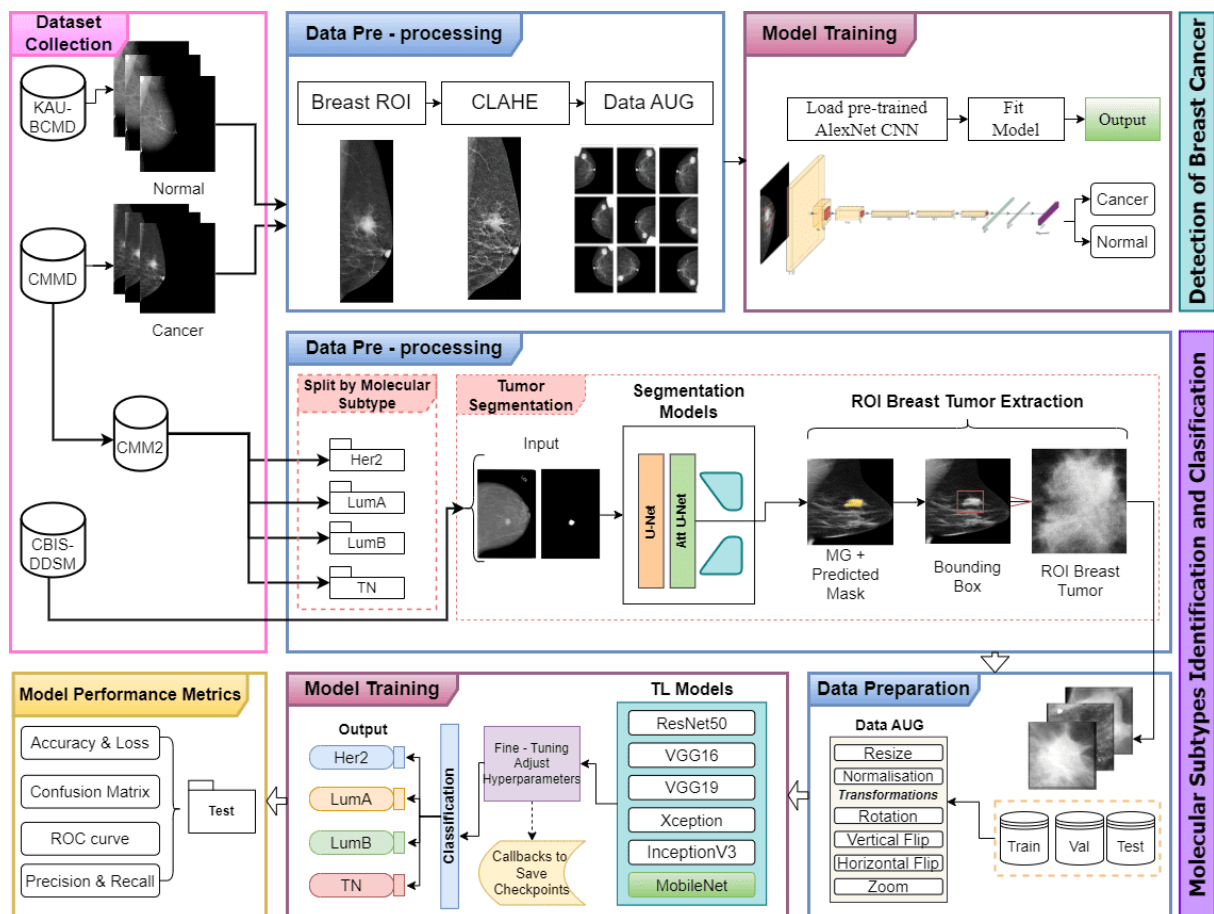


Figure 4.1: Flowchart of the proposed methodology.

4.1 Dataset description

Three mammography databases were used in this research, the data specifications of each one is summarized in Table 4.1 , which is described in detail below.

Table 4.1: Summary of the datasets used, with their specifications.

Dataset	CMMD		KAU-BCMD	CBIS-DDSM
	CMMD1	CMMD2		
Number of cases	1026	749	1416	1566
Number of images	2214	1498	5662	10239
Image type	DICOM		JPG	DICOM
Resolution	1914 x 2294 px		2816 x 3583 px	Varied
Categories	Benign Malignant	Her2-enriched Luminal A Luminal B Triple negative	BIRAD1 BIRAD2 BIRAD3 BIRAD4 BIRAD5	Benign Malignant
Size	22.86GB		607 MB	163.51GB
Approach	Breast cancer detection	Molecular subtypes classification	Breast cancer detection	Tumor segmentation
Ref.	[72]		[73]	[74]

4.1.1 CMMD

The Cancer Imaging Archive (TCIA) [\[75\]](#) provides The Chinese Mammography Database (CMMD) [\[76\]](#) contains 3728 mammographies from 1,775 patients (mean age: 47.56 years; range 18-87 years) from China with benign and malign cases from the left and right sides of the breast with CC and MLO views as it is shown in Figure 4.2.

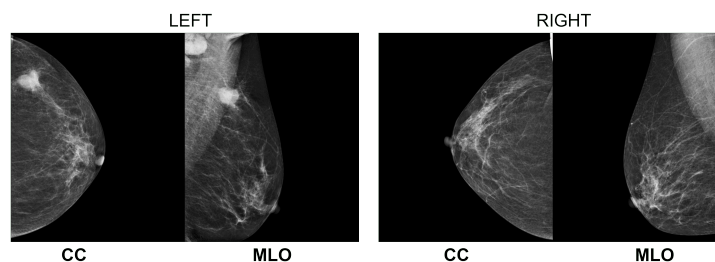


Figure 4.2: CC and MLO views of the left and right breasts of sample mammogram of CMMD dataset.

The database is divided into CMMD1 and CMMD2. CMM1 contains 1026 cases (2214 mammograms) that are distinguished between benign and malignant, on the other hand CMM2 includes 749 cases (1498 mammograms) that include malignant cases divided by molecular subtypes as follows:

- Luminal A: 152 cases (51.39 ± 11.21) years.
- Luminal B: 376 cases (49.30 ± 10.82) years.
- HER2-Enriched: 135 cases (50.21 ± 10.16) years.
- Triple negative: 86 cases (48.67 ± 10.79) years.

The distribution per class is shown in Figure 4.3.

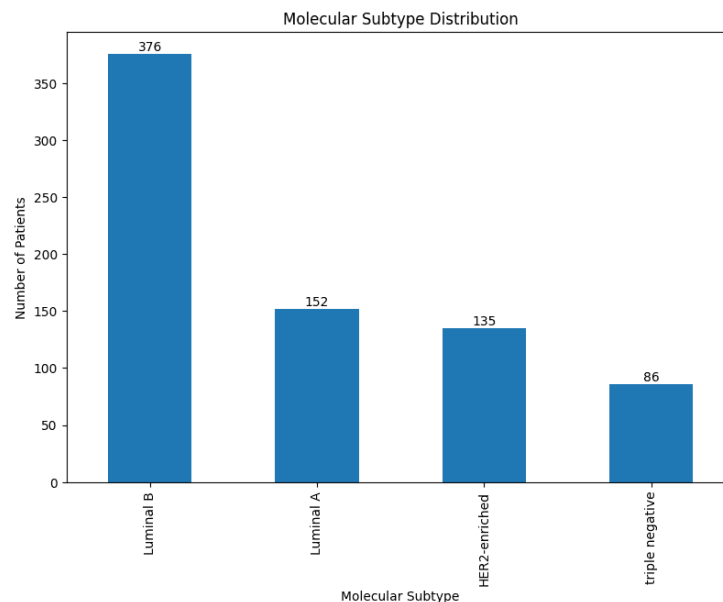


Figure 4.3: Distribution of molecular subtypes per class in CMMD2 dataset.

4.1.2 KAU-BCMD

King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD) [73], contains 1416 cases including 5662 DICOM mammograms with CC and MLO views of the left and right breast. It is divided according to BI-RAD categorical from BI-RAD0 to BI-RAD5. However, for this specific study, only cases classified as BI-RADS 1 were used, corresponding to patients who do not present suspicious lesions in the breast that includes 1865 mammographies.

4.1.3 CBIS-DDSM

The Curated Breast Imaging Subset (CBIS-DDSM) is a updated public dataset of mammograms extracted from the Digital Database for Screening Mammography (DDSM) containing 1566 cases with 62644 images of both breasts in both CC and MLO views [74]. Each image includes precise information about the tumor region of interest (ROI) and binary mask images that delimited the general position of the tumor. The database is divided by their abnormality type into mass and calcification, where tumors are labeled by their lesion type into benign and malignant. For this study, only cases with malignant tumors (1176 images) were selected, since the molecular subtypes to be analyzed correspond to malignant.

4.2 Detection of Breast Cancer

4.2.1 Data Preparation

The mammograms that are in DICOM format had been converted to Portable Network Graphics (PNG) format. The datasets were splitted into training set (1300 images) and the validation set (200 images) per class. Normal cases were labeled with 0 and cases with carcinoma with 1. The input images were normalized by dividing each pixel value by 255, scaling the values to a range of [0, 1].

4.2.2 Data Pre-processing

To improve the quality of mammograms images and highlight relevant features of abnormal lesions, for the detection of breast cancer. The procedure used in data pre-processing in this study is summarized in Figure 4.4. The techniques used are described in detail below.

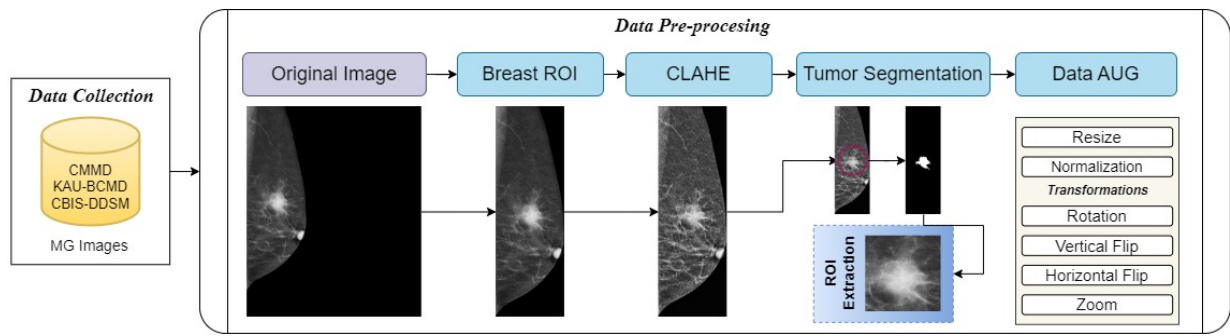


Figure 4.4: Block diagram for data pre-processing.

ROI Extraction of Breast

It is important to obtain a region of interest (ROI) that focuses on the breast area of mammograms, because most of these are large in size and only 30% corresponds to breast tissue. Analyzing the whole-mammogram can complicate the detection of breast cancer [65]. This procedure reduces noise and avoids analyzing irrelevant areas. Therefore, mammograms were processed in grayscale and the external contours are detected. Then, the contour with the largest area is selected, assuming that this corresponds to the breast, and a rectangle is calculated to delimit it.

Finally, the coordinates of the rectangle are adjusted to ensure that they are within the image limits, obtaining the breast region of interest (BROI). Thanks to this technique, the model will be focus on specific patterns of breast tissue, facilitating the identification of breast lesions. The algorithm 1 for this procedure is found in appendix section.

CLAHE

Contrast-Limited Adaptive Histogram Equalization (CLAHE) improves image analysis to adjust contrast in small regions, allowing the edges of tumors to be highlighted and easier to identify [77]. This method is especially useful for segmenting tumors in dense breasts, where edges of suspicious abnormalities can be difficult to distinguish [78]. The specific parameters used in the implementation of CLAHE were the clipping limit for histogram equalization '*clipLimit*' of 2.0, and the grid size '*tileGride*' of (12x12).

The algorithm 2 for this procedure is found in appendix section. For these computer vision procedures, it was necessary to use the '*CV2*' library of Python.

The mammogram processed with both previous techniques is shown in Figure 4.5.

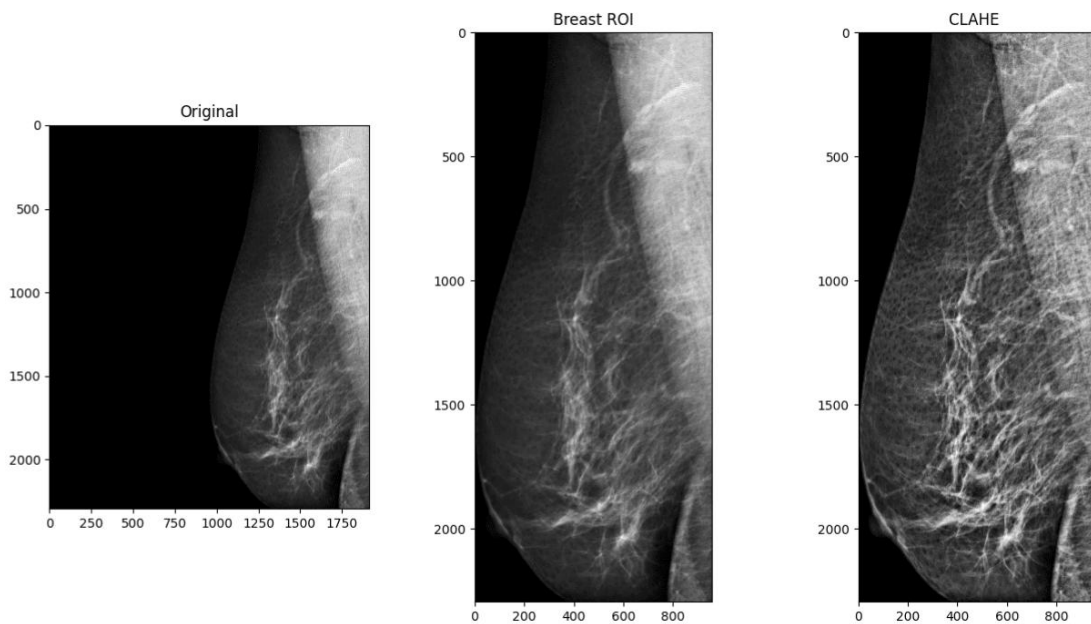


Figure 4.5: Breast ROI cropping and CLAHE application for mammogram image processing.

Data augmentation

For training deep learning models, it is crucial to have a robust data in order to increase the diversity of the training set and improve the ability of the model to generalize to new data and avoid overfitting. In this study, the augmentation technique is applied using ImageDataGenerator implemented in Tensorflow. This process included the following transformations:

- Rotation: 20°
- Shear range, probability of 0.2
- Horizontal flip, probability of 0.2
- Width shift, probability of 0.1
- Zoom range, probability of 0.2

4.2.3 Proposed model based on AlexNet

For breast cancer detection, the pre-trained AlexNet model was used, which is a CNN architecture very useful in image classification. However, some modifications were made in

the final layers of the network, to adapt it to a classification of normal and cancer cases. First, the activation function of the last classification layer was replaced by the *sigmoid* function, for this binary classification. The last FC containing 4096 neurons was reduced to 1024 neurons, which allowed to reduce the dimensionality of the representation. In Figure 4.6 the proposed model based on AlexNet architecture is shown.

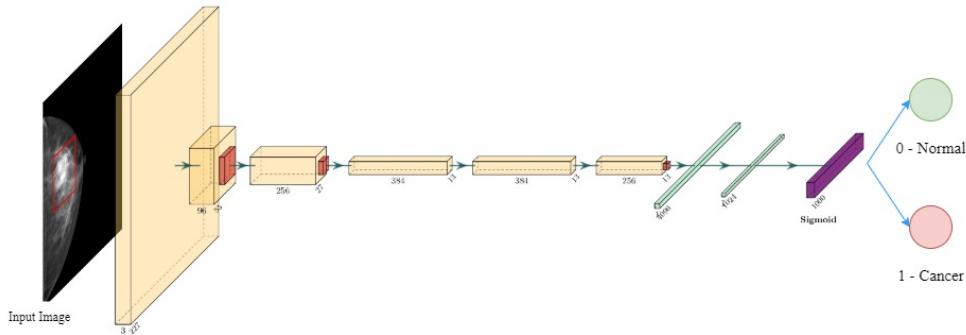


Figure 4.6: AlexNet model architecture proposed for breast cancer detection.

4.3 Identification and classification of breast cancer molecular subtypes

4.3.1 Data Preparation

In this section, stratification and division by molecular subtype was performed using the CMMD2 database. For this, the CMMD manifest and the Excel file containing the patients' clinical data were needed. The code developed by [79], for the classification between benign and malignant tumors, was useful to modify and apply it in the division by molecular subtype. Then, the dataset was randomly divided into Train, Validation, Test using the proportion of 80%, 20%, 20%, the number of images per subtype is shown in Figure 4.7.

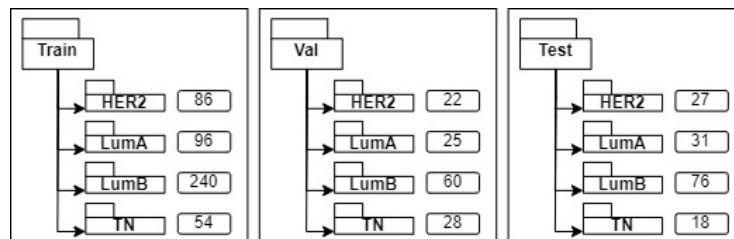


Figure 4.7: Data split into Train, Validation and Test for each molecular subtype.

4.3.2 Data Pre-processing

1. **Mammogram Enhancement** The image processing mentioned before, for CMMD2, the crop of breast region BROI and CLAHE were applied.
2. **Segmentation of Breast Tumor** In this part, two commonly used models for medical image segmentation were trained: U-net and Attention U-net.
 - (a) **U-net model**, is a type of CNN specialized for medical imaging segmentation. It was first introduced in 2015 by Ronneberger et al., [80], its name is based in "U" letter form of this architecture, and consists of two parts:
 - Encoder, the left side that captures context and spatial features, is a typical CNN architecture with repeated application of convolutions, ReLU activations, and max-pooling operations. As the path progresses, the spatial dimensions of the feature maps reduce while the number of feature channels increases.
 - Decoder, the right side, which aims to localize the objects, consists of up-convolution or transposed convolution layers that upsample the feature maps. It also includes concatenations with the corresponding feature maps from the encoder path (skip connections), which helps in retaining spatial information loss during downsampling.
 - (b) **Attention U-net model**, is an enhancement of the original U-Net architecture that incorporates attention mechanisms to improve segmentation performance and was developed by Oktay et al., [81]. They add attention gates (AG) in the skip connections of decoder path, that are capable to learn and focus on the most important features in the input data, and increase model accuracy and sensitivity, the model architecture is shown in Figure 4.8.

The training was based on an improved version of the previously mentioned models. Bhattiprolu [82], modified some parameters such as implementing a batch normalization on channels after a convolution, and the loss function used was Binary Focal Loss, since it is a binary classification, the *sigmoid* activation

function was used. The input size was 224x224x3, Batch size 8 and it was iterated for 50 epochs.

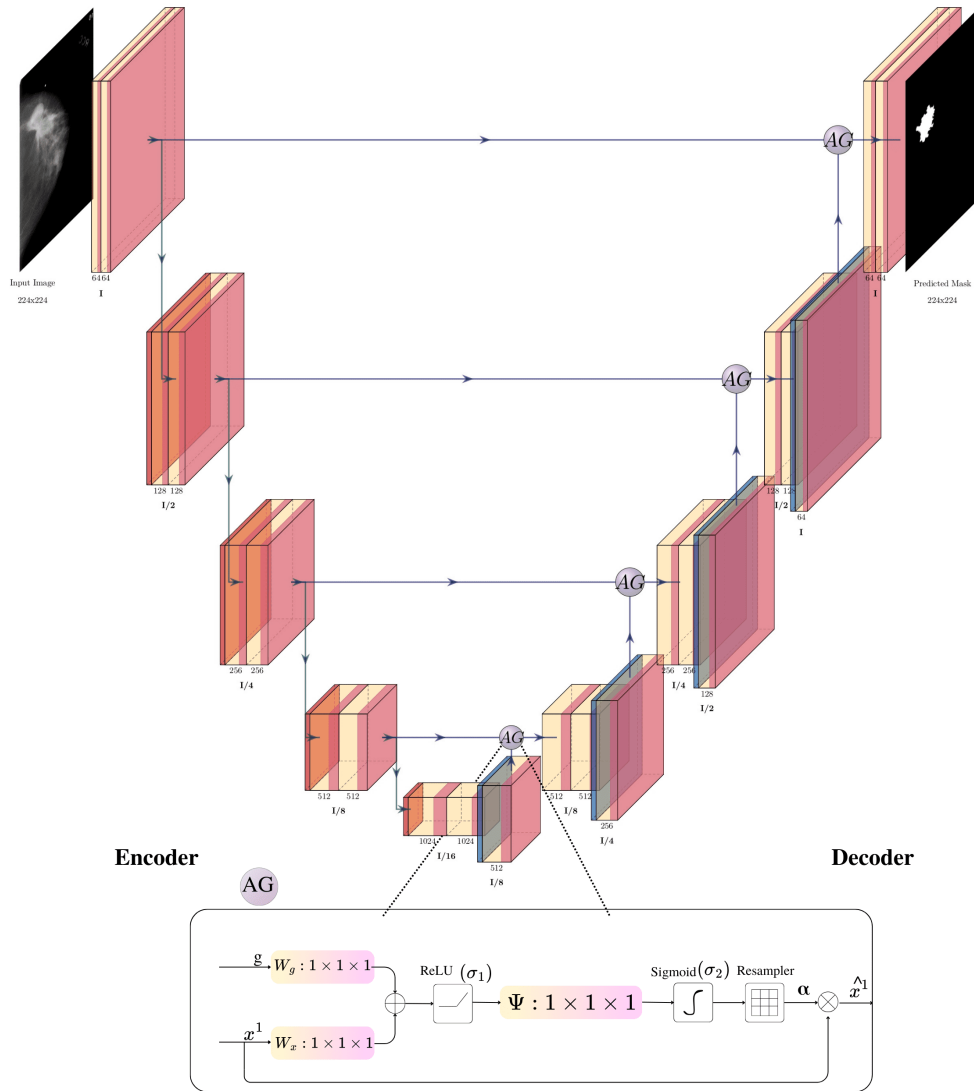


Figure 4.8: Attention U-net model architecture.

3. ROI Extraction of Breast Tumor

The segmentation model with the best performance will be used to obtain a binary mask (predicted) that shows the location of the breast tumors corresponding to each molecular subtype. To extract the tumor ROI, the Bounding Box technique used in object detection was used to determine the spatial location of the breast lesions.

For this process, which is explained in Figure 4.9, a function was designed in MATLAB, where the mammogram and its corresponding mask generated previously are

loaded. Then, the mask is verified as binary; if it is not, it is converted to binary using *'imbinarize'*. Then, the *'BoundingBox'* property that encloses the object (tumor) is searched for in the mask region, the coordinates and dimensions of the bounding box are extracted, and the tumor is cropped. Then, the images obtained were resized to 224x224 and converted to RGB (3 channels). The algorithm 3 of this process was applied to the entire CMMD2 dataset using the Image Batch Processor application from MATLAB R2023b as a function. The output images will be the input data for molecular subtyping classification.

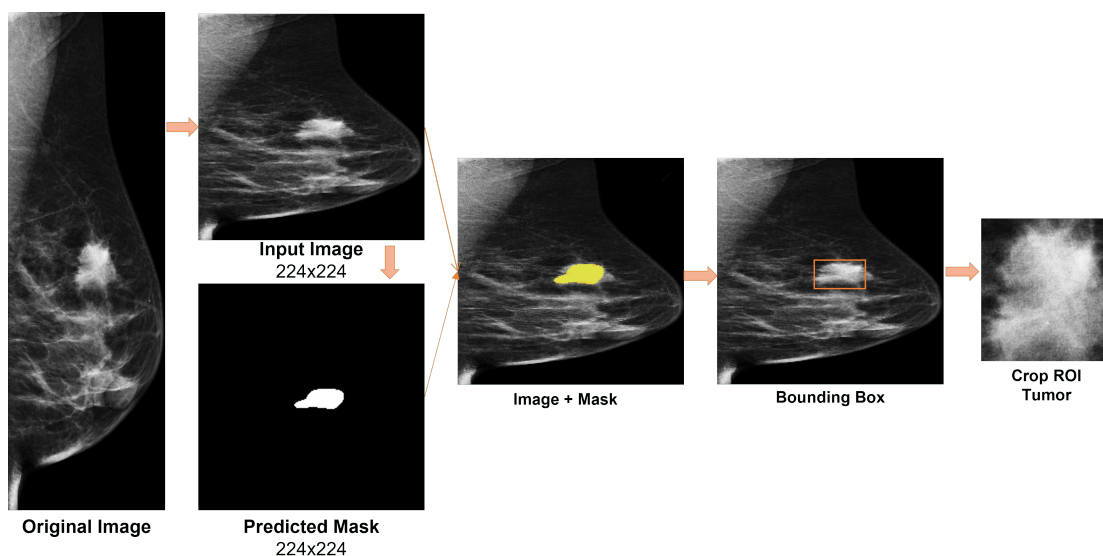


Figure 4.9: Breast tumor ROI extraction process.

4.3.3 Transfer Learning Models

Transfer learning (TL), is an efficient approach that allows improving performance in a new task by taking advantage of the knowledge acquired in a previous task [83]. This method is particularly useful when the data set is limited. Transfer learning consists of two stages: pre-training and fine-tuning. First, the model is trained on a general task, such as with ImageNet for 2D images, and then adjusting its weights to fit the new task, allowing it to adapt to new patterns and relationships in the data [84].

The TL models used for molecular subtype classification were selected according to the state-of-art of this study, these are: ResNet50, VGG16, VGG19, InceptionV3, Xception, and MobileNet. After this, fine-tuning was performed to adjust the hyperparameters and

retraining the network layers for the classification of four classes. Therefore, the last fully-connected layer configured for 1000 classes was replaced by a dense layer containing 4 neurons (4 molecular subtypes), and a 'softmax' activation function ideal for multiclass classification was implemented.

4.4 Experimental Setup

4.4.1 Software and Hardware

Training a deep learning model on a powerful hardware configuration combined with specialized software. The processor used is a 12th generation Intel® Core™ i7-12650H and GPU NVIDIA RTX 3070 of 8GB, which allows to accelerate calculations, in particular matrix operations and parallel processing, which are vital for training neural networks. To take advantage of the performance of this GPU, the CUDA Toolkit version 11.2 was used. According to the programming environment, Python 3.10 was used, which provides extensive libraries, mainly TensorFlow Keras 2.10, which provides high-level APIs to create and train complex models efficiently and quickly. In addition, some of the essential libraries used were the following: *Numpy*, *Pandas*, *Scikit-learn*, *Seaborn*, *OpenCV*.

4.4.2 Experiments for Breast Cancer Detection

- i. **Experiment 1: Comparison between the previous AlexNet model proposed and the fine-tuned AlexNet model.**

In addition, the impact of the image processing techniques used in this study on mammography images will be evaluated to improve the model performance. The proposed model by [85] previously presented overfitting problems, the pre-processing methods such as data augmentation were used, where Gaussian noise and morphological filters such Median blur filter and erode were employed randomly in order to avoid overfitting.

The training setup and the hyperparameters used in this experiment for new model proposed is:

- Batch size: 64
- Epoch: 20

- Dropout: 0.5
- Input size: 227x227x3
- Activation function: Sigmoid
- Loss: binary crossentropy
- Learning rate (LR): 0.0001

Additionally, a model Checkpoint was used to save the best model and an EarlyStopping with patience of 5 so that when the model starts to overfit, the training will be stopped.

4.4.3 Experiments for Molecular Subtypes Classification

- i. **Experiment 1: Comparison of Transfer Learning models (ResNet50, VGG16, Xception, InceptionV3) using BROI images.**

All training images use ImageDataGenerator for data augmentation, the transformation will be the same as experiment for BC detection. All input data will be rescaled and normalized to 1./255. Moreover, BROI images will be resized according to the input size for each convolutional neuronal network.

The training settings are:

- Batch size: 32
- Epoch: 50
- Input size: depends on each architecture
- Activation function: Softmax
- Loss: categorical crossentropy
- Learning rate (LR): 0.001
- Dropout: 0.5
- L2 Regularizer: 0.001

- ii. **Experiment 2: Tumor segmentation using Unet and Attention U-net.**

The process of selecting the best model for breast tumor segmentation is illustrated in

Figure 4.10 Using the CBIS-DDSM malignant cancer mammogram dataset and their respective masks, two segmentation models will be trained, then the evaluation is carried out with performance metrics. If the model has an Acc > 90% and Jaccard index >80%, the most appropriate model will be selected to perform tumor segmentation according to its molecular subtype providing a predicted mask as an output.

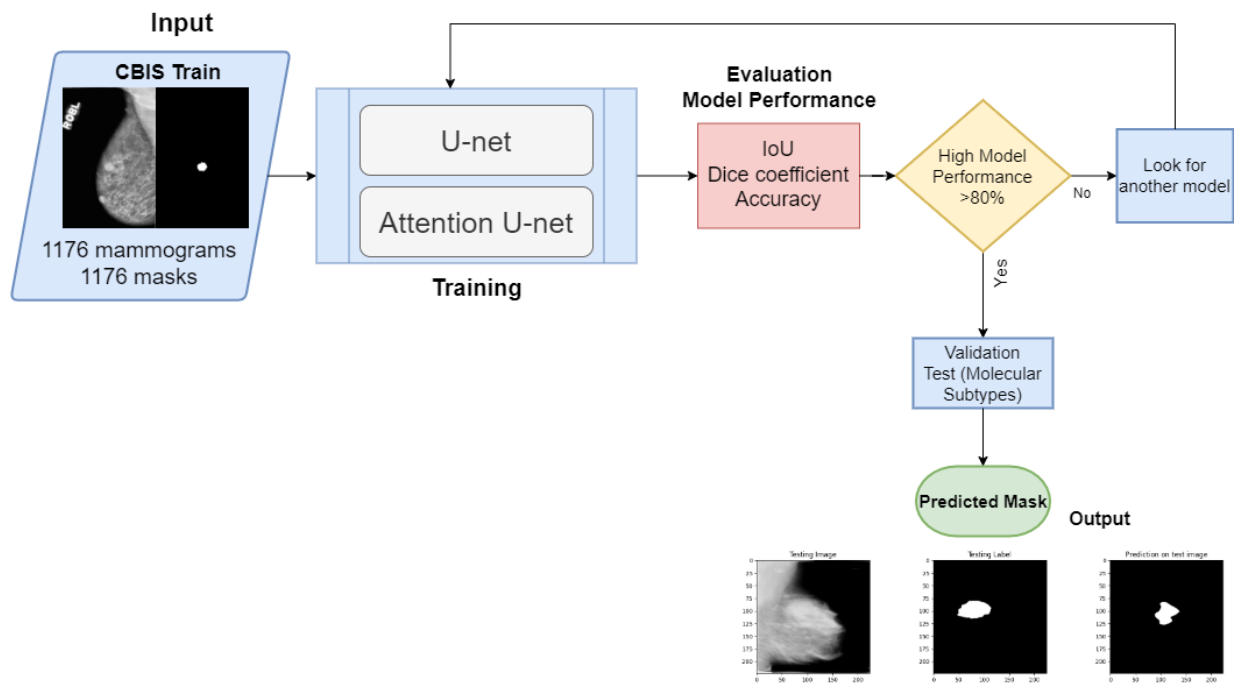


Figure 4.10: Block diagram of model segmentation selection process to obtained predicted mask with location of breast tumor.

iii. Experiment 3: Evaluation of TL models using segmented tumor images (ResNet50, VGG16, Xception, InceptionV3, VGG19, MobileNet).

For this experiment, the images obtained after tumor ROI segmentation and extraction were used. Some samples of the training dataset used are shown in Figure 4.11. The training set consists of 120 images, 30 images for Validation and 30 images for Testing. In addition, MobileNet and VGG19 models for transfer learning was added, the configuration was the same as used in Experiment 1.

Dataset of ROI of Breast Tumors

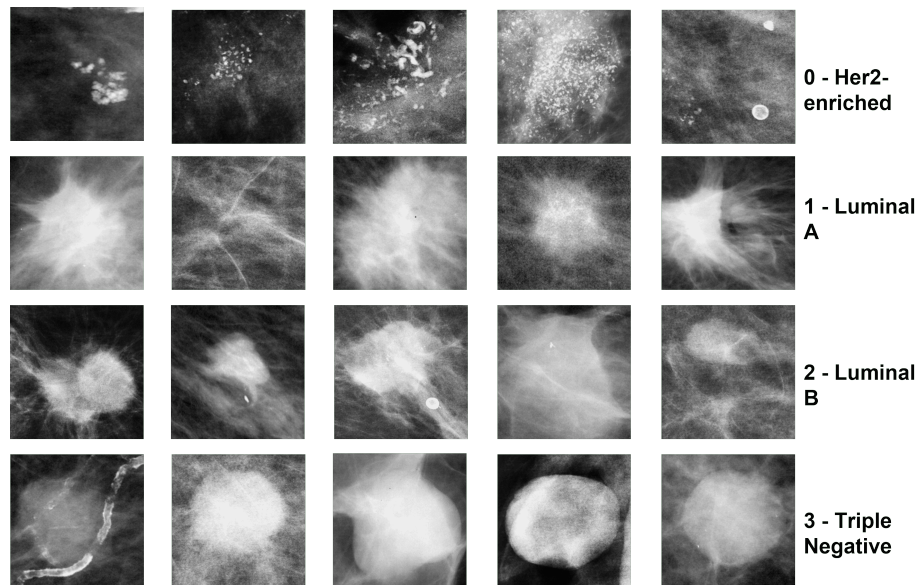


Figure 4.11: Samples training set of breast tumors images corresponding to HER2-enriched, Luminal A, Luminal B, and Triple negative subtypes cropped using ROI extraction.

The process of training and evaluation of different TL models in molecular subtypes classification is illustrated in Figure 4.12.

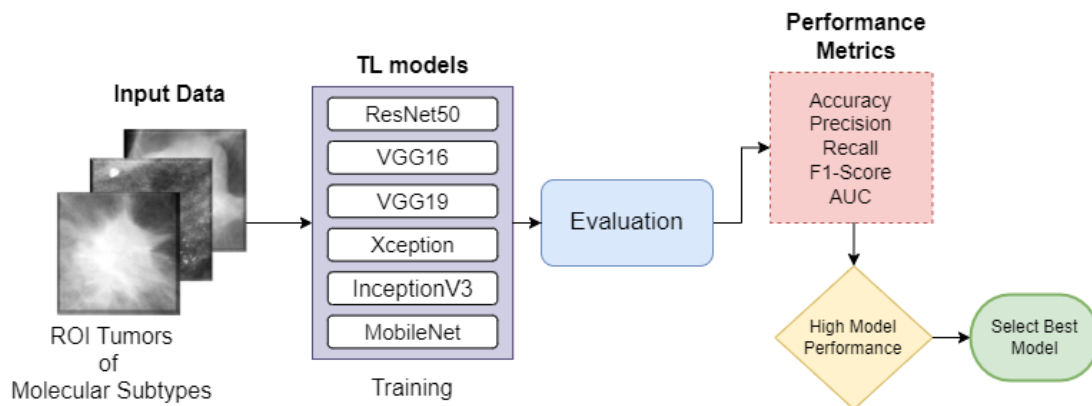


Figure 4.12: Flow-chart of evaluation of different CNNs models in classification of molecular subtypes using ROI breast tumors images.

iv. Experiment 4: Selection and evaluation of best TL model and Fine-Tuning.

After experiment 3, the best model with good accuracy is selected and adapted for the classification of ROI tumor images into four molecular subtypes of breast cancer using transfer learning and fine-tuning. The input ROI tumors images was augmented using ImageDataGenerator and the metrics described in Experiment 1.

The procedure for Fine-Tuning applied is explained below:

- Load pre-trained model.
- The classification layer was removed from the pretrained model, which was designed to classify 1000 ImageNet classes.
- A FC layer was added with 4 neurons, corresponding to the four molecular subtypes, using *softmax* as activation function in order to predict the probability of each class.
- The initial layers of pretrained model were frozen in order to allow that only layers added to be fine-tuned.
- Adjust hyperparameters such as learning rate, batch size, add dropout, L2-regularizer to avoid overfitting. Adjust the loss function to a categorical crossentropy.

After the training the model, the best model is saved through Callbacks and ModelCheckpoint. Finally, the model is evaluated through the metrics described in the next section.

The flow-chart illustrated in Figure 4.13 summarizes the process applied in this experiment.

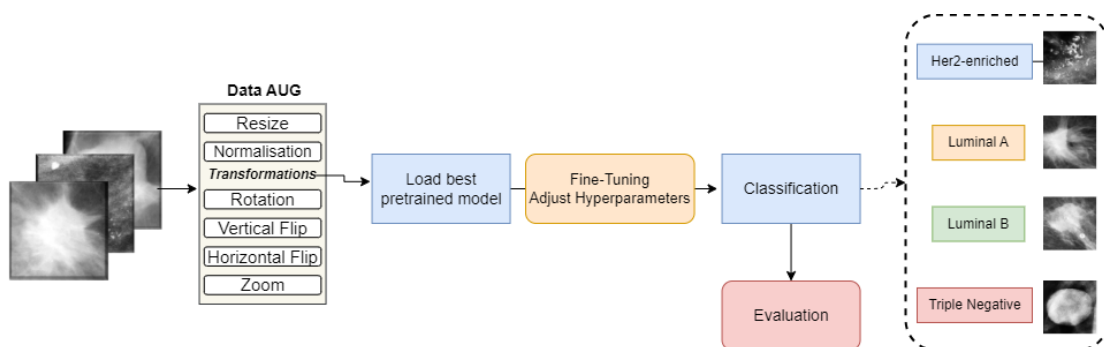


Figure 4.13: Flowchart of the Transfer Learning Process with best model for the Classification of Molecular Subtypes.

4.5 Model Performance Metrics

4.5.1 Evaluation metrics for Image Classification

In order to assess the performance of the proposed model, the confusion matrix is one of the fundamental tools for binary or multiclass classification tasks. This matrix represents the number of true values versus predicted results and divided into four components:

- **TP:** True Positive indicates the cases in which the model correctly predicted a class labeled as positive (e.g., correctly identifies breast cancer patients as positive).
- **TN:** True Negative indicates the cases in which the model correctly predicted a class labeled as negative (e.g., correctly identifies non-cancerous cases as negative).
- **FP:** False Positive indicates the cases in which the model incorrectly predicted a class labeled as negative, as positive (e.g., incorrectly identifying a healthy patient as having cancer).
- **FN:** False Negative indicates the cases in which the model incorrectly predicted a class labeled as positive, as negative (e.g., wrongly identify a cancer patient as healthy).

Based on the confusion matrix, there are several evaluation measures for BC detection, described below in detail:

Accuracy

Accuracy score measures the proportion of all correct predictions out of total cases, showing how well the model correctly identifies both normal (non-cancerous) and abnormal (cancerous) cases. It can be expressed by equation (4.1)

$$\text{Accuracy}(\%) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4.1)$$

Precision

Precision is determined by dividing the true positives predictions among all positive predictions made by the model. High precision indicates that the model predict correctly

classifying cancer patients among all positives cases, reducing false positives. It can be measured by the equation (4.2).

$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4.2)$$

Sensitivity or Recall

Recall calculates the proportion of actual positives (cancer cases) that are correctly identified by the model and it can be measured using the equation (4.3).

$$\text{Recall}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.3)$$

F1-Score

F1-score is used to measure the overall performance of classification model, it combines precision and recall into a balanced metric. This metric is especially helpful for imbalanced datasets, where a high score in one metric (e.g., sensitivity) might come at the expense of another (e.g., precision). It can be denoted by equation (4.4).

$$\text{F1-Score}(\%) = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.4)$$

AUC-ROC

The AUC-ROC metric represents the area under the Receiver Operating Characteristic (ROC) curve, which plots sensitivity against the false positive rate at different thresholds. A high AUC value (closer to 1) indicates better overall model performance across all thresholds, distinguishing between classes more effectively.

4.5.2 Evaluation metrics for Image Segmentation

Intersection over Union (IoU)

IoU or Jaccard coefficient, is a metric used to measure the overlap between the predicted segmentation and the ground truth segmentation. The equation of this metric is 4.5

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (4.5)$$

Where:

- $|A \cap B|$ is the area of intersection between the predicted and ground truth regions.
- $|A \cup B|$ is the area of their union.

Dice coefficient

This coefficient measures the similarity between the predicted label and the ground truth region. The corresponding equation is 4.6

$$\text{Dice coefficient} = \frac{2|A \cap B|}{|A| + |B|} \quad (4.6)$$

where:

- $|A \cap B|$, is the area of intersection between the predicted and ground truth regions.
- $|A|$, is the area of the predicted region.
- $|B|$, is the area of the ground truth region.

Chapter 5

Results and Discussion

5.1 Detection of Breast Cancer

5.1.1 Experiment 1: Comparison between the previous model based on AlexNet and fine-tuned proposed model.

In this section, accuracy and loss curves of previous pretrained AlexNet model and Fine-tuned proposed model is described below.

Firstly, both the accuracy and loss curves of the reference model and the proposed model show a similar trend; during training, the loss curve decreases and the accuracy curve increases. However, each there are some differences according to the validation and training curves. In Figure 5.1, the pre-trained model based on AlexNet shows the loss curve decreases close to zero, which indicates that the model is effectively learning to minimize the error in the training set. In the case of the accuracy curve, it increases rapidly and reaches very high values (0.95) since the beginning of training. However, the validation accuracy curve above the training curve and reaching values very close to 1 suggests overfitting. This means that the model has difficulty generalizing to new data.

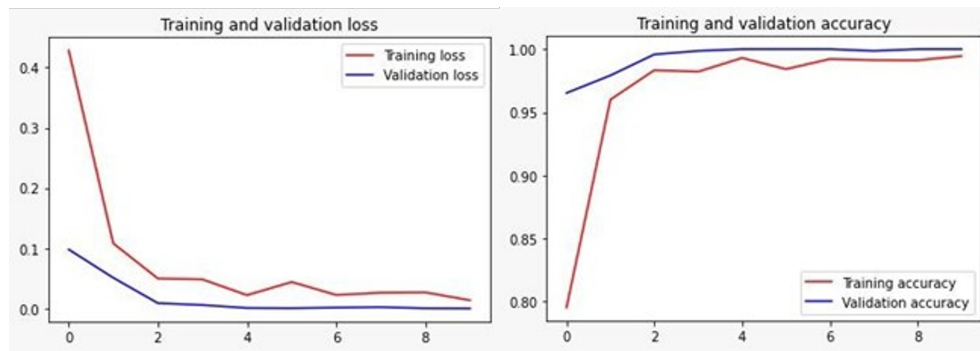


Figure 5.1: Accuracy and Loss curves of previous pre-trained AlexNet model.

In the case of the proposed AlexNet model (Figure 5.2), the loss curve also decreases steadily, but its starting point is slightly higher than that of the baseline model. Moreover, the validation loss curve lies above the training one, which might indicate that the model is not generalizing as well as expected. Tuning the learning rate, or using another optimizer suitable for this binary classification. Meanwhile, the accuracy curve increases and stabilizes around 0.90 to 0.99, indicating good performance. The two accuracy curves are almost aligned, but the validation curve is under the training curve. Therefore, it was necessary to employ another evaluation metric to complement the evaluation of the model's performance.

The test accuracy after training achieved was 98.60%.

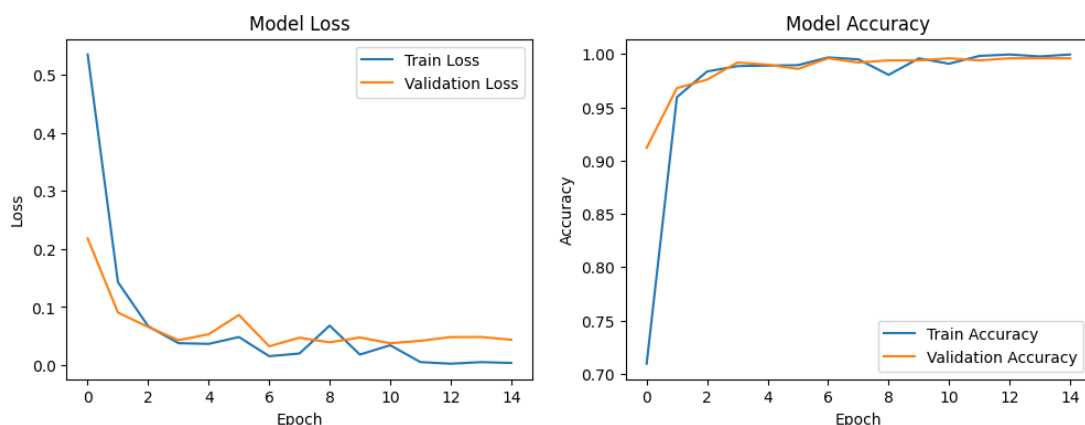


Figure 5.2: Accuracy and Loss curves of fine-tuning AlexNet model.

The confusion matrix shown in Figure 5.3 demonstrates that the breast cancer detection model proposed has excellent performance: it correctly identified 193 of 200 cancer cases (96.5% recall) and correctly classified the 200 normal cases. There were no false positives

and only 7 false negatives (patient with cancer classified as a normal). The model precision was 100% and the overall accuracy of 98.25%, indicating that the model correctly predicts in most cases, although there is a small chance that the model will fail.

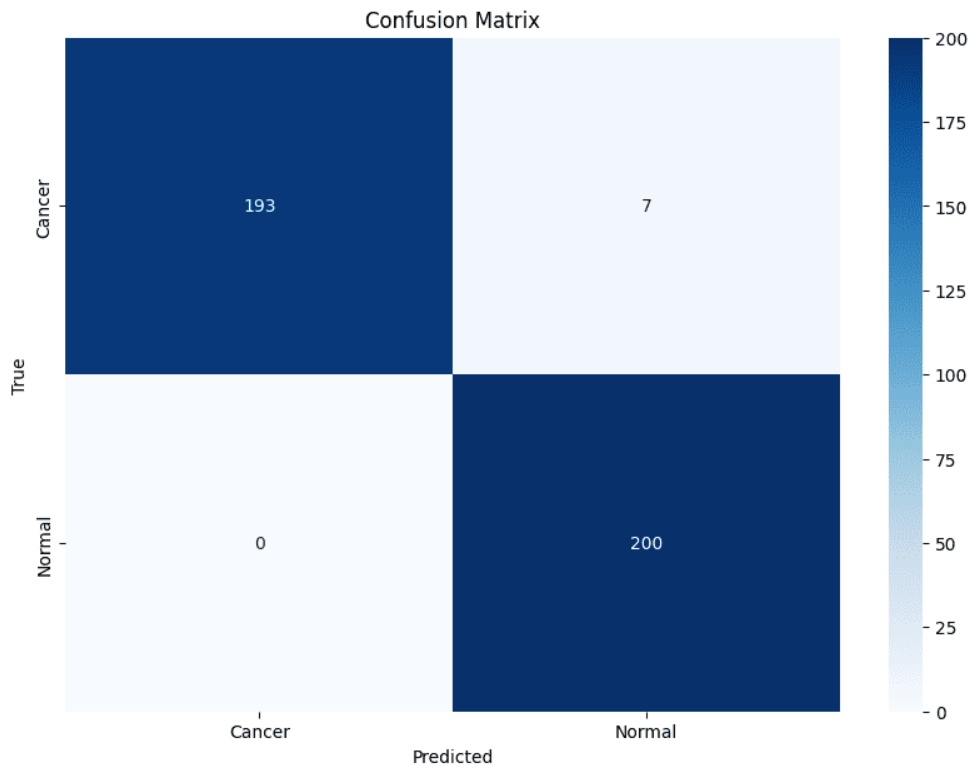


Figure 5.3: Confusion matrix of proposed Alexnet model applying fine-tuning.

Comparing our work with other studies in terms of performance (Table 5.1), it is observed that our breast cancer detection model presents good results. However, Salama et al., [67] report superior performance in terms of accuracy and recall, with values of 98.87% and 98.98%, respectively. This is because their approach focused exclusively on cancer detection using only the tumor region of the breast. Similarly, Mahmood et al., [65] highlights that the use of a ROI and the application of image enhancement techniques, such as CLAHE, significantly increases the performance of deep learning models. In his study, he achieved a accuracy of 98% and a loss of 0.07. This is consistent with Murtaza et al., [86] who state that pre-processing and enhancement techniques such as ROI extraction, image normalization, cropping and data augmentation, improve model performance.

Table 5.1: Comparison of proposed fine-tuned AlexNet model with related works

Ref.	Metrics			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[4]	98	-	86.7	-
[64]	95.10	-	70.8	-
[65]	98	-	99	-
[66]	95.70	94.6	95	-
[59]	87.20	85	76.3	80
[67]	98.87	98.79	98.98	97.99
Present work	98.60	100	96.50	98.22

5.2 Identification and classification of Molecular Subtypes with Transfer Learning

Initially, we evaluated and compare some CNNs suggested to molecular subtyping classification such as, ResNet50, VGG16, Xception, and InceptionV3 models with TL , mammograms of breast ROI or BROI was used as an input data for these models.

5.2.1 Experiment 1: Comparison of TL with ResNet50, VGG16, Xception, InceptionV3 models using BROI images.

In this section, we show the accuracy and loss results of the TL models for molecular subtype classification. As seen in the bar chart (Figure 5.4), all CNNs demonstrated very low performance below 50%. The ResNet50 model was the worst with an accuracy of 22.57%, followed by Xception with 40.28%, VGG16 with 43.75%, and finally InceptionV3 which showed a higher accuracy percentage of 45.49% compared to the other architectures. In the case of model loss, the architecture with the lowest value of loss was VGG16. This was due to the complexity and variability of the mammography images, which made it challenging to identify and classify molecular subtypes.

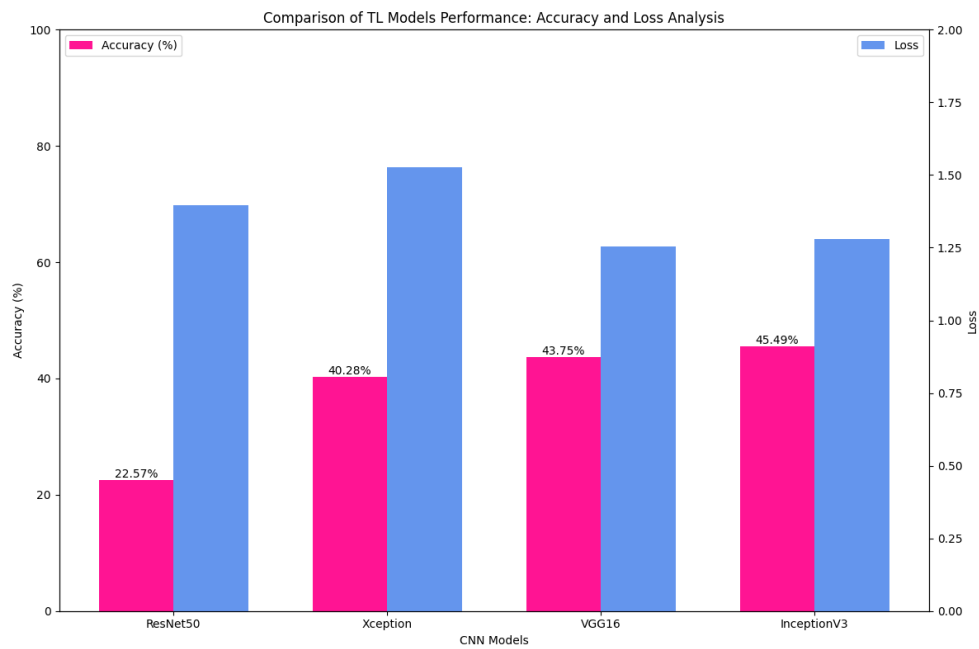


Figure 5.4: Comparison of Transfer Learning models performance (ResNet50, Xception, VGG16, InceptionV3) using BROI images.

5.2.2 Experiment 2: Comparison of Unet and Attention Unet in breast tumor segmentation.

To obtain and extract the breast tumor ROI, was necessary the use of segmentation model. For this, two segmentation models were trained, U-net and an optimized version called Attention U-net, and the best one was selected based on the one that shows better performance for segmenting tumors. Table 5.2 shows the results of the evaluation metrics for each model.

Table 5.2: Model performance of U-net and Attention U-net

Model	Metrics				
	Accuracy	Loss	Dice coefficient	Jaccard index	Mean IoU
Unet	0.9854	0.0107	0.6940	0.7498	0.5338
AttUnet	0.9881	0.0084	0.7005	0.8657	0.5423

According to Table above, the Att U-net segmentation model shows promising performance, with high accuracy 98.81% indicating correct classification of most pixels. However, the intersection between the predicted and true label, reflected by the Dice Coefficient 70% and Jaccard Index 86%, although good, still has room for improvement. The Mean IoU of

0.54 suggests inconsistencies in segmentation between different classes. Compared to U-net model, Att U-net delivers superior results, but it would be beneficial to continue fine-tuning the model, optimizing the image preprocessing to further improve segmentation. However, this model could help in segmentation of breast tumor. Furthermore, Figure 5.5 shows how segmentation is performed on a test mammogram, as seen on the top, U-net did not fully segment the tumor region and had an IoU of 0.74. However, the tumor segmentation on the predicted mask through the Attention U-net model obtained a higher IoU of 0.82. This complements the information in the previous table, demonstrating that Attention U-net is a more suitable in the task of tumor segmentation.

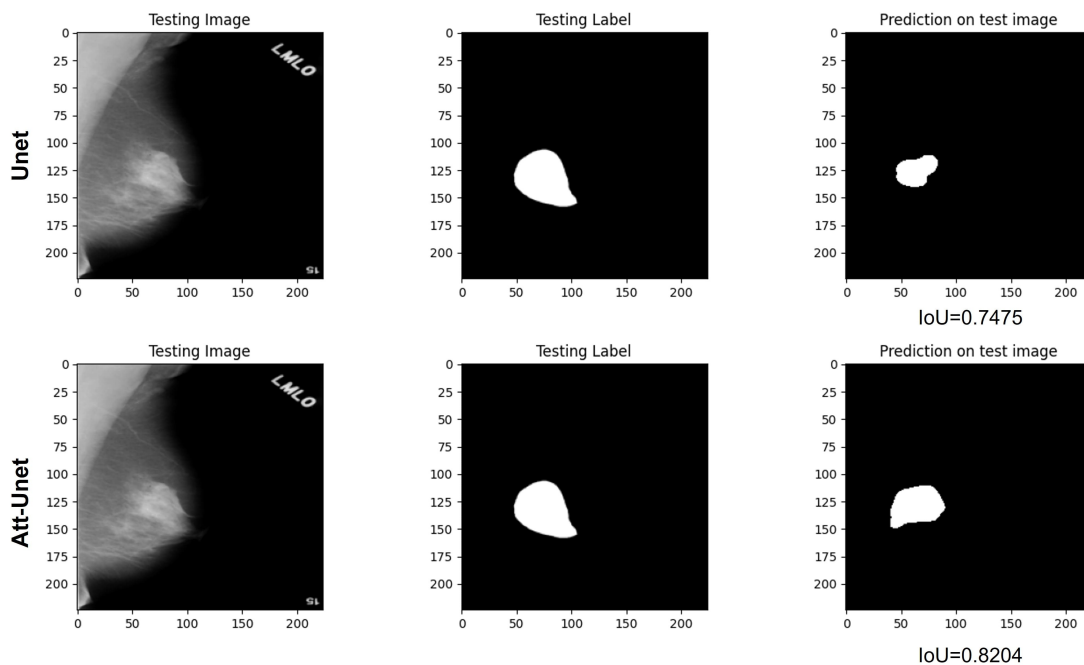


Figure 5.5: Comparison of sample segmentation results of U-net and Att-U-net models

5.2.3 Experiment 3: Evaluation of TL models for identification of molecular subtypes using breast tumors ROI images

As in Experiment 1, to select the best architecture for subtype classification, transfer learning was used with the same models mentioned in the previous experiment. In addition, two additional models were implemented: MobileNet and VGG19. The input data contains 200 images of breast tumors ROI.

Table 5.3: Performance comparison of different TL in molecular subtypes classification using breast tumors ROI.

Model	Metrics					
	Proposed	Acc(%)	Precision (%)	Recall(%)	F1-Score (%)	AUC (%)
VGG16		42.7	58.0	55.8	54.8	81.0
ResNet50		44.8	28.8	39.0	31.8	73.8
VGG19		52.1	61.0	55.8	55.5	83.0
Xception		64.6	63.0	62.8	62.0	86.5
InceptionV3		66.7	68.8	68.3	66.5	86.5
MobileNet		67.7	68.0	68.3	68	87.8

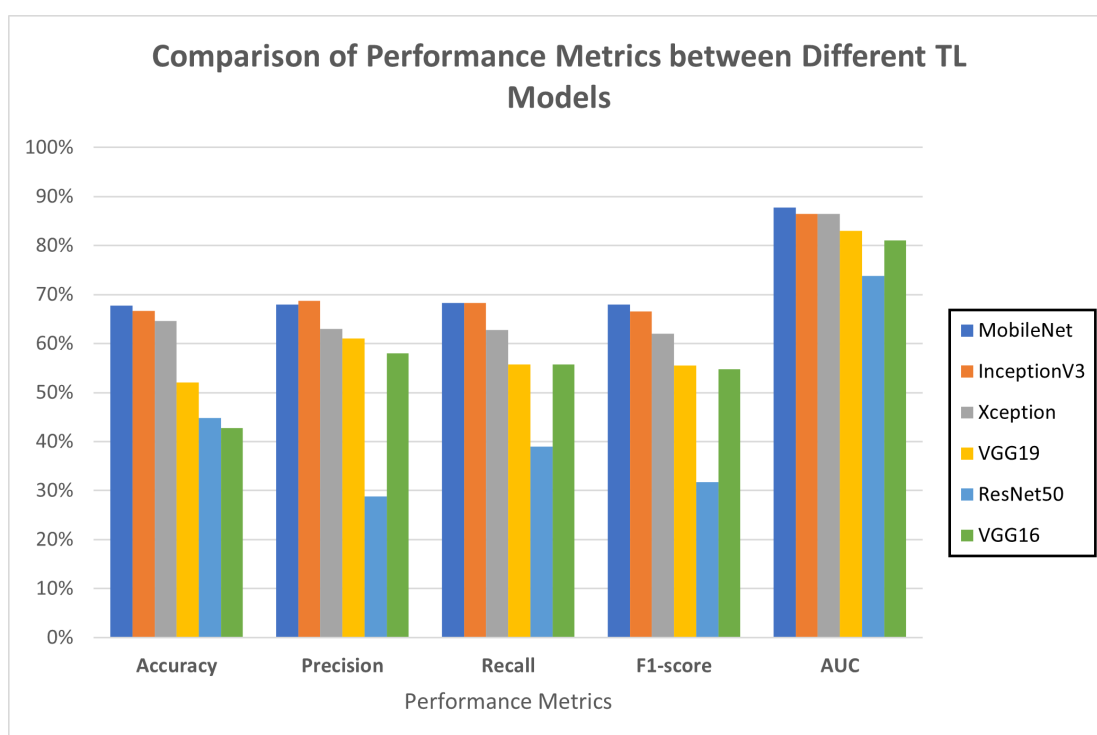


Figure 5.6: Bar chart-based performance comparative analysis of different CNNs in molecular subtypes classification.

After training and evaluating the performance of the architectures for classification, as seen in the Figure 5.6 and Table 5.3, the model that has the best results was MobileNet, reaching an higher accuracy of 67.7% and a low loss of 0.83, and precision 68%, recall of 68.3%, F1-score of 68, and AUC of 87.5%. In the case of other models such as: InceptionV3, Xception, VGG19, ResNet50, and VGG16 achieved 66.7%, 64.6%, 52.1%, 44.8%, and 42.7% accuracy, respectively, with corresponding losses of 0.94, 0.99, 1.03, 1.23, and 1.27. As a result, the model selected is MobileNet. Thus, MobileNet and InceptionV3

stand out as the best models on this dataset using tumor ROIs, with high values in accuracy, precision, recall, F1-score and AUC, and with the lowest loss values. In contrast, VGG16 and ResNet50 show relatively poor performance in several metrics, with VGG16 being especially low in precision and ResNet50 showing low recall and F1-score. Moreover, comparing with the accuracy results of the CNNs used in experiment 1, it was shown that the extraction of the breast tumor ROI allowed a significant improvement, compared to the breast ROI. This is because the tumor segmentation allows to reduce the complexity of the image and allows to focus on the most relevant features of the tumor [87]. Xception and InceptionV3 models improve the accuracy to 60%. On contrary, CNNs such as ResNet50 and VGG16 increased their accuracy, however it is still an accuracy below 50%.

5.2.4 Experiment 4: Selection and evaluation of best TL model for classification of breast cancer molecular subtypes

For the classification of molecular subtypes using the ROI tumor images as input, the transfer learning technique was used with the pre-trained MobileNet model, due to its better performance in the previous experiment. To adapt this model to our classification task, the following adjustments were made:

- Add FC layer of 1024 with ReLU as activation function,
- L2 Regularizer of 0.001
- Dropout: 0.5.
- Loss function: Categorical crossentropy.
- Batch size: 32
- LR: 0.001
- Optimizer: Adam
- Epoch: 100

The architecture of proposed model is illustrated in Figure 5.7

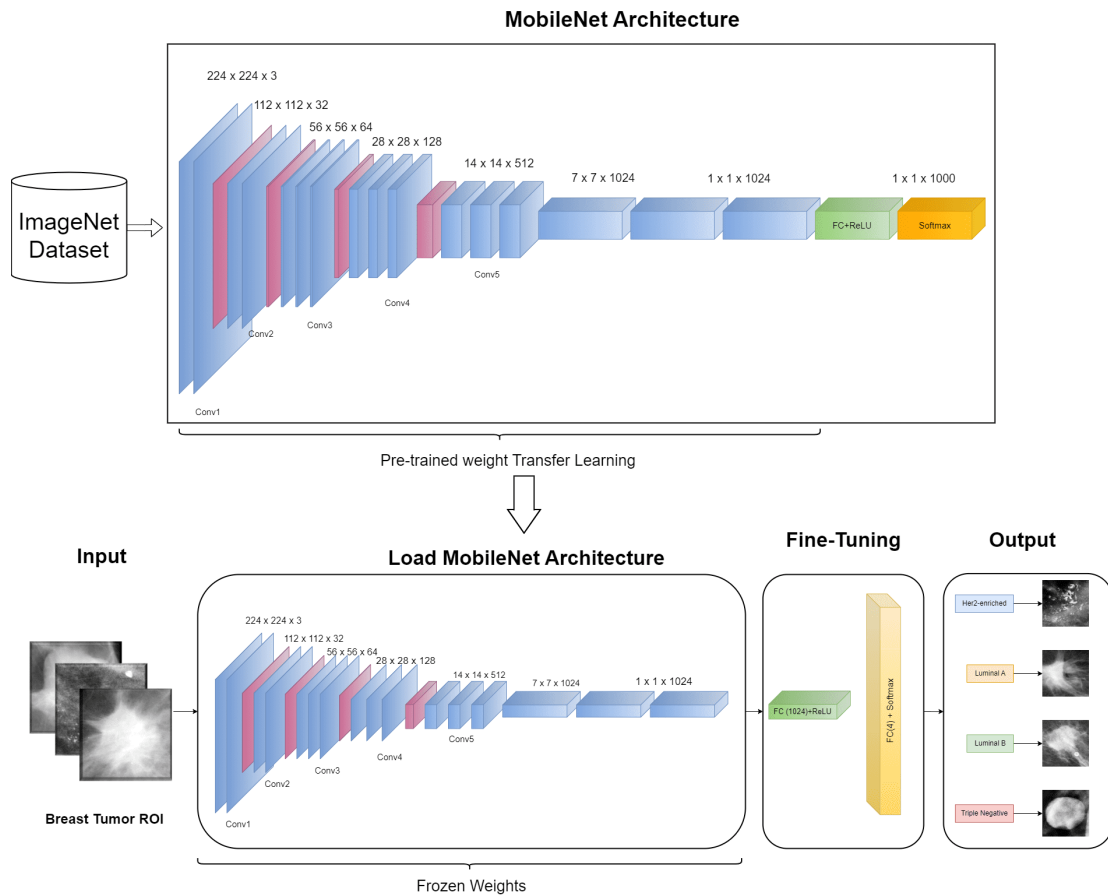


Figure 5.7: Architecture proposed of transfer learning model based on MobileNet.

After training the proposed model, the accuracy and loss curves shown in Figure 5.8 were obtained. Besides, the model achieves a test accuracy was 73.95%.

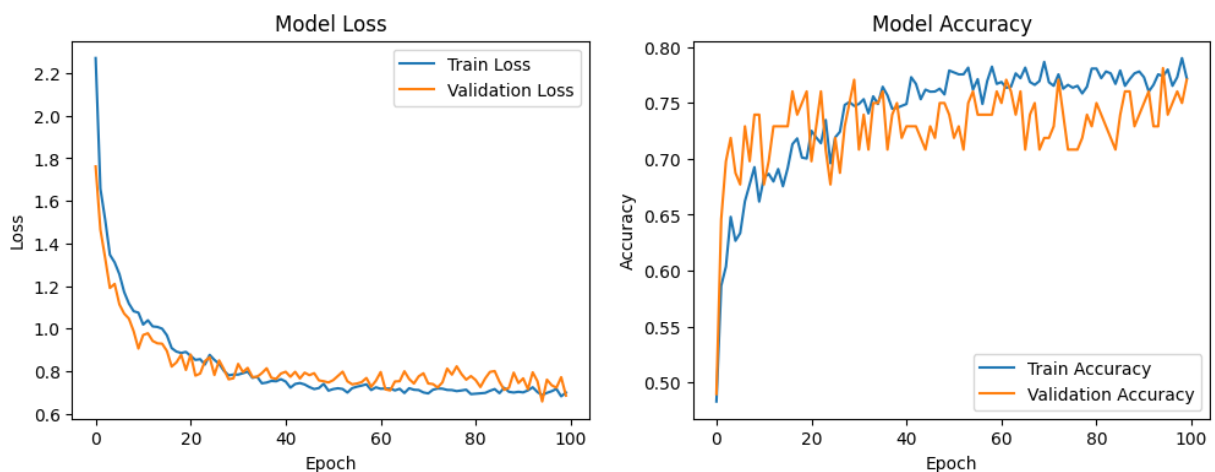


Figure 5.8: Accuracy and Loss curves of Fine-tuned MobileNet model.

The differences observed in each curve are explained below:

- The **loss curves** decrease constantly during training. The *training loss* curve decreases from an initial value of approximately 2.27 to reach 0.69 at epoch 100. The *validation loss* curve follows a similar trend, decreasing from 1.76 to 0.69, and remains aligned with the training loss curve, although always slightly above it. This indicates a reasonable fit of the model, but suggests that a lower loss could be achieved by adjusting some hyperparameters, such as the learning rate or the loss function.
- On the other hand, **the accuracy curves** indicate an increase during training, although with a more variable behavior. In fact, the *validation accuracy* curve is higher than the *training accuracy* curve for the first 35 epochs, but then remains below it until epoch 100. During this time, the accuracy varies from 0.67 to 0.77. This variability suggests that the model could benefit from further fine-tuning of the hyperparameters or additional regularization techniques to improve its stability and accuracy.

For the analysis of the confusion matrix (Figure 5.9) and the classification for each class, we have:

- **Her2-enriched (Class 0)**: The model performs well, with 28 out of 30 instances correctly classified (two misclassifications).
- **Luminal A (Class 1)**: The model also performs reasonably well but has some confusion with LumB, as six instances were misclassified.
- **Luminal B (Class 2)**: This class shows significant confusion with LumA and TN, indicating the model struggles more with this class.
- **Triple negative (Class 3)**: The model is relatively accurate but shows some confusion with the other classes, especially Class 3.

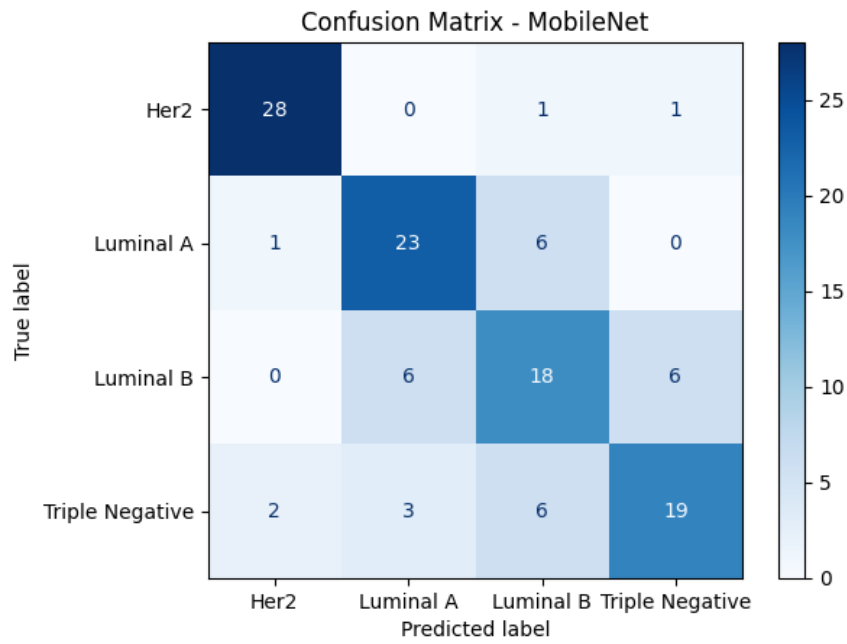


Figure 5.9: Confusion matrix of Fine-tuned MobileNet model.

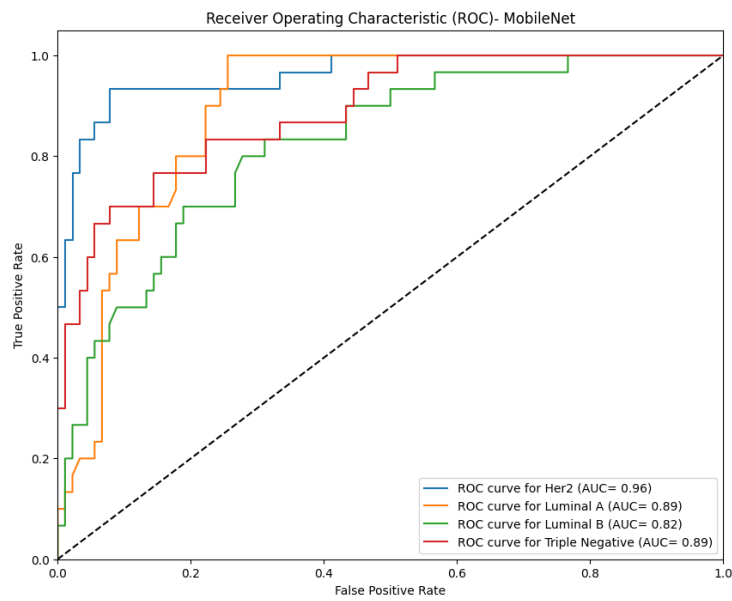


Figure 5.10: AUC-ROC curves of proposed Fine-Tuned MobileNet model

The results of the model’s performance from the confusion matrix are presented in Table 5.4, the table also includes AUC per class obtained from previous Figure 5.10.

Table 5.4: Transfer learning model based on MobileNet performance in multi-class classification.

Molecular Subtypes	Metrics			
	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Her2-enriched	90.32	93.33	91.80	96
Luminal A	71.88	76.67	74.19	89
Luminal B	58.06	60	59.02	82
Triple Negative	73.08	63.33	67.86	89

In Table 5.4, the 4-class classification performance metrics for the analyzed molecular subtypes are described. According to accuracy in predicting and classification different breast cancer molecular subtypes using mammograms, varies by subtype. Overall, the Her2-enriched subtype was the best ranked, presenting high values in all metrics: an accuracy of 90.32%, a recall of 93.33%, an F1-score of 91.80% and an AUC of 96%. In second place was the triple-negative subtype, with a precision of 73.08%, a recall of 63.33%, an F1-score of 67.86 and an AUC of 89. In third place, the Luminal A subtype showed a precision, recall, F1-score and AUC of 71.88%, 76.67%, 74.19% and AUC of 89%, respectively. Finally, the worst classified subtype was Luminal B, with an accuracy of 58.06%, a recall of 60%, an F1 score of 91.80%, and an AUC of 82%. The average precision obtained in the study was 73.33%.

The fact that the Her2-enriched molecular subtype was the best classified in this study is supported by the results obtained by Ueda et al., [71], who reported an AUC of 75%, and by Mota et al., [69], with an accuracy of 89.79% and an AUC of 73.31%. These authors argue that HER2 subtype, tend to have a higher predictive accuracy compared to others subtypes, such as Luminal subtypes. This is due to the tumor characteristics, especially due to the presence of microcalcification, while Luminal tumors, particularly Luminal A, may exhibit imaging features that significantly overlap with other subtypes, making them difficult to accurately differentiate. Luminal subtypes show greater variability in their imaging features, making it difficult for deep learning models to learn consistent patterns.

In contrast, these results differ from those obtained by Zhang et al., [70], who reported that Luminal subtypes were mostly identified with an AUC of 92.9%. Similarly, the study by Bhandary et al., [68] came to the same deduction, obtaining an accuracy of 81.88% in luminal cases. The authors state that this could be related to the relationship with

imaging features, where luminal subtypes are characterized by having an irregular and spiculated margins. In contrast, Her2 and TN subtypes present more regular and specific shapes; the first one is distinguished by the presence of microcalcifications, while TN is typically observed as a circumscribed and oval mass. These findings are consistent with that presented by Singh et al., [88], who argue that tumor morphology may be a determining factor in the classification of molecular subtypes.

In addition, the frequency of presentation of these subtypes in breast cancer cases and datasets used to train deep learning models often have a class imbalance, with some subtypes (such as Luminal A) overrepresented compared to others (such as HER2 and TN). This can bias the model toward predicting the most common subtypes and reduce the accuracy of less common subtypes.

To sum up, the higher accuracy in predicting some molecular subtypes is due to a combination of distinctive tumor features in the images, whereas the lower accuracy for other subtypes is related to feature overlap, class imbalance.

5.2.5 Overall Analysis and Comparison

In this section, the results of related works in deep learning models applied to the classification of molecular subtypes of from mammogram images with our proposed model are compared and analyzed, as it is summarized in Table 5.5.

Table 5.5: Comparison of our proposed model and state-art related works

Model	Metrics				
	Acc(%)	Precision (%)	Recall(%)	F1-Score (%)	AUC (%)
ResNet-18 [68]	-	-	-	66.88	66.93
ResNet-101 [69]	79.02	-	-	-	64.7
MDL-IIA [70]	88.5	87.8	85.4	86.2	92.0
VGG+Inception [71]	-	61.0	74.0	-	75.0
Our model	73.95	73.33	73.32	73.2	89.0

In terms of performance measurements, the multimodal approach proposed by Zhang et al., [70] presents the highest performance in molecular classification. Their model combined mammography and ultrasound images to predict molecular subtypes using a multi-modal deep learning model (MDL-IIA) incorporating the attention mechanism with ResNet50 as base architecture. The MDL-IIA model has achieved an accuracy, recall and

F1-score and AUC corresponding to 87.5%, 85.4%, 86.2% and 92%, respectively. These high values highlight how the incorporation of attention modules allows the model to selectively focus on the relevant characteristics of the images to detect patterns. In addition, it is highlighted that due to its multimodality, it achieved an MCC of 83.7% in the prediction of molecular subtypes of 4 categories.

In comparison with our proposed work, based on MobileNet architecture that achieved a satisfactory AUC of 89% but showed lower accuracy (73.95%), precision (73.3%), recall (73.32%), and F1-score (73.2%) relative to Zhang et al.'s multimodal model. Although the MobileNet model performed well with mammogram images alone, it may benefit from architectural adjustments and the inclusion of additional imaging modalities like ultrasound or MRI. Such as Wu et al. [89], who also observed improved results using MRI and MG images with an accuracy of 74.1%. This approach could further improve accuracy and generalizability, as demonstrated by previous studies leveraging multimodal data.

Furthermore, while many studies relied on residual networks like ResNet, such as Bandhary et al. [68] with ResNet-18, performance did not universally improve with this architecture. For instance, Bandhary's model achieved an F1-score of only 66.98% and an AUC of 66.93%, significantly lower than multimodal models. This suggests that ResNet-based models might have limited capacity to classify breast cancer subtypes accurately when used on single imaging modalities without further enhancements. Also, Mota et al., [69] employed ResNet-101 for both binary and multiclass classification of molecular subtypes using mammography images. In binary classification, accuracy exceeded 67% for all subtypes, with HER2 classification achieving the highest accuracy (89.79%) and AUC (73.31%) after oversampling and augmentation. However, multiclass classification showed lower accuracy (66.88%) and AUC (66.93%) values, possibly due to the challenges associated with class imbalance. In the case of a different approach proposed by Ueda et al. [71], who used a mix of VGG16, InceptionV3, ResNet52, and DenseNet121 to predict receptor expression (ER, PgR, HER2) from mammograms. Performance for ER and PgR was limited (AUCs of 67% and 61%, respectively), with HER2 prediction slightly better (AUC of 75%), in which better combinations was VGG16 and InceptionV3. This study showed some limitations such as limited mammograms images.

Overall, the studies demonstrate the potential of deep learning to predict breast cancer

molecular subtypes from images. Multimodal integration, transfer learning, and incorporation of additional clinical information are promising strategies to improve model accuracy. However, larger and more diverse datasets, strategies to address class imbalance, and rigorous clinical validation are needed to translate this technology into clinical practice.

Chapter 6

Conclusion and Future Works

6.1 Conclusion

In this study, two deep learning models with two specific objectives were developed: one for breast cancer detection and another for classifying molecular subtypes using mammographic images. For the first, we proposed a model based on AlexNet CNNs, which demonstrated outstanding performance, achieving an overall accuracy of 98.25%. Thanks to pre-processing techniques, the model successfully distinguished between normal mammograms and those with cancer with 100% precision, reaching a sensitivity of 96.50%. These results underline the high performance in the detection task compared to other models, providing a potential tool for early breast cancer diagnosis.

Regarding the identification and classification of molecular subtypes from mammograms, the MobileNet model with transfer learning achieved an overall accuracy of 73.95%. This process involved tumor segmentation and tumor region of interest (ROI) extraction, since using breast ROI images did not allow the model to focus on the imaging features characteristic of each molecular subtype.

Furthermore, it is highlighted that the best classified molecular subtype was the one enriched with Her2, with a precision of 90.32%, a recall of 93.3%, an F1-score of 91.80% and an area under the curve (AUC) of 96%, followed by the Triple Negative subtype with an accuracy of 73.08%, and the Luminal A and B subtypes with precision values of 71.88% and 58.06%. Luminal B was the worst classified, since it shares characteristics in its morphology similar to Luminal A, which complicates the distinction between these classes.

This is attributed to the fact that image features are distinctive in mammograms for each molecular subtype.

In brief, the proposed Deep learning models offer a promising alternative to assist radiologists and oncologists in the diagnosis and treatment of breast cancer through medical imaging.

6.2 Limitations

Regarding the breast cancer detection model, it offers a high performance, but adjustments must be made to the architecture and its hyperparameters to avoid overfitting.

Despite the promising results obtained, the study on the identification of molecular subtypes presents certain limitations that must be considered so that the model improves its ability to classify molecular subtypes.

One of the main limitations is the limited number of training images that were used in the model based on the region of interest (ROI) of the tumor. This fact suggests the need for a more extensive database to improve the generalization of the model that is classified by its molecular subtype that is supported by a biopsy or detailed clinical data, and the assistance of an expert radiologist who can correctly identify tumors in mammograms. Although the segmentation model had a great performance, some tumors were not segmented correctly, which reduces the amount of data. Furthermore, the imbalance of molecular subtype data due to the frequency of cases in breast cancer means that the model may be biased by the majority class, reducing the accuracy of less frequent subtypes. Another challenge of the present study regarding the model is that although deep learning models are powerful, their ability to represent subtle differences between molecular subtypes remains a challenge. Integrating more advanced feature extraction techniques and incorporating additional clinical and genetic data could improve model performance.

6.3 Future works

In future research, it is recommended to incorporate multi-omics data, such as gene expression, as well as additional images, such as histology, MRI, ultrasound, in order to develop a multi-modal model that offers a more complete view of breast cancer and its molecular

subtypes. Some future applications of these developments include:

- a. **Diagnostic assistance and treatment support:** Improve the accuracy and efficiency in the detection and diagnosis of breast cancer, by providing advanced tools that assist radiologists, clinicians identifying the optimal treatment for each patient based on precise classification of the molecular subtype.
- b. **Development of an Integrated Model:** Creating a model that combines the two processes discussed in this study: cancer detection and molecular subtype classification in one model could provide a more complete and efficient solution for breast cancer diagnosis.
- c. **Application to other types of cancer:** Extend the application of this approach to the detection of other cancer types or diseases that could benefit from a similar analysis, thus expanding the impact of the model in personalized medicine.

These proposals not only have the potential to enhance the accuracy and clinical outcomes of the developed model, but could also open new research avenues in breast cancer diagnosis and help reduce mortality rates.

6.4 Scientific Dissemination

Previous research presented at the 4th International Conference of Information Systems and Software Technologies ICI2ST 2023 with the work titled "Early Detection of Breast Cancer using Pretrained AlexNet Convolutional Neural Network" [85].

Bibliography

- [1] W. C. R. F. International, “Breast cancer statistics,” 3 2022. [Online]. Available: <https://www.wcrf.org/cancer-trends/breast-cancer-statistics/>
- [2] Surveillance, Epidemiology, and End Results (SEER) Program, “Cancer stat facts: Breast cancer,” 2024. [Online]. Available: <https://seer.cancer.gov/statfacts/html/breast.html>
- [3] W. H. Organization. Breast cancer. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>
- [4] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, “Deep learning to improve breast cancer detection on screening mammography,” *Scientific Reports*, vol. 9, p. 12495, 5 2019.
- [5] I. A. for Reseach on Cancer (IARC), “Ecuador fact sheets,” 2020. [Online]. Available: <https://gco.iarc.who.int/media/globocan/factsheets/populations/218-ecuador-fact-sheet.pdf>
- [6] A. C. Society. (2023) Breast cancer facts and statistics. [Online]. Available: <https://www.breastcancer.org/facts-statistics>
- [7] C. A. Almeida and S. A. Barry, *Cancer: Basic Science and Clinical Aspects*, 22nd ed., Wiley-Blackwell, Ed., 2011.
- [8] Y. Feng, M. Spezia, S. Huang, C. Yuan, Z. Zeng, L. Zhang, X. Ji, W. Liu, B. Huang, W. Luo, B. Liu, Y. Lei, S. Du, A. Vuppalapati, H. H. Luu, R. C. Haydon, T. C. He, and G. Ren, “Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis,” pp. 77–106, 6 2018.

- [9] C. Admoun and H. N. Mayrovitz, *The Etiology of Breast Cancer*. Exon Publications, 8 2022, pp. 21–30.
- [10] H. R. Brewer, M. E. Jones, M. J. Schoemaker, A. Ashworth, and A. J. Swerdlow, “Family history and risk of breast cancer: an analysis accounting for family structure,” *Breast Cancer Research and Treatment*, vol. 165, pp. 193–200, 8 2017.
- [11] Y. S. Sun, Z. Zhao, Z. N. Yang, F. Xu, H. J. Lu, Z. Y. Zhu, W. Shi, J. Jiang, P. P. Yao, and H. P. Zhu, “Risk factors and preventions of breast cancer,” pp. 1387–1397, 2017.
- [12] F. P. Turkoz, M. Solak, I. Petekkaya, O. Keskin, N. Kertmen, F. Sarici, Z. Arik, T. Babacan, Y. Ozisik, and K. Altundag, “Association between common risk factors and molecular subtypes in breast cancer patients,” *Breast*, vol. 22, pp. 344–350, 2013.
- [13] K. Lee, L. Kruper, C. M. Dieli-Conwright, and J. E. Mortimer, “The impact of obesity on breast cancer diagnosis and treatment,” *Current Oncology Reports*, vol. 21, p. 41, 5 2019.
- [14] I. Sechopoulos, J. Teuwen, and R. Mann, “Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art,” pp. 214–225, 7 2021.
- [15] F. Gilbert and K. Pinker-Domenig, *Diagnosis and Staging of Breast Cancer: When and How to Use Mammography, Tomosynthesis, Ultrasound, Contrast-Enhanced Mammography, and Magnetic Resonance Imaging*, 2019, pp. 155–166.
- [16] L. Wang, “Early diagnosis of breast cancer,” *Sensors (Switzerland)*, vol. 17, 7 2017.
- [17] D. Barba, A. León-Sosa, P. Lugo, D. Suquillo, F. Torres, F. Surre, L. Trojman, and A. Caicedo, “Breast cancer, screening and diagnostic tools: All you need to know,” 1 2021.
- [18] B. Niell, P. Freer, R. Weinfurtner, E. Arleo, and J. Drukteinis, “Screening for breast cancer,” pp. 1145–1162, 11 2017.

- [19] M. F. Mridha, M. A. Hamid, M. M. Monowar, A. J. Keya, A. Q. Ohi, M. R. Islam, and J. M. Kim, “A comprehensive survey on deep-learning-based breast cancer diagnosis,” 12 2021.
- [20] C. Coleman, “Early detection and screening for breast cancer,” pp. 141–155, 5 2017.
- [21] D. Eugênio, J. Souza, R. Chojniak, A. Bitencourt, L. Graziano, and E. Marques, “Breast cancer diagnosed before the 40 years: imaging findings and correlation with histology and molecular subtype,” *Applied Cancer Research*, vol. 37, 12 2017.
- [22] R. M. Mann, R. Hooley, R. G. Barr, and L. Moy, “Novel approaches to screening for breast cancer,” *Radiology*, vol. 297, pp. 266–285, 11 2020.
- [23] S. Temerik, S. Elwahab, M. Wahman, M. Ahmed, and M. Elwanis, “Relation between morphological features of initial breast mri and breast cancer molecular subtypes,” *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 54, 12 2023.
- [24] E. Horvath, “Molecular subtypes of breast cancer-what breast imaging radiologists need to know,” pp. 17–26, 2021.
- [25] B. Dołęga-Kozierowski, M. Lis, H. Marszalska-Jacak, M. Koziej, M. Celer, M. Bandyk, P. Kasprzak, B. Szynglarewicz, and R. Matkowski, “Multimodality imaging in lobular breast cancer: Differences in mammography, ultrasound, and mri in the assessment of local tumor extent and correlation with molecular characteristics,” *Frontiers in Oncology*, vol. 12, 8 2022.
- [26] R. Guo, G. Lu, B. Qin, and B. Fei, “Ultrasound imaging technologies for breast cancer detection and management: A review,” pp. 37–70, 1 2018.
- [27] S. Rashmi, S. Kamala, S. S. Murthy, S. Kotha, Y. S. Rao, and K. V. Chaudhary, “Predicting the molecular subtype of breast cancer based on mammography and ultrasound findings,” *Indian Journal of Radiology and Imaging*, vol. 28, pp. 354–361, 7 2018.

- [28] Z. Huang, L. Chen, Y. Wang, L. Fu, and R. Lv, “Molecular markers, pathology, and ultrasound features of invasive breast cancer,” *Clinical Imaging*, vol. 79, pp. 85–93, 11 2021.
- [29] P. Steyerova and A. Burgetova, “Current imaging techniques and impact on diagnosis and survival—a narrative review,” *Annals of Breast Surgery*, vol. 6, pp. 25–25, 9 2022.
- [30] C. Perou, T. Sørlie, M. Eisen, M. Rijn, S. Jeffrey, C. Rees, J. Pollack, D. Ross, H. Johnsen, L. Akslen, Fluge, A. Pergamenschikov, C. Williams, S. Zhu, P. Lønning, A.-L. Børresen-Dale, P. Brown, and D. Botstein, “Molecular portraits of human breast tumours,” *Nature*, vol. 406, pp. 747–52, 09 2000.
- [31] K. S. Johnson, E. F. Conant, and M. S. Soo, “Molecular subtypes of breast cancer: A review for breast radiologists,” pp. 12–24, 1 2021.
- [32] R. Gulzar, R. Shahid, and O. Saleem, “Molecular subtypes of breast cancer by immunohistochemical profiling,” 2018. [Online]. Available: <https://www.researchgate.net/publication/328744478>
- [33] X. Dai, L. Xiang, T. Li, and Z. Bai, “Cancer hallmarks, biomarkers and breast cancer molecular subtypes,” pp. 1281–1294, 2016.
- [34] S. W. Lam, C. R. Jimenez, and E. Boven, “Breast cancer classification by proteomic technologies: Current state of knowledge,” pp. 129–138, 2 2014.
- [35] E. Provenzano, G. A. Ulaner, and S. F. Chin, “Molecular classification of breast cancer,” pp. 325–338, 7 2018.
- [36] G. Pankotai-Bodó, O. Oláh-Németh, F. Sükösd, and T. Pankotai, “Routine molecular applications and recent advances in breast cancer diagnostics,” pp. 20–28, 1 2024.
- [37] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso, “Breast cancer,” *Nature Reviews Disease Primers*, vol. 5, 12 2019.
- [38] L. Yin, J. J. Duan, X. W. Bian, and S. C. Yu, “Triple-negative breast cancer molecular subtyping and treatment progress,” 6 2020.

- [39] A. Prat, E. Pineda, B. Adamo, P. Galván, A. Fernández, L. Gaba, M. Díez, M. Viladot, A. Arance, and M. Muñoz, “Clinical implications of the intrinsic molecular subtypes of breast cancer,” *Breast*, vol. 24, pp. S26–S35, 11 2015.
- [40] M. Wu and J. Ma, “Association between imaging characteristics and different molecular subtypes of breast cancer,” *Academic Radiology*, vol. 24, pp. 426–434, 4 2017.
- [41] T. W. M. Ian, E. Y. Tan, and N. Chotai, “Role of mammogram and ultrasound imaging in predicting breast cancer subtypes in screening and symptomatic patients,” *World Journal of Clinical Oncology*, vol. 12, pp. 808–822, 9 2021.
- [42] S. Shaikh and A. Rasheed, “Predicting molecular subtypes of breast cancer with mammography and ultrasound findings: Introduction of sono-mammometry score,” *Radiology Research and Practice*, vol. 2021, pp. 1–12, 2 2021.
- [43] J. Huang, Q. Lin, C. Cui, J. Fei, X. Su, L. Li, J. Ma, and M. Zhang, “Correlation between imaging features and molecular subtypes of breast cancer in young women (30 years old),” *Japanese Journal of Radiology*, vol. 38, pp. 1062–1074, 11 2020.
- [44] H. Li, Y. Zhu, E. S. Burnside, E. Huang, K. Drukker, K. A. Hoadley, C. Fan, S. D. Conzen, M. Zuley, J. M. Net, E. Sutton, G. J. Whitman, E. Morris, C. M. Perou, Y. Ji, and M. L. Giger, “Quantitative mri radiomics in the prediction of molecular classifications of breast cancer subtypes in the tcga/tcia data set,” *npj Breast Cancer*, vol. 2, 12 2016.
- [45] M. A. Mazurowski, J. Zhang, L. J. Grimm, S. C. Yoon, and J. I. Silber, “Radiogenomic analysis of breast cancer: Luminal b molecular subtype is associated with enhancement dynamics at mr imaging,” *Radiology*, vol. 273, pp. 365–372, 11 2014.
- [46] J. Kim, M. Jang, S. M. Kim, B. L. Yun, J. Y. Lee, E.-K. Kim, E. Kang, and S. Y. Park, “Clinicopathological and imaging features of breast cancer in korean women under 40 years of age,” *Journal of the Korean Society of Radiology*, vol. 76, p. 375, 2017.
- [47] A. G. Bitencourt, N. P. Pereira, L. K. França, C. B. Silva, J. Paludo, H. L. Paiva, L. Graziano, C. S. Guatelli, J. A. Souza, and E. F. Marques, “Role of mri in the staging

- of breast cancer patients: Does histological type and molecular subtype matter?" *British Journal of Radiology*, vol. 88, 2015.
- [48] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0016510720344667>
- [49] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *2017 15th International Conference on ICT and Knowledge Engineering (ICTKE)*. IEEE, 2017, pp. 1–6.
- [50] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65, pp. 386–408, 1958. [Online]. Available: <https://api.semanticscholar.org/CorpusID:12781225>
- [51] L. Vanneschi and M. Castelli, *Multilayer perceptrons*. Elsevier, 1 2018, vol. 1-3, pp. 612–620.
- [52] A. Jain, J. Mao, and K. Mohiuddin, "Ann tutorial-jain1996," *Artificial neural networks: a tutorial*, vol. 29, pp. 31–44, 1996.
- [53] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," pp. 2352–2449, 9 2017.
- [54] J. Schmidhuber, "Deep learning in neural networks: An overview," pp. 85–117, 1 2015.
- [55] R. C. Gonzalez, "Deep convolutional neural networks [lecture notes]," *IEEE Signal Processing Magazine*, vol. 35, pp. 79–87, 11 2018.
- [56] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," pp. 436–444, 5 2015.
- [57] N. O. Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," 2019.
- [58] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, and Z. Xie, "Deep learning and its applications in biomedicine," pp. 17–32, 2 2018.

- [59] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, “Breast cancer detection using deep convolutional neural networks and support vector machines,” *PeerJ*, vol. 2019, 2019.
- [60] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [61] L. G. Falconi, M. Perez, and W. G. Aguilar, “Transfer learning in breast mammogram abnormalities classification with mobilenet and nasnet,” in *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 6 2019, pp. 109–114.
- [62] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, “Deep learning for breast cancer diagnosis from mammograms — a comparative study,” *Journal of Imaging*, vol. 5, 3 2019.
- [63] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, “Rethinking the inception architecture for computer vision,” 2015.
- [64] X. Li, G. Qin, Q. He, L. Sun, H. Zeng, Z. He, W. Chen, X. Zhen, and L. Zhou, “Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification,” *European Radiology*, vol. 30, pp. 778–788, 2 2020.
- [65] T. Mahmood, J. Li, Y. Pei, F. Akhtar, M. U. Rehman, and S. H. Wasti, “Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach,” *PLoS ONE*, vol. 17, 1 2022.
- [66] E. L. Omonigho, M. David, A. Adejo, and S. Aliyu, “Breast cancer:tumor detection in mammogram images using modified alexnet deep convolution neural network,” in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*. IEEE, 3 2020.

- [67] W. M. Salama and M. H. Aly, “Deep learning in mammography images segmentation and classification: Automated cnn approach,” *Alexandria Engineering Journal*, vol. 60, pp. 4701–4709, 10 2021.
- [68] A. Bhandary, P. Madhu, and A. Maier, “Classification of luminal subtypes in full mammogram images using transfer learning,” 2023.
- [69] A. M. Mota, J. Mendes, and N. Matela, “Breast cancer molecular subtype prediction: A mammography-based ai approach,” *Biomedicines*, vol. 12, 6 2024.
- [70] T. Zhang, T. Tan, L. Han, L. Appelman, J. Veltman, R. Wessels, K. M. Duvivier, C. Loo, Y. Gao, X. Wang, H. M. Horlings, R. G. Beets-Tan, and R. M. Mann, “Predicting breast cancer types on and beyond molecular level in a multi-modal fashion,” *npj Breast Cancer*, vol. 9, 12 2023.
- [71] D. Ueda, A. Yamamoto, T. Takashima, N. Onoda, . S. Noda, S. Kashiwagi, . T. Morisaki, . T. Honjo, A. Shimazaki, and Y. Miki, “Training, validation, and test of deep learning models for classification of receptor expressions in breast cancers from mammograms,” *JCO Precision Oncology*, vol. 5, pp. 543–551, 2021.
- [72] H. Cai, J. Wang, T. Dan, J. Li, Z. Fan, W. Yi, C. Cui, X. Jiang, and L. Li, “An online mammography database with biopsy confirmed types,” *Scientific Data*, vol. 10, 12 2023.
- [73] A. S. Alsolami, W. Shalash, W. Alsaggaf, S. Ashoor, H. Refaat, and M. Elmogy, “King abdulaziz university breast cancer mammogram dataset (kau-bcmd),” *Data*, vol. 6, p. 111, 10 2021.
- [74] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, “Data descriptor: A curated mammography data set for use in computer-aided detection and diagnosis research,” *Scientific Data*, vol. 4, 12 2017.
- [75] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, “The cancer imaging archive (tcia): Maintaining and operating a public information repository,” *Journal of Digital Imaging*, vol. 26, pp. 1045–1057, 12 2013.

- [76] C. Cui, L. Li, H. Cai, Z. Fan, L. Zhang, T. Dan, J. Li, and J. Wang, “The chinese mammography database (cmmd): An online mammography database with biopsy confirmed types for machine diagnosis of breast,” *The Cancer Imaging Archive*, 2021.
- [77] N. Al-Najdawi, M. Biltawi, and S. Tedmori, “Mammogram image visual enhancement, mass segmentation and classification,” *Applied Soft Computing*, vol. 35, pp. 175–185, 10 2015.
- [78] P. Carneiro, C. Debs, A. Andrade, and A. Patrocínio, “Clahe parameters effects on the quantitative and visual assessment of dense breast mammograms,” *IEEE Latin America Transactions*, vol. 17, pp. 851–857, 5 2019.
- [79] C. Myles. (2021) Cmmd mammography classification. [Online]. Available: <https://github.com/CraigMyles/cggm-mammography-classification/>
- [80] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 5 2015. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [81] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention u-net: Learning where to look for the pancreas,” 4 2018. [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [82] S. Bhattiprolu. (2021) Segmentation models. [Online]. Available: https://github.com/bnsreenu/python_for_microscopists/blob/master/224_225_226_models.py
- [83] L. Abdelrahman, M. A. Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, “Convolutional neural networks for breast cancer detection in mammography: A survey,” 4 2021.
- [84] I. H. Sarker, “Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions,” 11 2021.
- [85] B. Guanulema, D. Osorio-Ordóñez, H. Vaca-Farinango, A. Mieles-Salazar, F. Villalba-Meneses, G. Cevallos-Bermeo, C. Cadena-Morejón, D. Almeida-Galárraga, and

- A. Tirado-Espin, “Early detection of breast cancer using pretrained alexnet convolutional neural network,” in *2023 Fourth International Conference on Information Systems and Software Technologies (ICI2ST)*. IEEE, 11 2023, pp. 81–88.
- [86] G. Murtaza, L. Shuib, A. W. A. Wahab, G. Mujtaba, G. Mujtaba, H. F. Nweke, M. A. Al-garadi, F. Zulfiqar, G. Raza, and N. A. Azmi, “Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges,” *Artificial Intelligence Review*, vol. 53, pp. 1655–1720, 3 2020.
- [87] A. Petrillo, R. Fusco, E. D. Bernardo, T. Petrosino, M. L. Barretta, A. Porto, V. Granata, M. D. Bonito, A. Fanizzi, R. Massafra, N. Petruzzellis, F. Arezzo, L. Boldrini, and D. L. Forgia, “Prediction of breast cancer histological outcome by radiomics and artificial intelligence analysis in contrast-enhanced mammography,” *Cancers*, vol. 14, 5 2022.
- [88] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig, and J. Torrents-Barrena, “Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network,” *Expert Systems with Applications*, vol. 139, 1 2020.
- [89] M. Wu, X. Zhong, Q. Peng, M. Xu, S. Huang, J. Yuan, J. Ma, and T. Tan, “Prediction of molecular subtypes of breast cancer using bi-rads features based on a “white box” machine learning approach in a multi-modal imaging setting,” *European Journal of Radiology*, vol. 114, pp. 175–184, 5 2019.

Appendices

.1 Appendix 1. Algorithms for Data Pre-processing

Algorithm 1 Breast ROI Extraction Algorithm

- 1: **Input:** Whole-mammogram image (img)
- 2: $\text{img} \leftarrow$ Read the image in grayscale
- 3: $(\text{ori_h}, \text{ori_w}) \leftarrow$ Get the dimensions of img
- 4: cnts, \leftarrow Find the external contours in img
- 5: $\text{areas} \leftarrow$ Array of contour area and calculate areas
- 6: $\text{select_idx} \leftarrow$ Index of the largest contour area
- 7: $\text{cnt} \leftarrow$ Select the index
- 8: $(x_0, y_0, w, h) \leftarrow$ Bounding rectangle of the selected contour
- 9: $x_1 \leftarrow$ Clamp $(x_0 + w)$ within $[0, \text{ori_w}]$
- 10: $y_1 \leftarrow$ Clamp $(y_0 + h)$ within $[0, \text{ori_h}]$
- 11: $x_0 \leftarrow$ Clamp x_0 within $[0, \text{ori_w}]$
- 12: $y_0 \leftarrow$ Clamp y_0 within $[0, \text{ori_h}]$
- 13: $\text{img_broi} \leftarrow$ Extract breast ROI from img using (x_0, y_0, x_1, y_1) coordinates.
- 14: **Return** img, img_roi

Algorithm 2 Apply CLAHE to BROI

- 1: **Input:** BROI cropped image, clipLimit, tileGridSize
- 2: $\text{clahe} \leftarrow$ Create CLAHE object with clipLimit and tileGridSize
- 3: $\text{img_clahe} \leftarrow$ Apply CLAHE to input image
- 4: **Return** BROI enhance image with CLAHE

Algorithm 3 Bounding box crop tumor

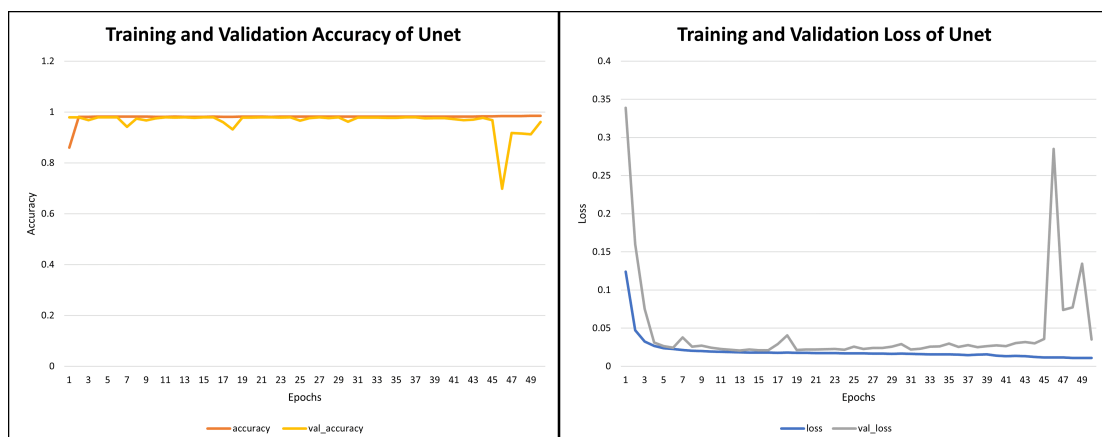
```

1: Input: Original image image, segmented mask mask Verify if mask is binary
2: if mask is not binary then
3:   mask  $\leftarrow$  imbinarize(mask)
4: end if
   Find properties of the region in mask
5: stats  $\leftarrow$  regionprops(mask, 'BoundingBox')
   Assume a single object (tumor)
6: boundingBox  $\leftarrow$  stats(1).BoundingBox
   Extract bounding box coordinates
7: x  $\leftarrow$  boundingBox(1)
8: y  $\leftarrow$  boundingBox(2)
9: width  $\leftarrow$  boundingBox(3)
10: height  $\leftarrow$  boundingBox(4)
   Crop the original image using bounding box
11: crop  $\leftarrow$  imcrop(image, [x, y, width, height])
   Resize the crop to 224x224
12: resizedCrop  $\leftarrow$  imresize(crop, [224, 224])
   Convert resizedCrop to RGB
13: rgbCrop  $\leftarrow$  cat(3, resizedCrop, resizedCrop, resizedCrop)
14: Output: Resized and RGB-converted image

```

.2 Appendix 2. Complementary results of segmentation models (plots)

.2.1 U-net model



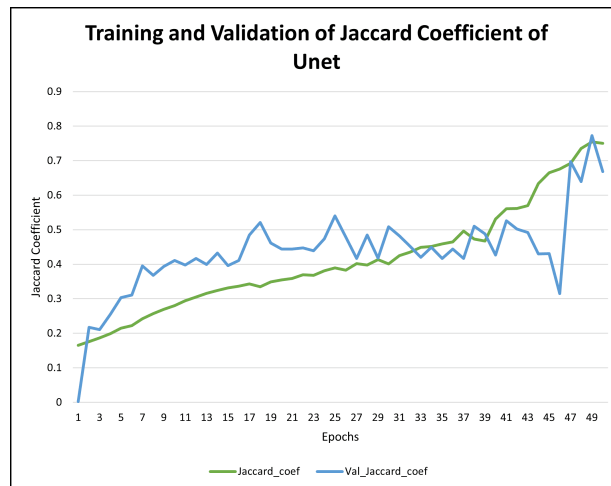


Figure 1: Accuracy, Loss (Top), Jaccard index (Bottom) curves of U-net model.

.2.2 Attention U-net model

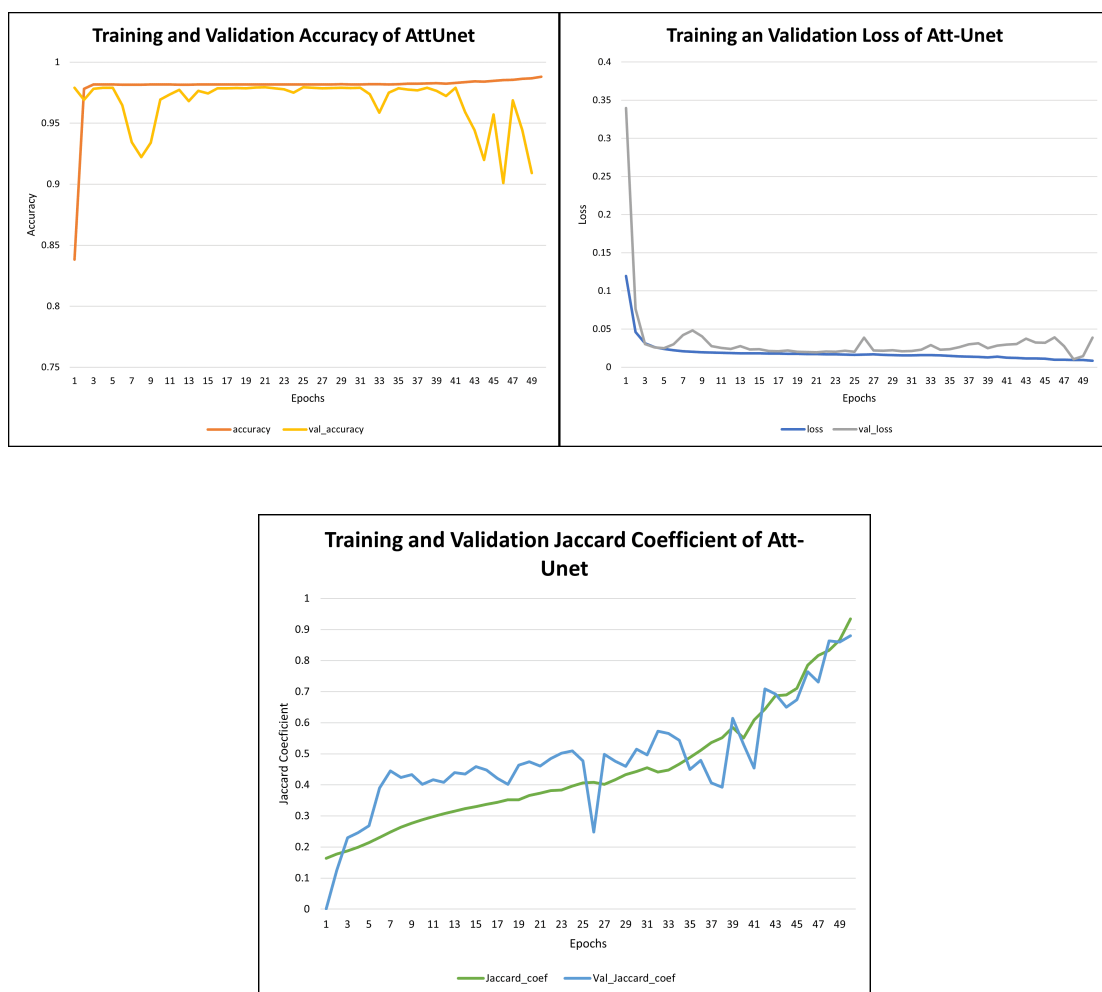


Figure 2: Accuracy, Loss (Top), and Jaccard index (Bottom) curves of Att-Unet model.