الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة سعيدة د. مولاي الطاهر

كلية الرياضيات و الإعلام الآلي و الاتصالات السلكية و اللاسلكية

قسم: الإعلام الآلي

Master's thesis in computer science

Speciality : Computer Modeling of Knowledge and Reasoning

T h e m e

Machine Learning For Early Detection Of Breast Cancer

- **Presented by:**

  AMINE Khadidja

- **Directed by:**

  YAHLALI Mebarka

Academic year 2024-2025

ملخص

لا يزال سرطان الثدي يمثل تحديًا صحيًا كبيرًا في جميع أنحاء العالم . يُعد الاكتشاف المبكر أمرًا حاسمًا لتحسين نتائج العلاج . تستكشف هذه الدراسة تطبيق تقنيات التعلمالآلي للتشخيص المبكر لسرطان الثدي باستخدام البيانات السريرية والتصوير الطبي .تم اقتراح نهجين متميزين : الشبكة العصبية الاصطناعية لتحليل البيانات المجدولة والشبكة العصبية التلافيفية لتصنيف الصور .وقد حقق نموذج الشبكة العصبية الاصطناعية دقة بنسبة100% في تحديد سرطان الثدي، بينما حقق نموذج الشبكة العصبية التلافيفية دقة بنسبة 90% مع بيانات التصوير .تسلط هذه النتائج الضوء على إمكانات الأدوات التي تعتمد على الذكاء الاصطناعي في تعزيز دقة التشخيص . يهدف العمل المستقبلي إلى دمج هذه النماذج في نظام تشخيص متعدد الوسائط والتحقق من صحة أدائها على مجموعات بيانات أكبر ومتنوعة لضمان قابلية التطبيق السريري

الكلمات المفتاحية : التعلم الآلي، التعلم العميق، الكشف عن سرطان الثدي، التشخيص المبكر، الصحة الإلكترونية، الشبكات العصبية الاصطناعية، الشبكات العصبية التلافيفية

Abstract

Breast cancer remains a major health challenge world wide. Early detection is crucial for improving treatment outcomes. This study explores the application of machine learning (ML) techniques for early breast cancer diagnosis using clinical data and medical imaging. Two distinct approaches are proposed: an Artificial Neural Network (ANN) for tabular data analysis and a Convolutional Neural Network (CNN) for image classification. The ANN model achieved 100% accuracy in breast cancer identification, while the CNN model attained 90% accuracy with imaging data. These results highlight the potential of AI-driven tools in enhancing diagnostic precision. Future work aims to integrate these models into a multimodal diagnostic system and validate their performance on larger, diverse datasets to ensure clinical applicability.

Keywords: Machine Learning, Deep Learning, Breast Cancer Detection, Early Diagnosis, E-health, Artificial Neural Networks, Convolutional Neural Networks

Résumé

Le cancer du sein reste un problème de santé majeur dans le monde entier. La détection précoce est cruciale pour améliorer les résultats des traitements. Cette étude explore l'application des techniques d'apprentissage automatique pour le diagnostic précoce du cancer du sein à l'aide de données cliniques et d'imagerie médicale. Deux approches distinctes sont proposées : un réseau neuronal artificiel (ANN) pour l'analyse des données tabulaires et un réseau neuronal convolutif (CNN) pour la classification des images. Le modèle ANN a atteint une précision de 100% dans l'identification du cancer du sein, tandis que le modèle CNN a atteint une précision de 90% avec les données d'imagerie. Ces résultats mettent en évidence le potentiel des outils pilotés par l'IA pour améliorer la précision du diagnostic. Les travaux futurs visent à intégrer ces modèles dans un système de diagnostic multimodal et à valider leurs performances sur des ensembles de données plus vastes et diversifiés afin de garantir l'applicabilité clinique.

Mots Clés: Apprentissage Automatique, Apprentissage Profond, Détection du cancer du sein, Diagnostic précoce, E-health, réseaux de neurones artificiels, réseaux de neurones convolutifs

Acknowledgments

I am grateful to Allah for allowing me the strength and direction to accomplish this thesis effectively.

I would like to thank my supervisor Dr.YAHLALI Mebarka for her support and feedback. I also thank all the professors for their help and advice.

Lastly, I would be remiss not to mention my family, especially my parents, their support has kept my motivation high during this process. I thank every member of my family who has been supportive, even with a small word.

Finally, i would like to express my heartfelt appreciation to my cousin and friends for late night feedback sessions and moral support.

Dedication

To my loved ones who patiently stood by me during this journey. May this work contribute to a future where early detection saves lives.

Acronyms

AI : Artificial Intelligence

ANN : Artificial Neural Network

BC : Breast Cancer

BSE : Breast Self Examination

CAD : Computer Aided Detection

CBIS-DDSM : Curated Breast Imaging Subset of Digital Database for Screening Mammography

CC : Cranio-Caudal

CNN : Convolutional Neural Network

DDSM : Digital Database for Screening Mammography

DL : Deep Learning

GELU : Gaussian Error Linear Units

ML : Machine Learning

MLO : Mediolateral Oblique

PCA : Principal Component Analysis

RELU : Rectified Linear Unit

ResNet : Residual Neural Network

SGD : Stochastic Gradient Descent

VGG Visual Geometry Group

WDBC : Wisconsin Diagnostic Breast Cancer

Contents

List of Figures

List of Tables

General Introduction

Context and problematic

Breast cancer (BC) is one of the most prevalent cancers leading to moratility among women worldwide, with over 2.3 million new cases annually (WHO, 2023)[1].

Early detection plays a crucial role in the prognosis of the disease. Breast screening is performed in women without any symptoms of breast to detect the disease as early as possible. Early detection increases treatment options and improves survival rates, which can be as high as 90% or more with timely intervention. Various screening methods, such as mammography and breast self-examination, play a crucial role in identifying BC in its early stages. However, these techniques still have certain limitations:

1. **Accuracy Limitations:** Mammograms miss about of 20% of cancers (NIH, 2022) [2], leading to delayed treatment.
2. **Resource disparities:** Low-income countries lack access to radiologists, with only 5% of women screened annually (ICO, 2023) [3].
3. **Subjectivity or Human Error in Diagnosis:** Since mammogram results depend on radiologists interpretations, inconsistencies in detection can occur.

To address these challenges, computer aided detection (CAD) systems have become increasingly common in recent years, helping to manage the growing number of BC screenings and the demand for diagnostic accuracy.

Recently, deep learning (DL) –a subfield of machine learning (ML)– has emerged as a promising technology in medical image analysis. By enabling automated feature extraction and improved performance in image analysis, DL methods can assist radiologists in detecting early-stage BC with high accuracy across diverse datasets and imaging modalities.

[1] World Health Organization
[2] National Institutes of Health
[3] International Cancer Organizations

Among these methods, convolutional neural networks (CNNs) have demonstrated exceptional remarkable success in various computer vision tasks, including image classification. Unlike traditional ML approaches, which often depend on manually extracted features and are limited by the quality and quantity of labeled data, DL models can automatically learn complex, hierarchical representations from high-dimensional data. This ability makes them especially well-suited for analyzing medical images and has the potential to significantly improve the accuracy and efficiency of BC detection. The limitations of traditional methods underscore the growing importance of adopting more advanced techniques like DL in medical diagnostics.

Motivation

Growing need for accurate, timely diagnosis of BC, along with the inefficiency of current approaches, requires investigation of DL techniques that can improve diagnostic accuracy, reduce human error, and even save lives with early intervention.

Objective

The primary objective of this thesis is to develop and evaluate a DL-based models for the early detection of BC using multimodal data:

1. **Structured clinical data** (e.g. tumor characteristics, biomarker test results).
2. **Medical imaging data** (e.g. mammograms).

The proposed models aim to classify tumors as benign or malignant, in order to assist radiologists in identifying abnormalities more accurately and efficiently. This work is intended to improve diagnostic accuracy, reduce false positives and false negatives, and support clinical decision-making.

By exploiting DL techniques, capable of automatically learning from both image and clinical data, this thesis seeks to overcome the challenges of limited traditional methods can result in misdiagnosis or inconsistent outcomes, and contribute to more reliable and comprehensive BC detection.

Thesis Organization

The thesis consists of three chapters as follows:

Chapter I: Breast cancer and early detection
Chapter II: Machine Learning & Deep Learning
Chapter III: Experimentation, results & discussion

Chapter 1

Breast Cancer & Early Detection

1.1 Introduction

One of the most prevalent cancers affecting women globally is breast cancer (BC). While BC can also affect men, it is significantly more prevalent among women accounting for nearly 25% of all female cancer cases [11]. In order to improve treatment results and survival rates, early detection is essential, and treatment is frequently simpler when the disease is discovered early. This chapter explores an overview of electronic health (E-health), BC, and common early detection techniques with recent treatment approaches.

1.2 What is E-health ?

The use of information and communication technology for health is called e-Health. It is a general word that encompasses various technologies like mobile devices, smartphones, EHRs [1], prescribing systems, and clinical decision support tools. E-Health aims to improve patient access to care, safety, quality, and cost-effectiveness, especially in rural areas where there is a shortage of doctors. Cancer patients in these areas face challenges due to travel distance and financial burdens. E-Health tools enable remote communication between cancer patients and expert healthcare staff, transforming healthcare delivery [15].

1.2.1 Digital tools for early BC diagnosis

Screening mammography is currently the most effective way to detect BC early. Using models can enhance targeted screening but may increase radiologist workloads. Artificial intelligence (AI) can assist by triaging mammograms, as shown in a trial called MASAI [2] where AI performed equally to traditional methods while reducing workload by 44.3%. This approach raised concerns about

[1] Electronic Health Records

[2] Mammography Screening with Artificial Intelligence

overdiagnosis (can be expensive for health systems). Digital health tools, such as apps and mobile mammography vans, could improve screening in low-income countries but need further research to ensure effectiveness. Overall, these innovations could enhance care and reduce BC fatalities globally [22].

1.3 Breast cancer

BC is a serious disease that arises when the tissue of the mammary glands produces tumors due to aberrant and uncontrollable cell growth [27].

The lobules (The glands that make milk), ducts, or connective tissue of the breast are typically where BC starts. Milk is drained from the lobules to the nipple via the ducts, which are tiny tubes. Everything is held together by the connective tissue, which is made up of fatty and fibrous tissue. The majority of breast malignancies start in the ducts, known as DCIS [3]. Non-invasive tumors do not spread to other tissues (non-invasive malignancies can be precancerous). When breast tumors develop into healthy tissue, they become invasive. Eventually, they may spread through blood and lymph vessels to other parts of the body. Successful BC treatment is more likely when it is discovered early, before it has spread (treatment becomes more challenging, women's chances of surviving the disease are significantly reduced) [36].

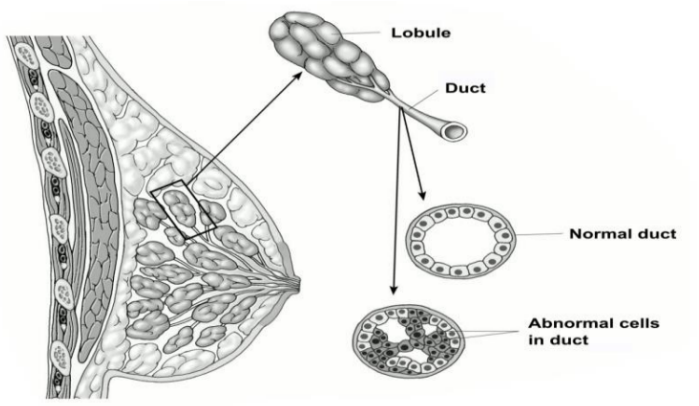

Figure 1.1: Anatomy of female breast [31]

1.3.1 Incidence of cancers for females :

The incidence and fatality rates of BC vary by geographic location, making it one of the leading causes of death for women worldwide. This disease affects women globally and is a major public health concern [27].

[3] ductal carcinoma in situ

In 2022, BC became the most frequently diagnosed disease in women in Algeria and is now the most common cause of cancer death for women in country, accounting for more than 41% as shown in 1.2 of all female cancer mortality according to the GCO [4].

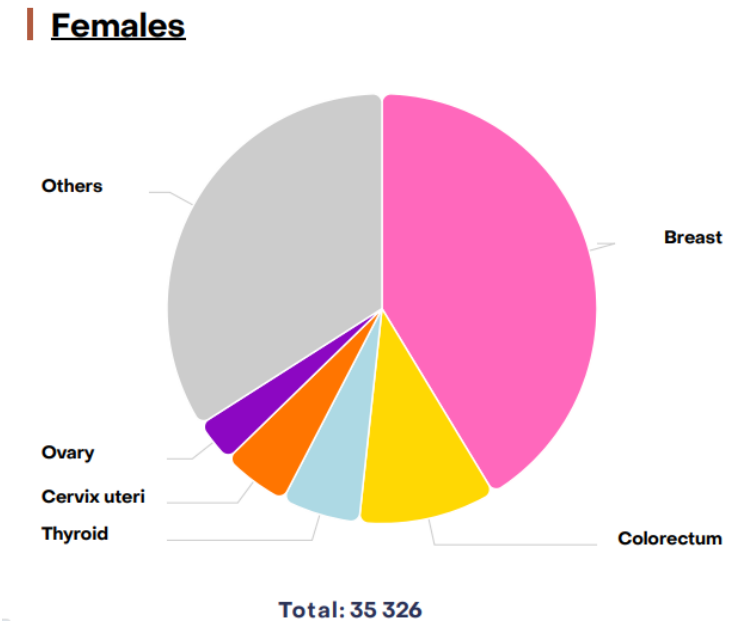

Figure 1.2: Estimated incidence of female cancer cases in Algeria 2022 [1]

1.3.2 Tumor types :

A tumor is a bodily mass of cells that is not normal. Cells that divide excessively or fail to die when they should are the causes [25]. We distinguish 2 categories of breast tumors:

1.3.2.1 Benign tumor

A benign breast tumor is a non-cancerous mass, which means that it has no chance of spreading. Occasionally, these tumors can continue to grow, pressing on the organs and causing discomfort or other complications, despite the fact that they are often not aggressive against the surrounding tissues. When the tumor is removed in these cases, the associated discomfort or problems subside [17]. Certain types of benign tumor have the potential to develop into malignant tumors. These are closely watched and might need to be removed surgically [25].

1.3.2.2 Malignant tumor

Uncontrollably growing cells that spread locally or to other locations are characteristics of malignant tumors. Tumors classified as malignant are cancerous, they use the lymphatic or circulatory

[4] Global Cancer Observatory

systems to travel to far-off places. Metastasis is the term for this spread. Treatment is necessary to prevent the rapid spread of malignant tumors [25]. The physician performs a biopsy to assess the severity of the tumor when it is suspected of being malignant [17].

1.3.3 Diagnostic Screening methods for BC :

Breast cancer screening involves the medical evaluation of women to achieve an earlier diagnosis. The premise is that early diagnosis can lead to better outcomes. Various screening methods have been utilized shown below in 1.4.

1.3.3.1 Clinical breast examination/ Breast Self Examination

CBE is a way of assessing the breasts that combines visual inspection and palpation. It is performed by a health professional and often includes examination of the breast in a variety of standing, sitting, and recumbent postures. The CBE test's sensitivity ranges from 40% to 59%. BSE, like CBE, assesses the breast through palpation and visual inspection. BSE is performed by the lady herself, whereas CBE is performed by a doctor or nurse. The overall test sensitivity for BSE in screened women is approximately 26%; however, this changes with age. Some organizations continue to support BSE as a means of raising awareness about breast health issues [37].

1.3.3.2 Mammography

The mammogram is a method that employs low amounts of radiation to make images of the breast. Mammograms can be used to screen and diagnose both symptomatic and asymptomatic patients. A 2D mammography captures images of only the front and side of the breast. Tomosynthesis, commonly known as 3D mammography, produces X-ray images of the breast by rotating and examining in many dimensions. Combining 3D and 2D mammography can significantly improve detection accuracy when performed by a specialist. Women with dense breasts should not receive a mammography because the overlap of normal fibro-glandular tissues makes false alarms more likely [16].

This is the main imaging method for the detection of breast cancer. The sensitivity of mammography, which is approximately 67.8%, can change based on personal characteristics, age, and race. It is crucial to remember that the operator's proficiency and the operation of the equipment can both impact how effective mammography is. For women over 40, mammograms are highly advised since they enable the early diagnosis of tumors before they cause symptoms [27].

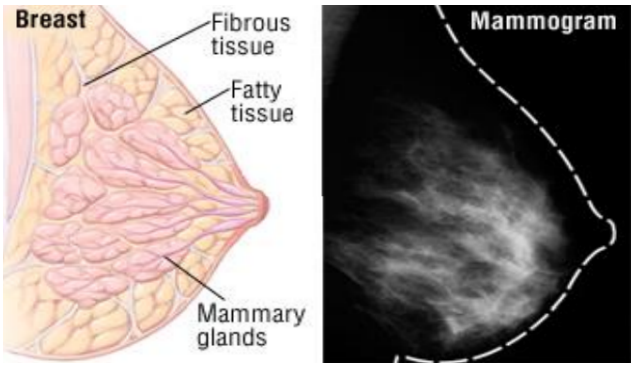

Figure 1.3: Mammogram illusatration [31]

1.3.3.3 Ultrasound imaging

Especially for younger women, pregnant women, women with dense breasts, and women with breast implants, ultrasound is frequently used as a supplement to mammography. Although this method has a 98% specificity, its significant operator dependence may affect the accuracy of the diagnosis. Ultrasound is important for separating benign from malignant lesions because it can distinguish between solid and cystic tumors [27].

1.3.3.4 Magnetic Resonance Imaging (MRI)

For women who are at a high risk of developing breast cancer, especially those who have a family history of the disease or genetic predispositions, this is advised. With a sensitivity range of 70% to 96% and a specificity range of 67% to 100%, magnetic resonance imaging (MRI) is especially helpful in identifying malignancies in dense breast tissue. Detailed imaging is possible with this technique, revealing structural abnormalities that are difficult to see with conventional methods [27].

1.3.3.5 Positron Emission Tomography (PET) and Computed Tomography (CT)

These tests are often saved for more serious cases in order to track the response to therapy or find metastases. The sensitivity and specificity of CT scans are 91% and 93%, respectively. PET scans have a sensitivity of 61% and a specificity of 80%, making them very useful for evaluating the response to treatment. These techniques are essential to detect and stage the course of the disease [27].

1.3.3.6 Biopsy

Of the many biopsies, core needle biopsy gathers larger tissue samples and frequently requires local anesthetic, while fine needle aspiration involves taking a tiny number of cells [27].

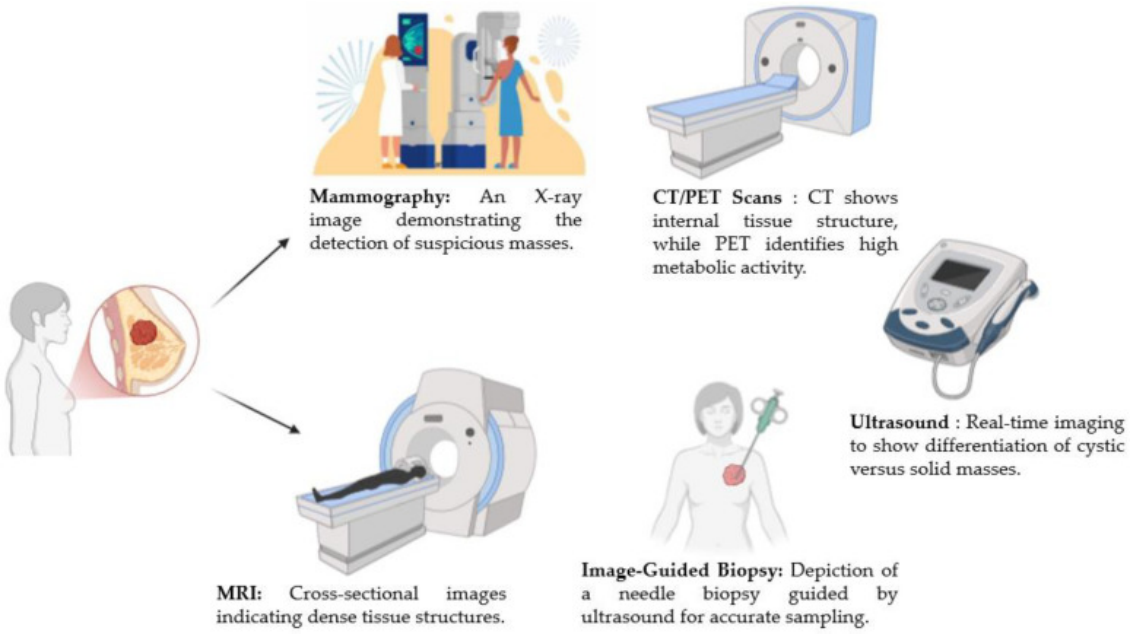

Figure 1.4: Overview of imaging modalities in breast cancer detection [27]

1.3.4 Abnormality types in breast :

The broad use of breast cancer screening has led to the detection of many purely radiological abnormalities.Mammography uses a variety of viewpoints, including the craniocaudal (CC) and medial-lateral oblique (MLO), to better understand these abnormalities, which consist of masses and calcifications.

1.3.4.1 Calcifications

There are two categories of calcifications, that appear on mammograms as tiny, luminous dots:

- **Microcalcifications:** These are microscopic (<0.5 mm) calcium deposits in the breast that resemble sand grains. They are made up of 90% calcium phosphate, which can be either benign or malignant, and 10% calcium oxalate crystals, which are always benign.Because of their high density, they are easy to see on mammograms.They are easily visible on mammograms. Analyzing the shape of microcalcifications during screening is crucial for differentiating between benign and malignant calcifications [17].
- **macrocalcifications:** Sizable calcium deposits in the breast. Women over 50 are more likely to have them. They are frequently linked to benign illnesses or changes that are taking place in the breast, like previous lesions [17].

1.3.4.2 Masses

A mass is a Large opaque that take up space in the breast and show up as white spots on mammograms. They may be a solid lesion, such as breast cancer, or a cyst, which is a non-cancerous accumulation of fluid. For instance, doctors can assess their density and shape based on many parameters. Doctors can define it using several characteristics, such as its density, shape, and contour, to determine its nature [17].

1.3.5 BC Staging and Classification (grades):

Upon diagnosis, cancer is categorized into stages according to its level of progression. Staging is useful for guiding the treatment strategy and evaluating the prognosis of cancer. In general, stages of breast cancer can be classified as invasive or in situ. Stages can be assigned a number (0 through IV) [32].

1.3.5.1 Stage 0, Non-invasive Breast Cancer

The most common way to identify DCIS (a type of pure, non-invasive cancer) is by looking at mammograms that show microcalcifications limited to the breast ducts. Up to 40% of DCIS instances will develop into invasive breast cancer if treatment is not received [35].At stage 0 survival rate during 5 years is almost 100% [23].

1.3.5.2 Stages I–III, Early Invasive and Locally Advanced, Non-metastatic Breast Cancer

A breast cancer tumor that is less than 2 cm in size is considered to be in its first stage. And that the tumor has either not yet metastasized (Stage 1A) to the lymph nodes or other organs, or it has only been able to generate micrometastases in 1-3 lymph nodes in the underarm region (Stage 1B). During the first stage, the five-year survival rate is nearly 100%.

Stage 2A tumors are smaller than 2 cm, with small metastases in mammary glands' lymph nodes. Stage 2B tumors are larger than 2 cm but less than 5 cm, with small cancer metastases in three lymph nodes in the axilla and/or mammary gland. The 5-year survival rate at stage 2 is almost 93%.

Stage 3 tumors have three types: Stage 3A, which is less than 5 cm in diameter, spreads to 9 lymph nodes in the axilla, forms metastases in mammary glands but not internal organs; Stage 3B, which has grown into chest wall or skin but doesn't form metastases in internal organs; and Stage 3C, which has metastases in more than 10 lymph nodes under the arm or above the collarbone.Survival rate in stage 3 is approximately 72% over 5 years [23].

1.3.5.3 Stage IV, Metastatic Breast Cancer

The tumor's ability to spread to other organs, regardless of its size, is the fourth stage. Most of the time, breast cancer spreads to the liver, brain, lungs, or bones. In the fourth stage, the average survival rate for five years is 22% [23].

The histoprognostic grade is a tool used to assess the aggressiveness of a tumor by considering the appearance of cancer cells, the features of cell nuclei, and the number of cells undergoing mitosis. The more aggressive a tumor is, the more it has transformed from normal cells. The histoprognostic grade ranges from 1 to 3, with low grade (I) scores indicating the least aggressive tumors, II scores indicating the least aggressive tumors, and high grade (III) scores indicating the most aggressive tumors. The more aggressive a tumor is, the faster it divides and the more easily the disease can spread [4].

1.3.6 Managing Breast Cancer:

1.3.6.1 Surgery

In tumor treatment, options include lumpectomy (removing only the lump) or mastectomy (removing the entire breast), Depending on the stage and type of the tumor. Surgeons check for clear margins to ensure cancer is fully removed; if not, more tissue may need to be taken. Sentinel lymph node dissection, which removes fewer lymph nodes, is increasingly used and has improved in accuracy from 80% to between 92% and 98% with new techniques [33].

1.3.6.2 Radiation Therapy

Radiation therapy uses high-energy X-rays or gamma rays to target tumors or tumor sites, to eliminate leftover cancer cells after surgery. It effectively kills remaining cancer cells. Other methods, like brachytherapy, have been replaced by electron beam radiotherapy. Usually, radiation is done after surgery over five to seven weeks, with treatments five days a week lasting about 15 minutes each [33].

1.3.6.3 Chemotherapy

Chemotherapy uses anti-cancer drugs to treat cancer cells. Treatment for breast cancer depends on health, history, age, cancer type, and tolerance to medications. It is administered in cycles and can be given before surgery to shrink tumors or after surgery, often every two or three weeks [33].

Other recent approaches exist in management of BC, including: Gene Therapy, Oncogenes Inactivation, Augmentation of Tumor Suppresser Genes, Cell-Target Suicide, Immunomodulation, ect.

1.4 Why mammographic images ?

Mammography,even though it has its limitations, it is widely preferred for breast cancer screening and detection due to several key advantages:

- **Proven Effectiveness in Early Detection and Mortality Reduction:** Studies have shown that digital and film-screen mammography are reliable and beneficial for early BC detection [31], Years before symptoms appear or a lump can be felt during a physical examination, mammograms can identify it.
- **Capacity to Identify Particular Cancer Indicators:** Microcalcifications are especially detectable by mammography. It's possible that other imaging techniques won't notice these minute variations as well.Additionally, it is capable of identifying malignant masses and architectural distortions.
- **Developed Screening Criteria:** For women in particular age ranges, major health organizations advise routine mammography screening. These widely accepted recommendations show the demonstrated advantages of mammography and are supported by substantial research.
- **Ongoing Technological Developments:** 3D Mammography (Digital Breast Tomosynthesis) [16], a more recent technology that allows radiologists to examine thin tissue slices by creating a 3D reconstruction of the breast. It has been demonstrated to lower false positives and increase cancer detection rates, particularly in women with dense breasts.

1.5 Related work

Various works have been introduced in this domain from machine learning to deep learning approaches. Recent articles (from 2021 to 2025) regarding the application of ML/DL in BC detection are reviewed.

H.Alshayeji et al.[6] used an artificial neural network model with a hidden layer to predict BC using the Wisconsin breast cancer (WBCD) and the Wisconsin diagnostic breast cancer (WDBC) datasets without employing feature optimization or selection algorithms, for BC detection using WDBC, it achieved an average accuracy of 99.47%, precision of 98.71%, and 99.56% AUC.

Bokhare et al.[10] used classifiers: k-nearest neighbor, decision tree classifier, SVM, random forest, SVM kernels, logistic regression, Naïve Bayes, to test their effectiveness in WDBC dataset. They compared them with each other and found that the decision tree classifier gets the highest accuracy 97. 08% among all models.

E.Poornima et al. (2023) [28] predicted BC using a hybrid ML methodologies (SVM and PCA) in the WDBC data set and achieved an accuracy of 96%.

La Moglia et al.(2025) [20] analyzed a BC data set with 11 features, using 8 machine learning classifiers.The results showed that logistic regression achieved the highest testing accuracy 91.67% without feature selection. Classifiers like LGBM improved, with a notable 90.74 % accuracy After applying feature selection.

H Shekhar et al.(2023) [14],compared the performance of his proposed shallow CNN architectures against pre-trained deep CNN architectures in CBIS-DDSM and INBreast data sets, Proposed Xception classifier with fine-tuning with range of 87–91 accuracy for CBIS-DDSM and 91 to 1.00 for the INbreast dataset.

Yusof et al.(2024) [39] developped CNNs architecture & compared its accuracy accuracy with deep architectures, ResNet50 V2, VGG19 with the CBIS-DDSM dataset, reaching the accuracy 64%, 70%, 72%, respectively.

C Murthy et al. (2024) [24] developed a hybrid model consisting of CNNs and stochastic gradients in the CBIS-DDSM and Wisconsin Breast Cancer Databases, reaching an accuracy of 96% precision of 95%, 95% recall, and an AUC-ROC value of 0.96.

M Keddas et al.(2025) [18] combined CNN with LSTM to improve classification accuracy. They compared the proposed model with other DL models, such as CNN, LSTM, Gated Recurrent Units (GRUs), VGG-16, and RESNET-50. The CNN-LSTM model achieved superior performance with accuracies of 99. 17% and 99. 90% in the Hispathological images data set, the Breahis data set, respectively.

Chapter 2

Machine Learning & Deep Learning

2.1 Introduction

For decades, machine learning (ML) has focused on developing models to extract patterns from data. However, it required extensive understanding of statistical methods, computing, and data, which were not readily available. But, during the previous two decades, powerful deep learning (DL) models with great accuracy can be created.

In this chapter, The goal is to give a basic theoretical knowledge of ML and DL, covering essential aspects. Specifically, we explore the principles of Artificial Neural Networks (ANNs) and their evolution into Convolutional Neural Networks (CNNs), which have revolutionized the field of computer vision. We begin by outlining the fundamental concepts of ANNs. Then, we transition to a detailed examination of CNNs, emphasizing their unique architectural components. This chapter will also provide an overview of deep CNN architectures. we discuss the specific techniques used within our DL framework, addressing aspects such as data pre-processing, image augmentation, ect.

2.2 ML & DL

The figure 2.1 depicts the relationship between artificial intelligene (AI), ML and DL.

ML is the study of computer algorithms that automatically improve with experience . It simplified programming by requiring only training data as input, consisting of pairings of data (features) and expected outputs (labels). While ML may identify patterns and rules on its own, the ultimate goal is to improve accuracy, which may result in errors [7].

Deep networks differ from neural networks in that they have more neurons and deeper layers, have more intricate connections with thousands to millions of parameters, and automatically extract features. The architecture has changed over the previous two decades, and fresh research into

deep networks is ongoing [8].

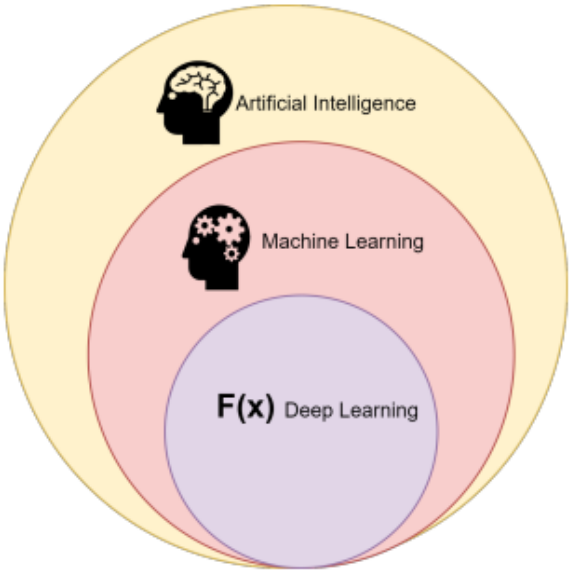

Figure 2.1: AI vs ML vs DL [34]

2.2.1 Supervised learning

Supervised learning is a ML task that involves mapping inputs to outputs using example input-output pairs. Supervised ML techniques require external support. The input data set is separated into training and testing sets. The train data set includes an output variable that needs to be predicted or classified. Algorithms employ patterns learned from training datasets to predict or classify data on test datasets [21].

2.2.2 Unsupervised Learning

Unsupervised learning differs from supervised learning in that it does not require accurate responses or a tutor. Algorithms are autonomous in identifying and presenting interesting data structures.Unsupervised learning algorithms only learn a few features from the data. When new data are introduced, it leverages previously learned features to classify them. It is used primarily for clustering and feature reduction [21].

2.2.2.1 Principal Component Analysis

Principal component analysis (PCA) is a statistical process that applies an orthogonal transformation to convert correlated observations into principal components uncorrelated linearly. This

reduces data dimensions for faster and easier computations. This method explains the variance-covariance structure of variables using linear combinations. This approach is widely used to reduce the dimensionality [21].

2.2.3 Computer Vision

a field of AI and the science of acquiring, processing, and analyzing digital images to extract meaningful information and perform actions. Computer vision mimics the human visual system in an accurate and efficient way. Computer vision has three layers as follows: Low-level (basic image processing functions such as recognizing object edges), Intermediate-level (recognizing various visual objects, such as animals, people, and digits (see Figure 2.2)), High-level (high-level descriptions for the images) [9].

Computer vision is commonly used in marketing, healthcare (to detect cancerous tumors in healthy anatomy), retail, and automotive.

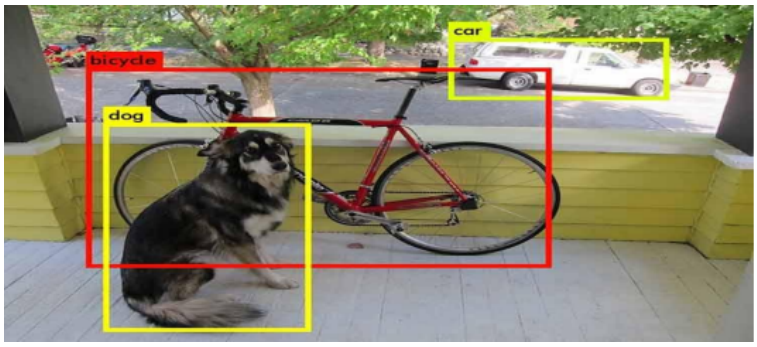

Figure 2.2: Example of computer vision detection

2.2.3.1 Image classification, object detection, segmentation

Image classification [34], object detection [34] or recognition, and object segmentation [34] are examples of common computer vision applications, see in Figure 2.3.

1. **Image classification :** is the task of classifying an image independently of where the object appears in the image. Classification networks typically receive a single object as input and output the label or name of the most visible object on the image.

2. **Object detection :** is a computer vision challenge that involves predicting the location and class of known objects in a picture and drawing a rectangle bounding box around them. Simply described, it is a classification with localization, but the image may comprise many objects. The second image, displays an example object detector output.

3. **Segmentation:** is a type of pixel-level image classification that assigns same color value to each object of the same class or instance, resulting in a color blob that resembles the

objects. Instance segmentation (assigns different colors to each object of a comparable class),.Semantic segmentation (distributes comparable colors to pixels of the same class, regardless of whether they correspond to different objects within the same class).

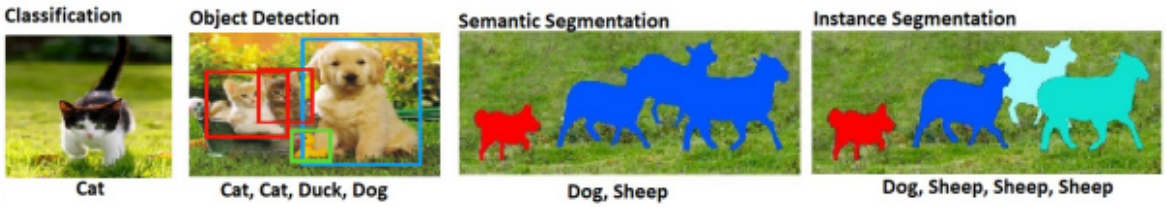

Figure 2.3: Illustration of different tasks of computer vision [34]

2.2.4 Artificial Neural Networks

ANNs, inspired by the human brain and biological neural networks, are part of AI. ANN learns gradually by minimizing its cost function. Layers are stacks of artificial neurons with varying sizes and transformation functions applied to inputs. Artificial neurons produce a nonlinear function based on the sum of their inputs. The inputs are multiplied by the weights of the neurons. The sum of the products is then provided to the activation function f as shown below.

$$f(x_i, w_i) = \phi\left(\sum_{i=1}^{n} w_i x_i\right)$$

x_i: the ith input
w_i: the ith weight of the neuron
ϕ: the activation function
n : the number of inputs

Figure 2.4 shows that the nonlinear function is partially weighted and modulates as the signal travels over a connection. This is an iterative procedure; from the first to the last layer, the weights of the neurons are optimized throughout each iteration [8].

The feedforward process involves computing new neurons from preceding layer neurons, weights, and biases as they traverse the network forward.

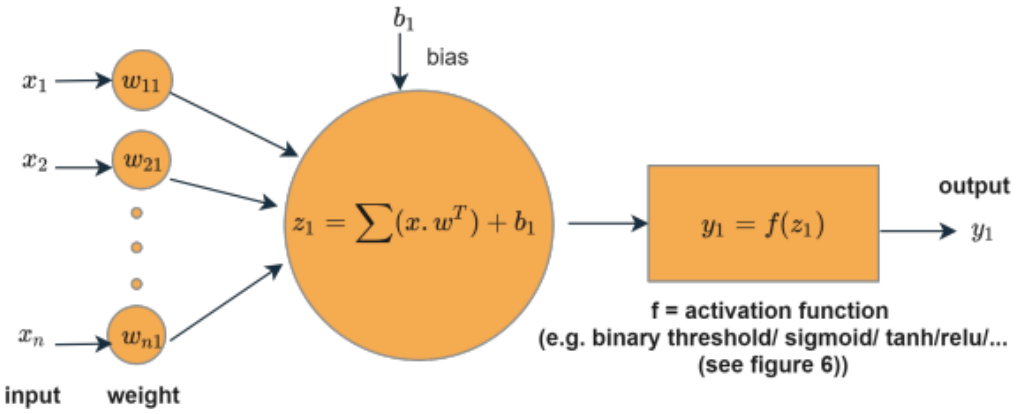

Figure 2.4: Structure of a neuron in ANN [34]

2.2.4.1 Multilayer Perceptron

The term "multilayer" refers to a feed-forward ANN that has three different types of layers: input, output, and hidden layers, as seen in Figure 2.5. The model is intended to tackle linearly inseparable problems and generates information from input to output direction. The output layers finish tasks like classification and prediction after the input layers process the input signals. Between the input and output layers, which are used for calculation, comes the most crucial process: the hidden layers [38].

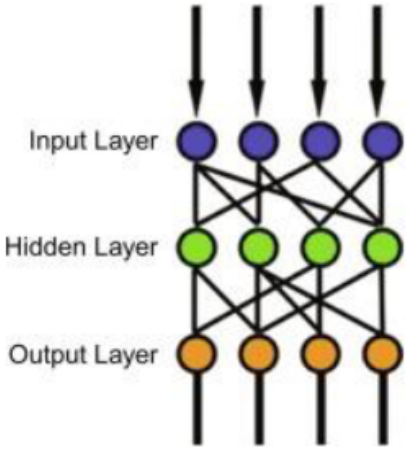

Figure 2.5: Multilayer Perceptron [38]

2.2.4.2 Activation functions

Activation functions are important for neural networks because they allow them to learn non-linear relationships. The first ones used were the sigmoid and Tanh functions.These saturated activation functions were widely used because of their computational simplicity, and they produced excellent results. They found that nonsaturated activation functions improve neural network robustness by

decreasing the vanishing gradient problem caused by backpropagation in feedforward networks. Since introducing Relu function, various non-saturated versions have been introduced [19].

We will get familiar with these activation functions (shown in figure 2.6 and talk about them in the following:

1. **The sigmoid activation function:** One of the first activation functions utilized. its output : $(0 \leq sigmoid(x) \leq 1)$. The sigmoid function is deemed saturated due to its boundaries [19].

$$f(x) = \frac{1}{1+e^{-x}}$$

   x : the input to the activation function

2. **Tanh's Activation Function:** Tanh functions are widely utilized in neural networks. It spans from -1 to +1(-1 $\leq tanh \leq$+1) [19]

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

3. **ReLU Activation Function:** Nair et al. (2010) invented the Rectified Linear Unit (ReLU) to improve neural network robustness by learning complex nonlinear functions. It removes negative values. the ReLU range is $(0 \leq Relu(x) \leq \infty)$.This is one of CNN's most significant activation layer as it eliminates negative gradients when the threshold is zero [19].

$$ReLU(x) = \max(0, x)$$

4. **Leaky ReLU Activation Function:** the Leaky Relu adds a new hyperparameter α to this one and allows for a tiny negative slope to address the issue of dying neurons when employing ReLU. Its range is $-\infty \leq LeakyRelu(x) \leq \infty$ [19].

$$LeakyReLU(x) = \begin{cases} \alpha x & if x < 0 \\ x & if x \geq 0 \end{cases}$$

5. **Softmax Activation Function :** Softmax is an activation layer at the end of a network that generates a discrete probability distribution vector [19] shown in equation.

$$Softmax(\mathbf{z_i}) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

   Other activation functions: GELU, SELU, ELU, ect.

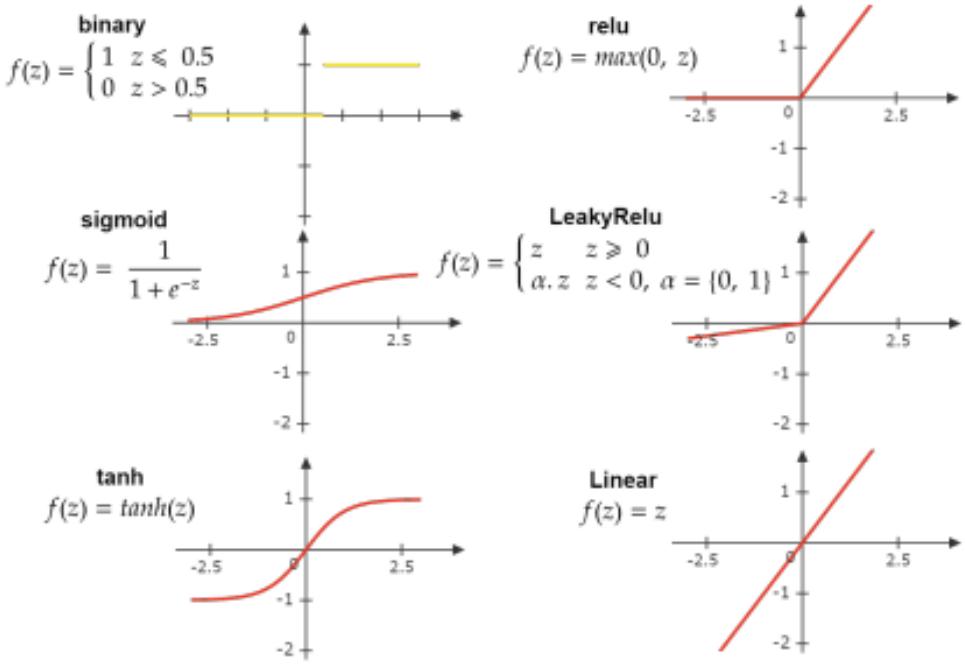

Figure 2.6: Plot of different activation functions [34]

2.2.4.3 Back propagation

We calculate the error at the final output layer by comparing it to the expected target output and then send this error backward through the network. Neurons that cause this error adjust their weights and biases, helping to trigger the right output next time the same input is provided. Backpropagation is an important benefit of ANNs over standard ML techniques allowing end-to-end learning without human help. It uses partial derivatives to understand how each layer's weights and biases impact the final error [34].

To reduce the loss function and so enhance the accuracy of the neural network, weights and biases must be adjusted. This technique calculates the error contribution of each weight. Backpropagation refers to the process of computing gradients, not weight updates [29].

2.2.4.4 Error functions (Loss Function) :

The loss function calculates the difference between predicted and real values.

1. **The Mean Squared Error (MSE):** also known as L2 loss, is commonly used in regression situations. This calculates the average of the squared distances as defined below where y : the true value and y^: the predicted value [29].

$$L(y,\hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

2. **Cross-Entropy loss :** It is mostly used for classification and defined as:

$$L(y_i,\hat{y}_i) = -\sum_{c=1}^{M} Y_{i,c}\log(f_k)(p_I,c))$$

yi,c indicates if the class label c is true for the i-th observation, pi,c predicts the probability of class c, and M represents the number of classes [29].

2.2.5 Image Processing

The most significant challenge when using a fully connected MLP for image processing is Excessive number of parameters, for this reason Convolutional neural networks (CNNs) has been introduced, we will dive into it in this section:

2.2.5.1 Convolutional Neural Networks

CNNs are distinguished from other neural networks by their high performance with image, speech, or audio signal inputs. Similarly, CNN can be used for text analysis by recognizing characters and analyzing words as discrete textual units. However, CNN is better recognized for image recognition. Figure 2.7 depicts CNNs used in computer vision [8].

The CNN's complexity grows with each layer, allowing it to identify more of an image. Earlier layers focus on basic elements like colors and borders. As visual input travels through CNN layers, it recognizes larger forms and eventually identifies the target object [9].

Figure 2.7: Computer vision and CNN [8]

CNNs may align neurons in three dimensions based on length, height, and depth, which can be mapped to image width, height, and RGB channels. CNNs use multiple connected layers to alter input images and output a class probabilities. All CNN architectures share some similar layers, as seen in the figure 2.8

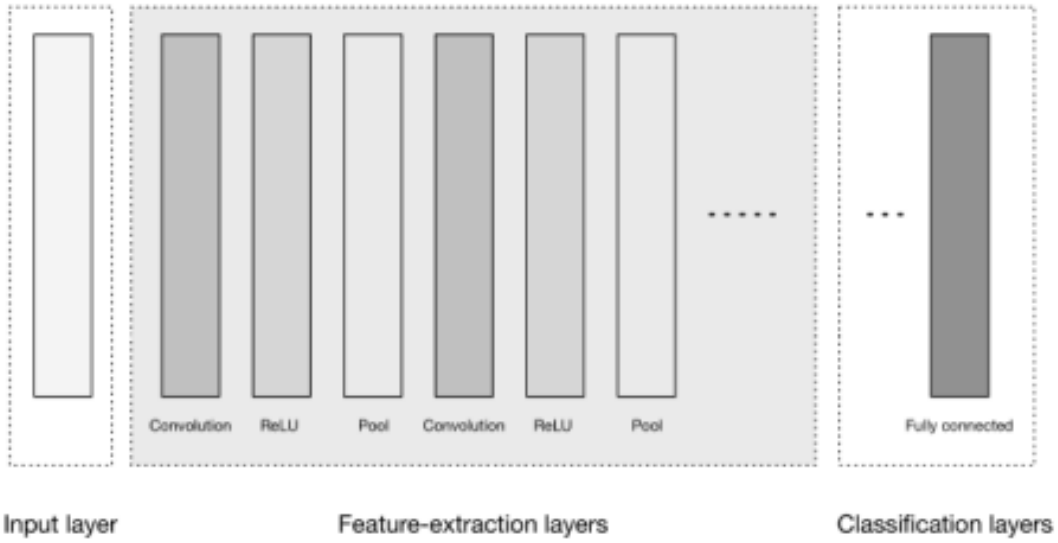

Figure 2.8: CNN general architecture [8]

2.2.5.2 CNN layers

We will discuss of these layers with details in the following:

1. **Convolution layers :**

   Convolutional layers are the primary building blocks of CNN architectures.They apply filters or kernels to the input image. The layer generates the feature map by performing a dot product between the input layer's neurons and filters [8]. It applies the kernel to the pixel values in different locations and slide the window of the filter over the entire image .Figure 2.9 depicts a convolution layer with input and output volumes.

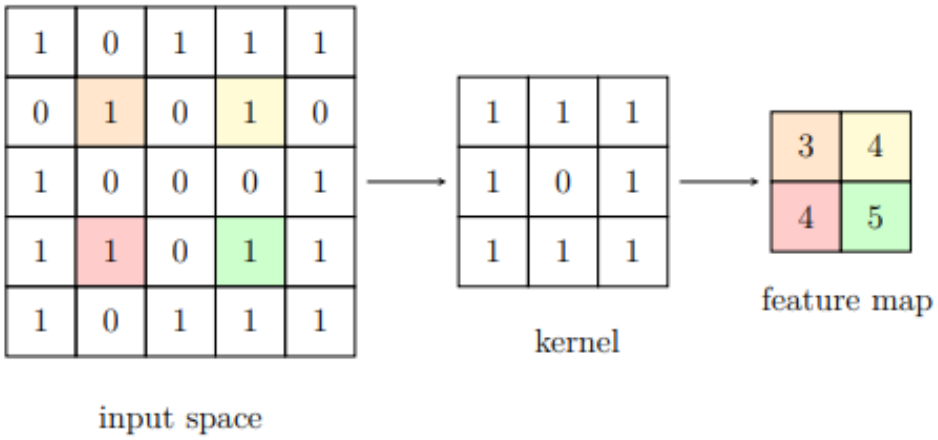

Figure 2.9: Example of a convolution operation using a 3x3 kernel and a stride of 2 [29]

After each convolution operation, a CNN applies a Rectified Linear Unit (ReLU) transformation to the feature map, introducing nonlinearity to the model [9].

Stride

is defined as the kernel's jump from the current to the next pixel center, an image's kernel center location changes by S in both dimensions. Higher strides lead to smaller feature maps [29] .it means how much we want to move the filter after applying it to one location.

Padding

Padding is employed when reducing the dimension of a feature map will result a loss of information at the input space's borders. Pixels are added around the input space to boost the feature map's dimension. The pixels have values of 0, which is referred to as zero padding [29] In Figure 2.10.

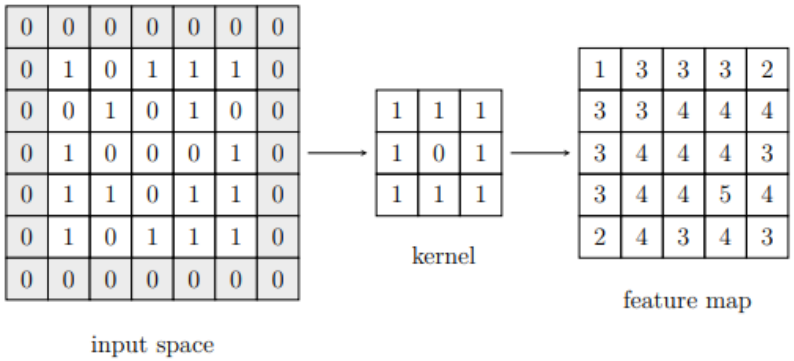

Figure 2.10: Example of a 3x3 kernel convolution operation with a zero padding and 1 stride [29]

Padding is used to control the output size to avoid noise. The standard for padding is zero padding (By default). big advantage is that we can apply our filter to all locations so that the output effectively has the same size as the input.

Convolution filters:

Kernels are the two-dimensional slices of the filters. These filters provide translation invariance and parameter sharing [7], that means that once the filter learned a feature, it can recognize it anywhere in the image. the main advantage of using multiple filters in a convolutional layer is that It enables the detection of different features simultaneously.

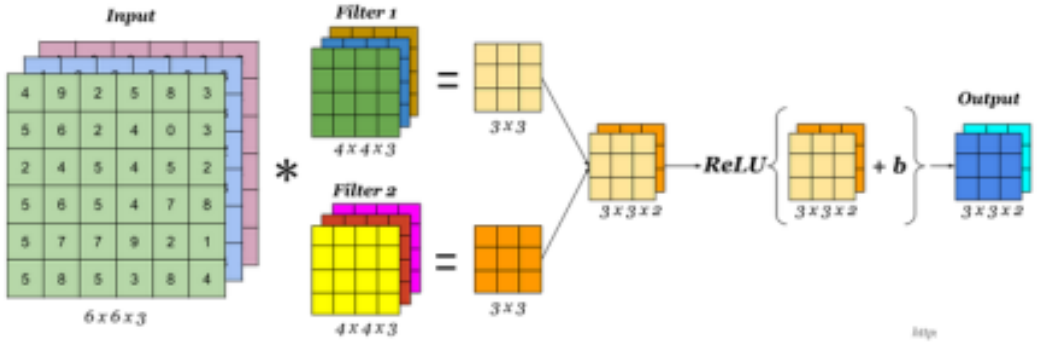

Figure 2.11: Visualization of the use of filters in Convolution Operation [7]

by default filters are applied to all input channels,this allows network to learn different features from same layer.

2. **Pooling layers (downsampling):**

Pooling layers are introduced following each convolutional layer. They remove noise and extracts required pixels from the convolution layer's output feature map. The pooling layers have two key parameters: filter size and stride [19] Pooling layers help to decrease data representation while also controlling overfitting[8]

(a) max-pooling : retains the maximum value for each grid. most important information is preserved and This method is more commonly employed than average pooling. Shown in 2.12

(b) average-pooling :The filter calculates the average value in the receptive field and sends it to the output array

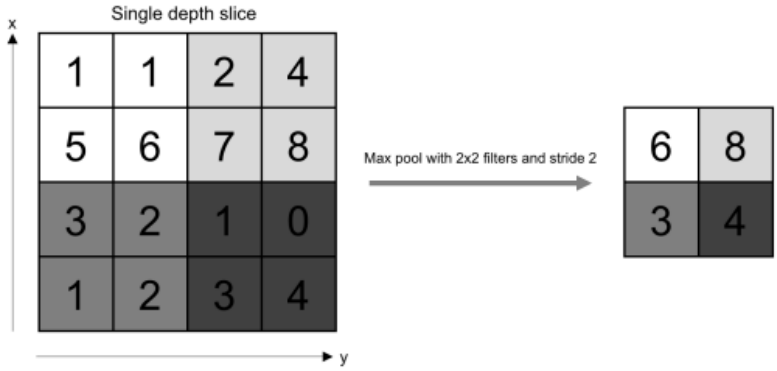

Figure 2.12: max pooling example [13]

3. **Flattening Layers:**

   The pooling layer's output is flattened to a 1D vector, as future dense layers can only accept 1D vectors. Figure 2.13 shows a flattening layer. The resulting vector's dimension is:

$$Dim_{Flat} = Dim_{img} * Dim_{img} * num_{color}$$

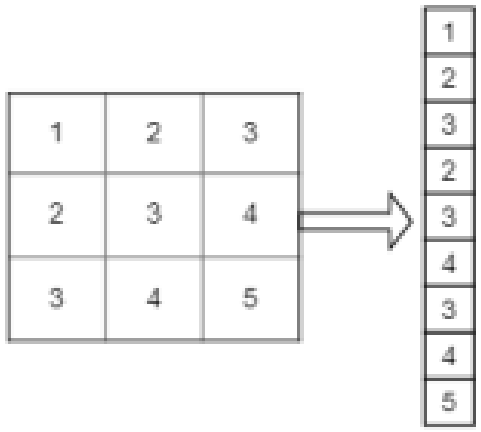

Figure 2.13: Flattening 2D feature maps to 1D vector [19]

4. **Fully Connected Layer (dense layers):**

   they are often placed at the end of a network and take the output of feature extraction layers as input. The dense layer uses features gathered from preceding layers to classify the source image [19]. The fully-connected layer connects each node in the output layer to a node in the previous layer (fig 2.14). FC layers use a softmax or sigmoid activation function to classify inputs, producing probabilities ranging from 0 to 1 [9].

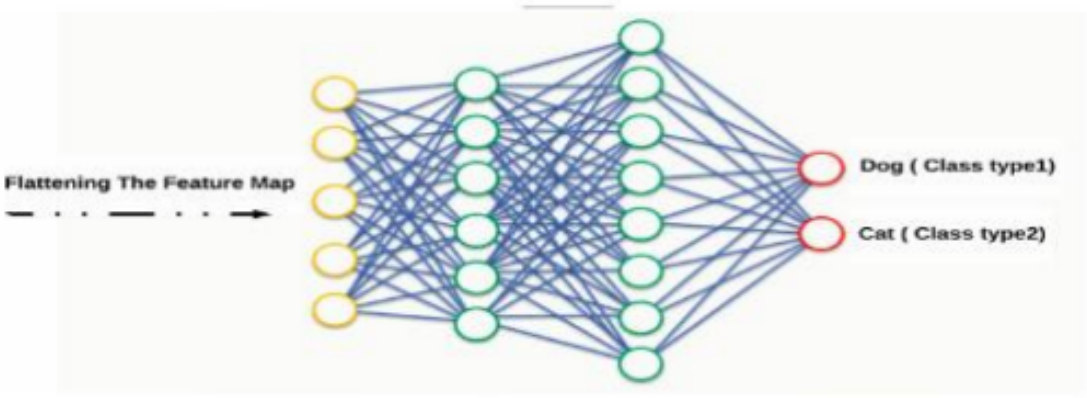

Figure 2.14: fully connected layer architecture [9]

5. **Batch normalization:**

   mini-batches of training data are utilized in forward and backpropagation to optimize the weights of a network. It is used as a network layer, this layer normalizes the input to ensure the mini-batch has a fixed mean and variance across training instances. Batch normalization is typically performed after or before an activation function. It helps with vanishing and exploding gradients. Batch normalization keeps the layer inputs stable, it can achieve convergence, escape local minima, and have a regularization effect [29].

6. **Dropout layer:**

   A dropout layer is a regularization layer, it can be applied at any network layer. During network training, some neurons are inhibited based on a preset dropout-rate probability (P). It can be compared to bagging in neural networks [19].

   We randomly switch off units during training, So in different epochs different parameters are trained and shared .Dropout technique helps reducing overfitting

2.2.5.3 Overfitting vs underfitting:

Overfitting occurs when a neural network performs well on its training set but struggles to generalize predictions to new samples. This is characterized by low bias and high variance. Underfitting occurs when a neural network fails to predict appropriately for both training and validation sets. This is characterized by high bias and variance [7].

2.2.5.4 Optimization and regularization

Training neural networks is a difficult process that might lead to many complications such as overfitting,vanishing and exploding gradients, convergence difficulties, local optima ... Regularization

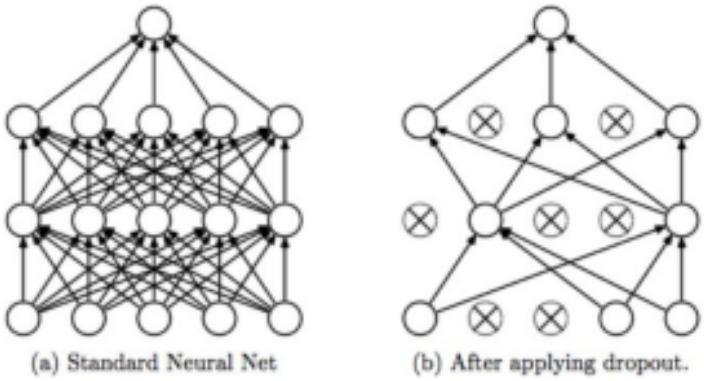

Figure 2.15: NN before and after dropout

is applied to objective functions in unclear optimization problems and its solutions aim to reduce test error in machine learning algorithms [7].

Here are some beliefs and approaches for tackling these challenges:

1. **Stochastic Gradient Descent :** SGD is based on the concept of gradient descent. It is the most commonly utilized optimization algorithm in machine and deep learning. Unlike gradient descent, stochastic gradient descent uses mini-batches instead of the entire dataset. Weights are updated based on the loss function's gradient. This significantly reduces computing time. The learning rate is an important parameter for stochastic gradient descent, it determines the step size for the steepest descent [29].

2. **Adaptive (optimizer) Learning Rate:** Adaptive learning rates adjust for each parameter [29]. Various ways have been introduced over time, the most common one is adam that performs the best and set by default, There is other ones like : Adamax, Adadelta, RMSProp, ect.

3. **L2 Regularization :** L2 regularization penalizes values with a squared term. As a result, it causes the weights to approach zero. It does, however, favor values larger than or less than one over values in the middle. L2 regularization is defined as [12].

$$L_2 = \lambda \sum_{w \in W} w^2$$

4. **Early Stopping:** This can help prevent overfitting in DL, this technique terminates if the validation set error doesn't reduce after a certain number of epochs . The validation dataset is included in the training dataset but is not used during training. If this is the case, the algorithm delivers parameters from the epoch with the lowest validation set error, rather than the most recent epoch [29].

5. **Dataset Augmentation:** Overfitting occurs when an algorithm fails to generalize to previously unknown data, One reason could be a small training dataset. Data augmentation can help to tackle the challenge of Obtaining additional "real" data . It involves transforming data points in several ways without affecting their labels. We will see some typical data augmentation techniques including rotation, flipping, and scaling.

    - **Rotations:** Images can be rotated either by a defined angle or randomly. The minimum and maximum angles are pre-defined.
    - **Flips :** flips mirror an image along either the horizontal or vertical axis.
    - **Scaling:** A defining factor causes the image's size to increase or shrink, this results in an increase in the number of pixels rather than zooming into an image.

In Deep Learning, there are two different approaches for performing data augmentation : offline and online data augmentation. Offline one involves using an augmentation technique to an image to create a new image; these must be kept on disk, while online (real-time) data augmentation involves augmenting images during training. Although images are not saved on disk,the advantage is that there is less control over the augmentation process [29].

2.2.5.5 CNN architectures:

New architectures and algorithms are developped due to the evolution of image classification. There are various architectures of CNNs, which are key in building algorithms [8] In this section, we go through these different architectures:

LeNet-5

LeCun et al. introduced one of the earliest CNN architectures in 1998. It has 5 weighted (trainable) layers: three convolutional and two FC. The first two convolution layers are followed by a max-pooling layer that sub-samples the feature maps. The last convolution layer is followed by two fully connected layers. The final completely connected layer serves as the classifier for the ten digits [9].

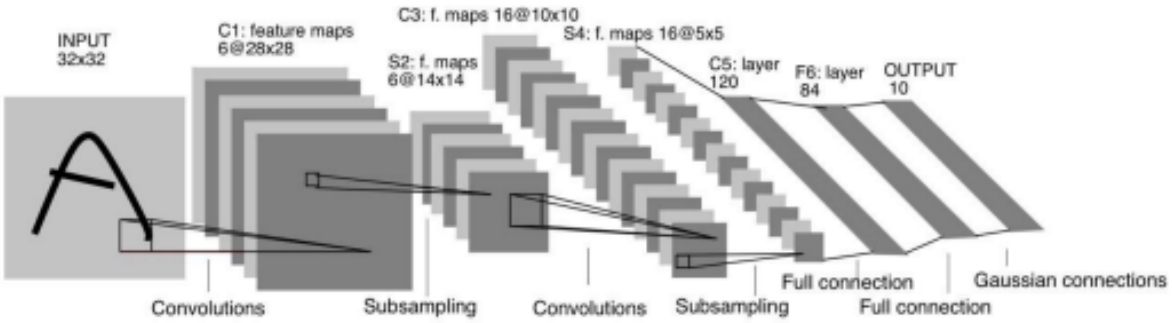

Figure 2.16: LeNet architecture [13]

AlexNet

In 2012, Krizhevky et al. created AlexNet, the first large-scale CNN model, to classify ImageNet data, drawing inspiration from LeNet. The final output layer uses 1,000 units to classify input images into one of the ImageNet dataset's thousand classes.[9] The AlexNet design includes five convolution layers, and three dense layers. It introduced two innovative features: the ReLU activation function and dropout to prevent overfitting [19].

VGGNet

Simonyan and Zisserman introduced VGGNet, a widely used CNN architecture, in 2014. The authors provided six different CNN configurations, with the most successful being VGGNet-16 (configuration D) and VGGNet-19 (configuration E) [9]. VGG19 differs from VGG16 in that it has 19 convolution layers instead of 16. This increases the number of parameters from 138,357,544 to 143,667,240. The fundamental shortcoming of this network is that the authors allocate more weights to the classifier section rather than the feature extraction portion [19].

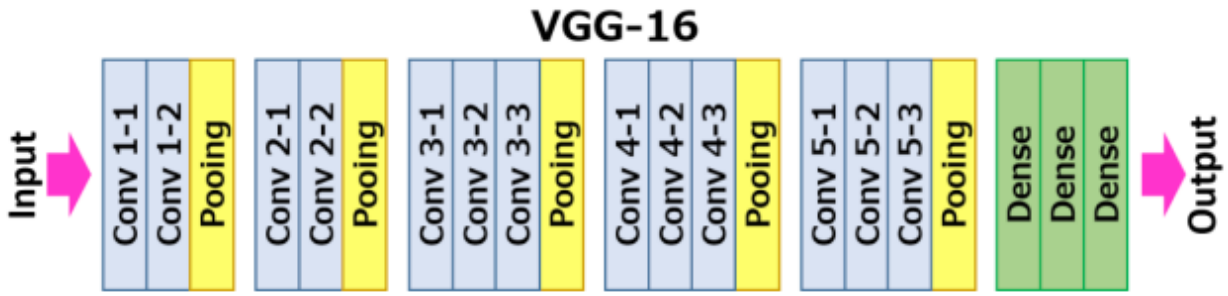

Figure 2.17: VGGNet architecture [9]

GoogleNet (InceptionNet)

GoogleNet was suggested by Szegedy et al. in 2014. The GoogleNet consists of 22 weighted (learnable) layers, with the "Inception Module" serving as the network's main component. The network processes this module in parallel. The resultant feature maps are then combined to produce a high-dimensional output [9]. Following the success of InceptionV1, the authors released additional versions, including InceptionV2 and InceptionV3,wich are the most popular GoogLeNet architectures. An InceptionV1 block and InceptionV3 block are shown in Figure 2.18.

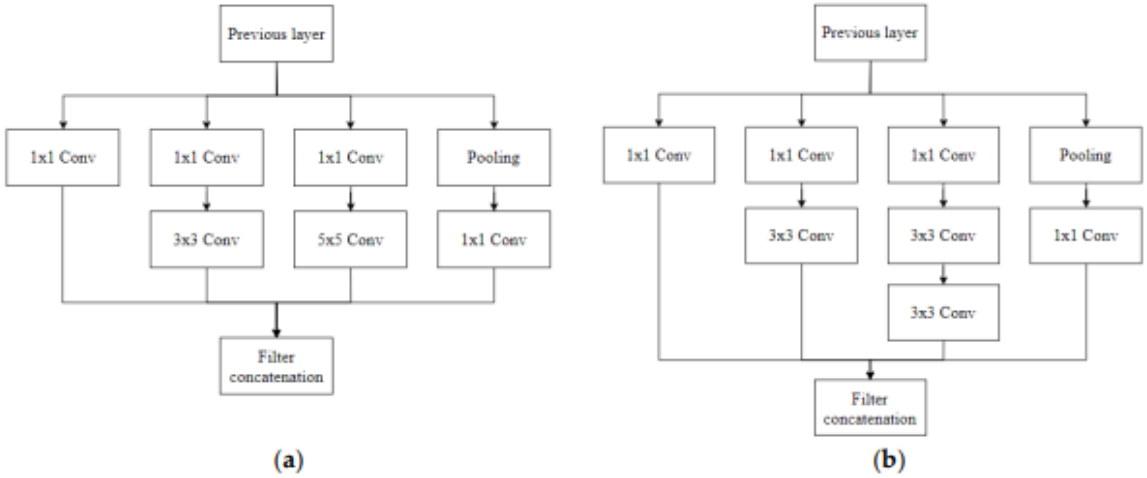

Figure 2.18: Inception blocks. (a) InceptionV1. (b) InceptionV3 [19]

ResNet

Stands for residual network, introduced by He et al. and won the 2015 ImageNet competition with a top-five accuracy rating of 94.29 %. The residual connection figure 2.19 connects convolutional layers before passing to the ReLU activation layer. It ensures that weights from preceding layers do not vanish (get lost) during backpropagation. This network has three versions, each with a different number of layers: ResNet50, ResNet101, and ResNet152.

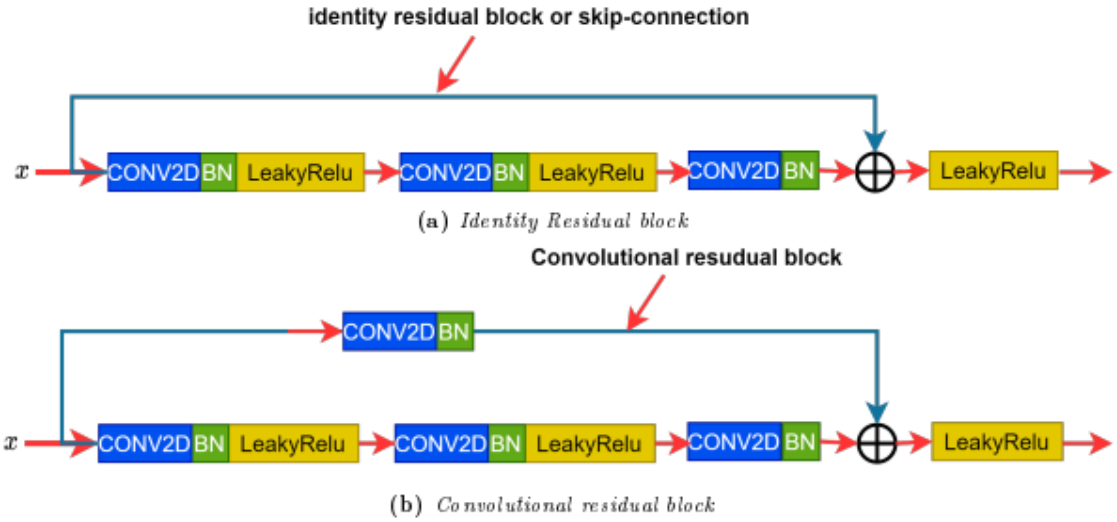

Figure 2.19: Residual blocks [34]

XCeption

Stands for extreme inception, introduced by Chollet (2017) and was motivated by the InceptionV3 architecture. The main benefit of this network over other deep networks is its computational efficiency due to its deep architecture and limited number of parameters [19]. It introduces the concept of Dephtwise Convolution in fig 2.20 of depthwise separable convolution for the inception module, obtaining results in image classification better to VGG, Resnet and Inception V3 [26]

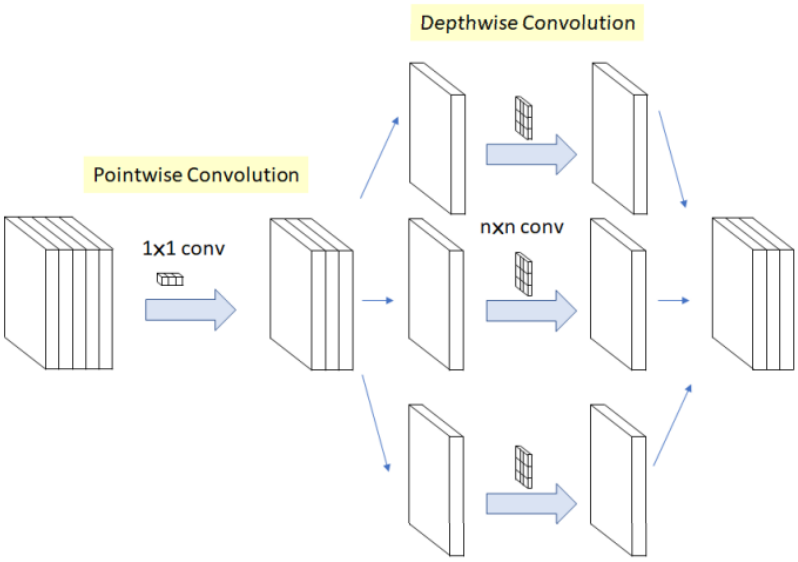

Figure 2.20: Depthwise convolution-Xception [26]

2.2.5.6 CNNs and Transfer Learning

Transfer learning is useful for low-volume datasets. A big dataset is utilized to train a neural network from scratch. The final fully connected layer is updated for the new dataset. The last layer's dimension is determined by the number of classes, which can vary between classification problems. The neural network is now trained on a smaller dataset using pre-trained weights. The literature distinguishes between two techniques: transfer learning and fine tuning. Transfer Learning involves retraining only the last layer. This is useful for tiny datasets and ensures faster training time [29]. One of the main reasons for using pre-trained models in deep learning is To reduce the need for labeled data.

Chapter 3

Experimentation, results & discussion

3.1 Introduction

In this chapter, we discuss matters related to the methodological approaches we adopted in regard to the application of deep learning (DL) in breast cancer (BC) detection. Regarding the scope of this work, we outline the research setting, the methods and procedures used for implementation, as well as the methods for analysis and evaluation which comprise the basis of our study. Finaly, a critical comparative analysis positions our findings against recent advancements (2022-2024), highlighting methodological improvements and performance gains. This structured approach ensures reproducibility while clarifying our contribution to AI-driven oncology diagnostics.

3.2 Proposed architectures :

3.2.1 The First Architecture : PCA-ANN

We propose a simple yet effective ANN architecture for BC classification using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. To enhance feature representation, we apply Principal Component Analysis (PCA) for dimensionality reduction before training the model.

The network architecture (illustrated in Figure 3.1) consists of:

1. An input layer receiving 15 principal components (PCA-transformed features) with 128 neurons
2. Two hidden layers utilizing ReLU activation functions (128, 64 neurons respectively)
3. A single-neuron output layer with sigmoid activation for binary classification (malignant/benign)

Key training parameters include:

- Binary cross-entropy loss function for optimization
- Adam optimizer for efficient parameter updates
- ReLU activation in hidden layers to address vanishing gradient concerns

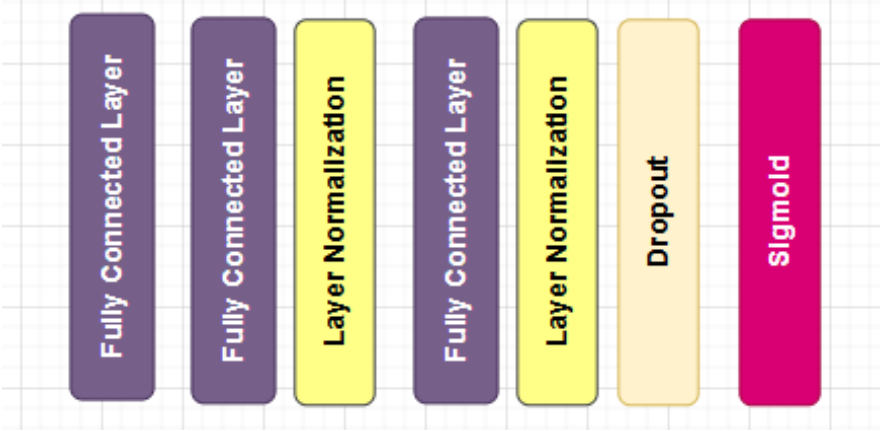

Figure 3.1: The proposed ANN architecture for WDBC dataset

3.2.2 The Second Architecture : A shallow CNN

To address the limitations of existing customized Convolutional Neural Network models, we introduce a novel architecture that achieves better accuracy while remaining efficient to run. The model was developed and trained using preprocessed images from the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) dataset, with careful optimization of hyperparameters including learning rates and optimizer selection.

As illustrated in Figure 3.2, the architecture comprises:

1. **Feature Extraction Module:**
    - Four convolutional layers for feature extraction (128, 128, 32, 32 filters respectively)
    - Batch normalization for stable training
    - Average pooling layers for progressive dimensionality reduction
2. **Classification Module:**
    - A fully connected layer integrating extracted features (32 neurons)
    - Dropout layers (rate of 50%), L2 regularization (Lambda=0.0001) and early stopping (Patience=5) to prevent overfitting
    - Softmax activation in the final layer for probabilistic classification

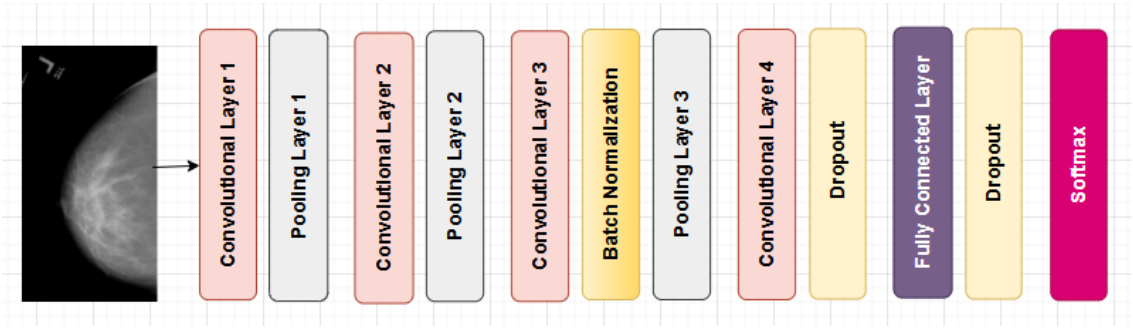

Figure 3.2: proposed CNN architecture for CBIS-DDSM dataset

3.3 Experimental Environment

3.3.1 Google colaboratory

Or Google Colab, is a free cloud-based platform that enables users to develop and execute Python code collaboratively in a Jupyter Notebook environment. The Google Collaboration Notebook is designed to simplify machine learning and data science tasks by providing a virtual environment, with free GPU resources [5].

3.3.2 Kaggle

Kaggle is an online platform for data science and machine learning, It provides a huge archive of publicly available datasets covering numerous fields, allowing users to practice their abilities with real-world data and also provides a free, cloud-based environment (Notebooks/kernels) for users to build, execute, and publish their data science programs and gives access to powerful computational resources (such as GPUs and TPUs).

3.4 Programming Language

3.4.1 Python

Python is a popular high-level programming language in DL/ML because of its readability, simplicity, and rich library ecosystem (which includes scikit-learn, TensorFlow, and PyTorch) that makes data processing, model building, training, and deployment easier.

3.5 Implementation Tools

Currently, various libraries, such as Tensorflow , allow you to build strong models in less time than ever before. Several tools were used in this thesis which helped implementing work

3.5.1 TensorFlow

TensorFlow is an end-to-end open source platform for ML. It offers a comprehensive and flexible ecosystem of tools, libraries, and community resources that enable researchers to advance the field of machine learning and developers to easily create and deploy applications [9]

3.5.2 Keras

Keras is a Python-based deep learning application programming interface running on top of TensorFlow. It was created with the intention of facilitating rapid experimentation [9]

3.5.3 Matplotlib

This package is the most widely used Python module for creating 2D plots and other data visualizations. Since visualization tools are necessary for data analysis, this library is the most appropriate for this use case [9]

3.5.4 OpenCV

Open Source Computer Vision Library): OpenCV is an open-source library in Python, used for computer vision and image processing tasks, such as object detection, face recognition, image filtering, and video analysis. It provides tools to work with images and videos efficiently in ML/DL applications.

3.5.5 Numpy

The foundation of numerous other Python libraries that have grown out of it is this library, which translates to "numerical Python." Since NumPy offers data structures and high-performance functions that the Python basic package is unable to offer, it is, in fact, the cornerstone library for scientific computing in Python [9]

3.5.6 Pandas

Pandas is an open-source Python library used for data manipulation and analysis. It provides powerful data structures like DataFrames and Series to handle structured data, making it easy to clean, filter, reshape, and analyze datasets in ML/DL workflows.

3.6 Data sets Description

3.6.1 The WDBC dataset :

Cancer Wisconsin (Diagnostic) Data Set, taken from kaggle [2] is a classic and widely used data set in the machine learning (ML) community, particularly for binary classification tasks (malignant vs. benign).

It consists of 569 instances (357 begnin and 212 malignant cases), each representing a breast mass analyzed through digitized images of fine needle aspirates (FNA). Containing 30 real-valued features that quantify various characteristics of the cell nuclei present in the images. These features include measurements such as radius, texture, perimeter, area, smoothness, compactness, concavity, and symmetry, ect [6]

3.6.2 The CBIS-DDSM dataset :

The CBIS-DDSM [3], is an updated version of the Digital Imaging and Communications in Medicine (DDSM).It contains over 3,000 mammography cases, each with standard views: Cranio-Caudal (CC) and Mediolateral Oblique (MLO) for both breasts. The dataset includes images along with lesion type (mass or calcification), and pathology (benign or malignant)

In this project, a subset with 3286 samples is selected to evaluate the classification model. Specifically, the subset contains 1931 benign samples and 1355 malignant samples.

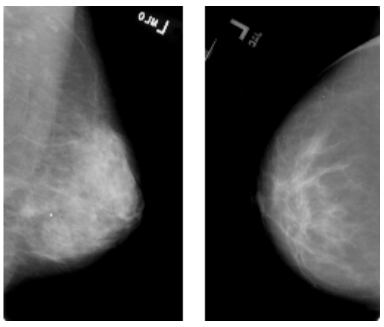

Figure 3.3: Full mammography images from CBIS-DDSM dataset (Kaggle)

3.7 Data Preprocessing

Data preprocessing is the essential first step in ML task, it ensures that data is suitable for model training.

3.7.1 Preparing WDBC dataset :

Several techniques were applied to the WDBC dataset, table 3.1 below depicts them with their purpose

Technique	Purpose
Dropping columns	Irrelevant to target/Columns with too many missing values
Standardization	standardizes features (PCA sensitive to feature scale)
Label Encoding	convert categories into numbers
Principal Component Analysis	For numeric features, to reduce redundancy
Splitting	Dividing data into train/test sets

Table 3.1: Pre-processing techniques used

3.7.1.1 Feature extraction with PCA :

Feature extraction reduces dimensionality to avoid redundancy and correlation among features, Although 30 features are not huge, some may be redundant (or correlated). This prevents overfitting and speeds up training.

Dealing with so many attributes becomes difficult because it is difficult to visualize data in 30 different dimensions.The aim in this step is to reduce the attributes from 30 to 15 without losing the key components of the data. PCA combines correlated features into fewer principal components, capturing the most variance while filtering out noise from less significant components. Different numbers of components ranging from 10 to 25 were used to find the best number with the best performance. Fig 3.4 shows the separation of the data on PC1 and PC2, we can see the malignant and benign class linearly separated.

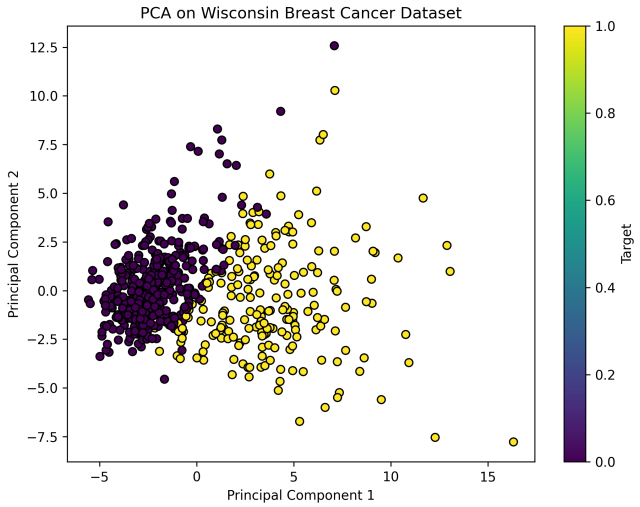

Figure 3.4: PCA with 15 component from wisconsin dataset

3.7.2 Preparing CBIS-DDSM dataset

This part explains how the images were prepared and cleaned before being used to train the model. The table 3.2 below shows the preprocessing techniques applied. 3-fold stratified cross-validation was used to split the training and validation sets.

Technique	Purpose
Contrast	Enhaces/Reduces differences in intensity
Brightness	Simulates lighting variations
Flipping	Increases spatial diversity
Resizing	Matches input size required by the model
Normalization	Speeds up training/improves convergence
Encoding	converting categorical into numerical
Class mapping	transforming class labels to numerical
Splitting	Dividing data into sets

Table 3.2: Preprocessing techniques used

3.8 Evaluation Measures

3.8.1 Confusion Matrix

Confusion matrix is basically a tabular summary showing how well the model performs [30], we compare its predictions to the actual results as shown in 3.5

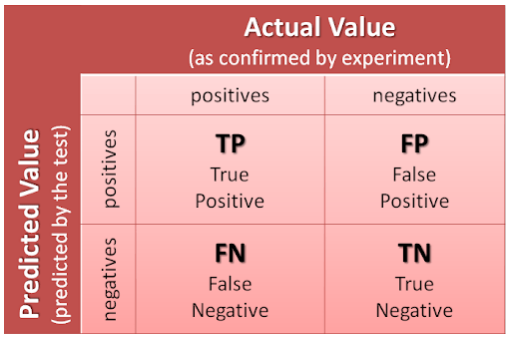

Figure 3.5: Confusion Matrix [30]

1. **True Positive (TP):** Correctly classified cases where the prediction is positive compared to the true values.
2. **True Negative (TN):** Situations where the prediction is negative compared to the actual value.
3. **False Positive (FP):** Situations in which the prediction is positive compared to the true value.
4. **False Negative (FN):** indicates the number of malignant cases incorrectly labeled as begnin.

Confusion matrix also helps calculate key measures like accuracy, precision, and recall, which give a better idea of performance:

3.8.2 Accuracy

It is the percentage of correct predictions among all predictions.

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

3.8.3 Precision

how many correct ones (positive predictive value) among all positive predictions.

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

3.8.4 Recall (Sensitivity)

It is the percentage of correct predictions of actual positives (true positive rate). This means it looks at true positives and false negatives (which are positives that have been incorrectly predicted as negative).

$$Recall = \frac{TP}{TP + FN}$$

3.8.5 F1-score

It is the balance between precision and recall.

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

3.8.6 AUC (Area Under ROC Curve)

As one of the most widely used ranking-type metrics, AUC can be utilized to develop an optimal learning model, for evaluating the performance of a classifier and selecting an optimal solution during classification training, the AUC is superior to the accuracy metric . This criterion can be calculated using Equation [16].

$$AUC = \frac{Specifity + Sentivity}{2}$$

3.9 Training & results

3.9.1 The PCA-ANN :

The model was trained using RELU activation function. After experiments, we come to the parameters configuration in table 3.3.

Optimizer	Learning rate	batch size	epochs
Adam	0.0001	8	35

Table 3.3: final configuration of ANN model

The Curves 3.6 & 3.7 represents the accuracy and the loss function for our first model after the experimentation with and without applying PCA.

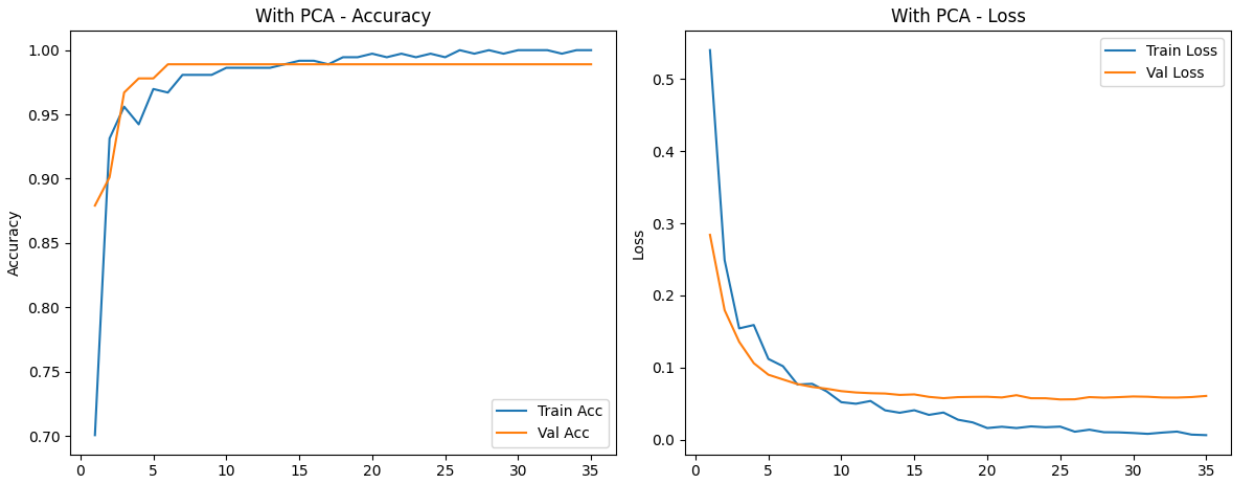

Figure 3.6: Validation results through The accuracy and the loss function ANN with PCA

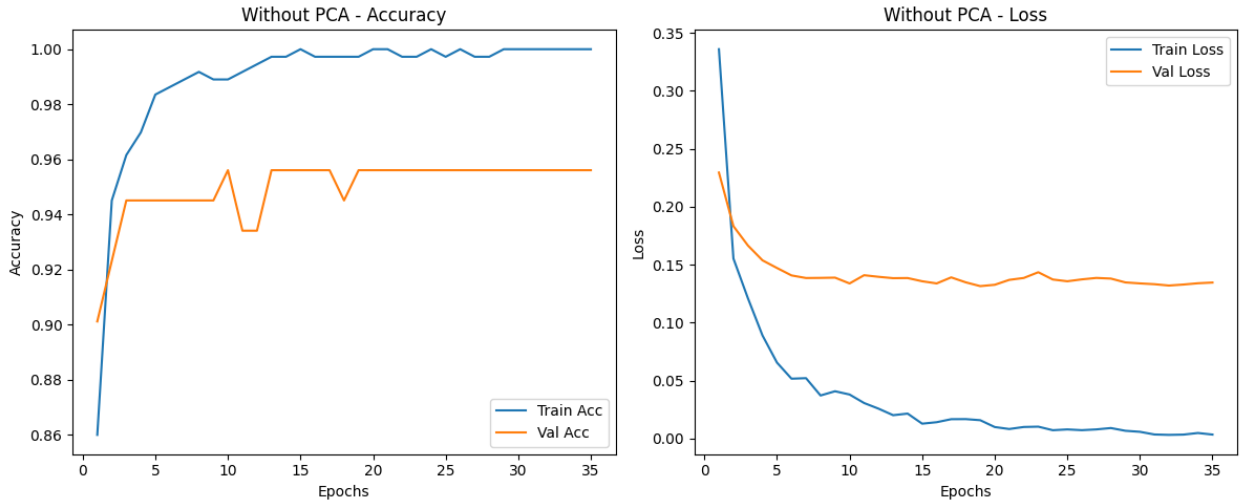

Figure 3.7: Validation results through The accuracy and the loss function ANN without PCA

- The final results are provided in table 3.4

Accuracy	Precision	Recall	F1-score	AUC	loss
100%	100%	100%	100%	100%	0.01

Table 3.4: The accuracy,precision,AUC and loss after the experimentation

- The Area Under ROC Curve in fig 3.8

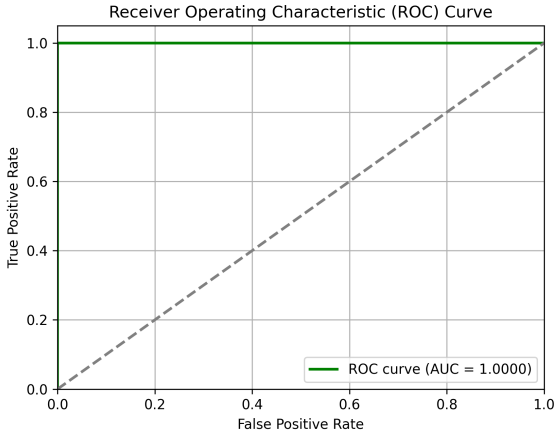

Figure 3.8: Area Under Curve plot for first model

- Confusion matrix : 114 instance was left for final test to evaluate the performance of the model with PCA in 3.9, confusion matrix below shows performance with 66 TPs, 48 TNs, 0 FPs, 0 FNs. Obtained results from metrics are depicted in table 3.6.

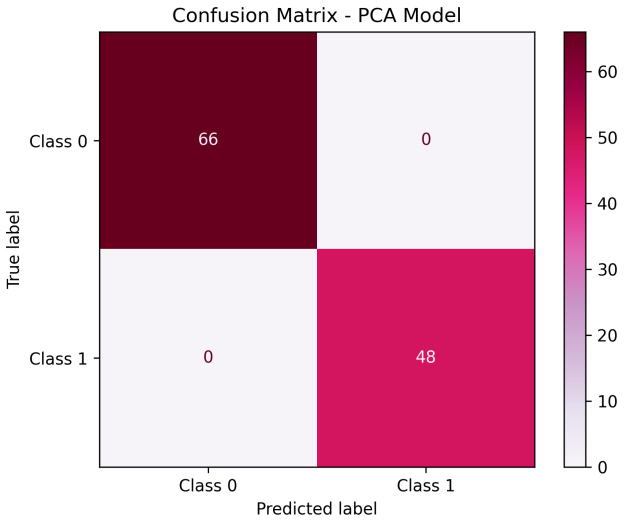

Figure 3.9: ANN model's confusion matrix with PCA

3.9.2 The shallow CNN :

The model was trained using GELU activation function, with different Optimizers and learning rates. Stratified K-fold technique was used for better generalizability with early stopping, to prevent overfitting with patience of 5.

- Optimizers : In the following figure 3.10, we have an experiment with several types of optimizers SGD, Adam, Adamax

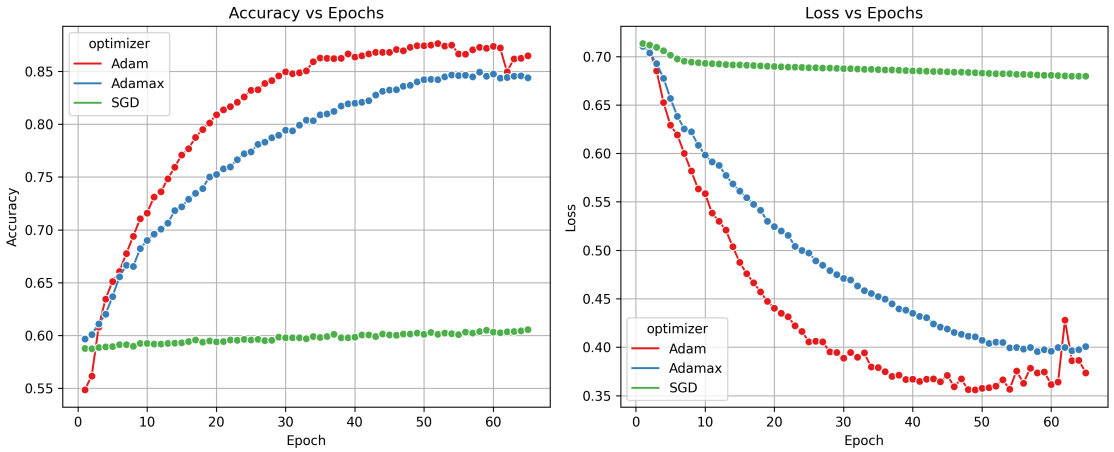

Figure 3.10: Accuracy and loss over epochs with different optimizers

- Learning-Rate : we have experimented with different values of learning rate also (1e-4, 2e-4, 3e-4, 4e-4, 5e-4).

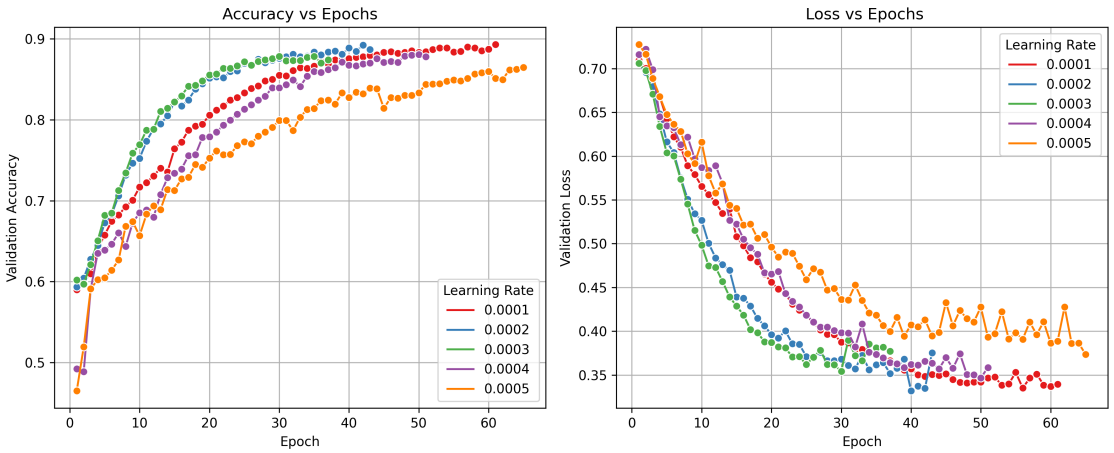

Figure 3.11: Accuracy and loss with different learning rates in each fold

- Final configuration results : The figure 3.12 represents the accuracy and loss curves for our second model after the experimentation over 3 folds, configuration parameters below in table 3.5

Optimizer	Learning rate	Batch size
Adam	0.0001	64

Table 3.5: Parameters of best configuration

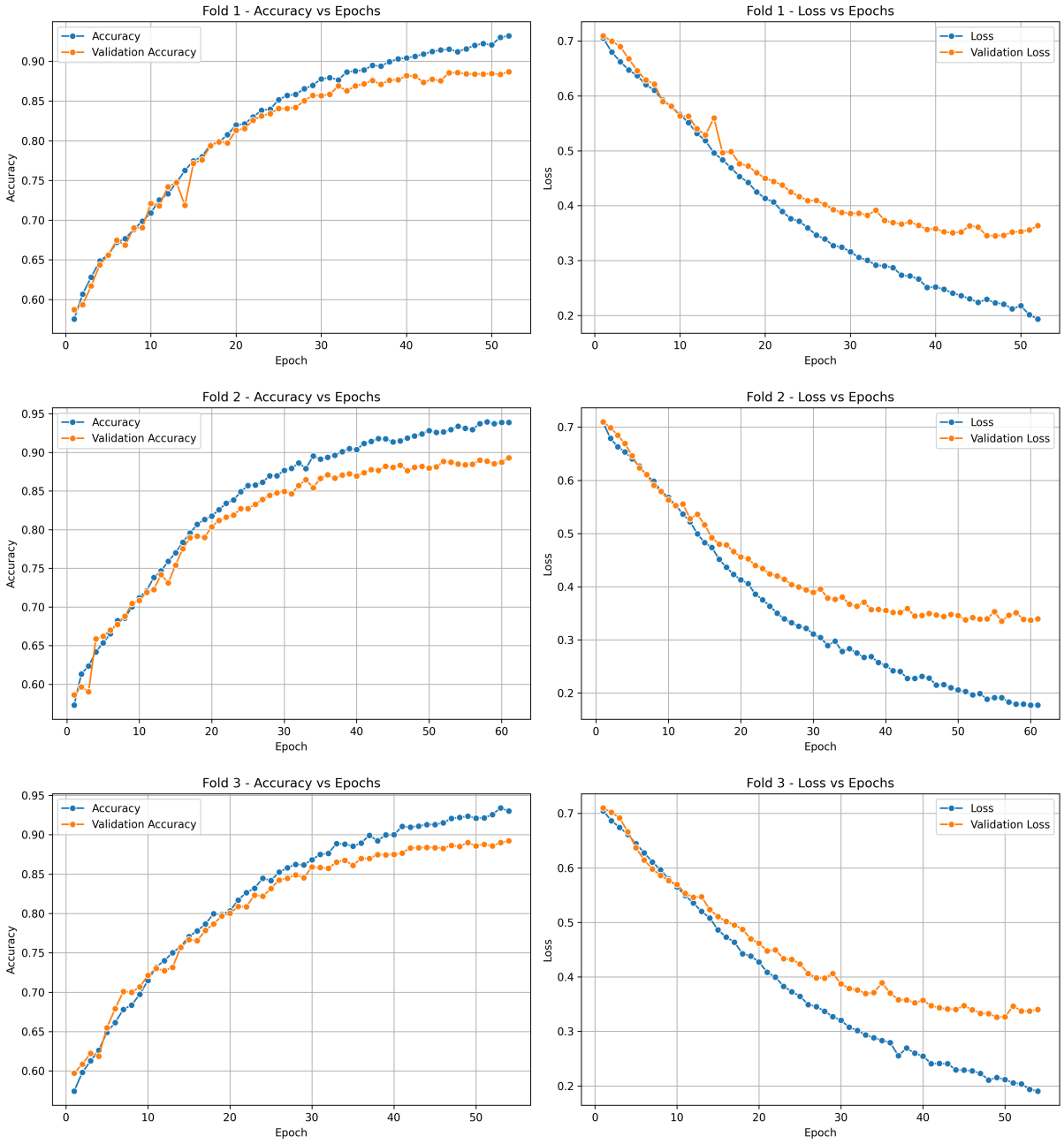

Figure 3.12: Accuracy and loss over epochs of the three folds

- The AUC-ROC is a reliable statistic that measures the model's ability to identify between benign and malignant instances at various thresholds, as illustrated in fig 3.13 The findings give detailed figure of the model's performance, with a higher AUC-ROC.

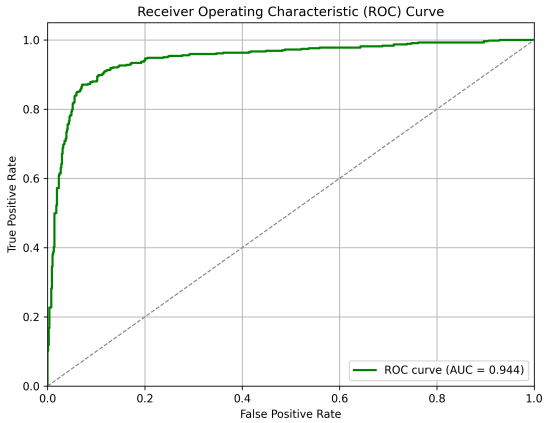

Figure 3.13: Area Under Curve plot of CNN-model

- Confusion matrix : 1315 images were left for final test to evaluate model, confusion matrix below shows performance with 718 TPs, 469 TNs, 55 FPs, 73 FNs. Obtained results from metrics are depicted in table 3.6

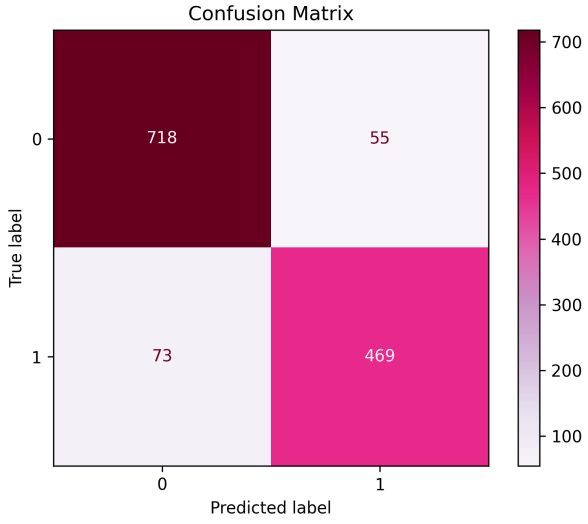

Figure 3.14: Confusion matrix

Accuracy	Precision	Recall	F1-measure	AUC	loss
90%	90.2%	90.3%	90.2%	94.4%	0.32

Table 3.6: Evaluation metrics performance of second model

3.10 Evaluation and discussion

For a robust evaluation of our models, their efficacy will be assessed through a comparative analysis with previously published studies involving the same subject and utilizing the same datasets. Tables 3.7 and 3.8 below contain overall evaluation with different metrics for each.

- **The PCA-ANN model comparision :**

  H.Alshayeji (2022), used a shallow artificial neural network model with one hidden layer using the Wisconsin breast cancer dataset and the Wisconsin diagnostic breast cancer dataset without employing feature optimization or selection algorithms, in the task of classifying benign and malignant tumours. The datasets are divided into 80% for training and 20% for testing using five-fold cross-validation. His model demonstrated significant potential for diagnosing BC using WBCD and WDBC without the need for feature optimization or selection algorithms.

  A. Bokhare (2023), presented the classification of same data set using various ML models. The models considered for the study were k-nearest neighbor (kNN), decision tree classifier, support vector machine (SVM), random forest (RF), SVM kernels, logistic regression, Naïve Bayes. These classifiers were tested, analyzed and compared with each other. The classifier, decision tree, got the highest accuracy with 97.08% among all other ones.

  E. Poornima (2023), implemented a hybrid algorithm with PCA and SVM and optimized SVM with k-fold cross-validation for predicting BC using same data set. The model with SVM and k-folds cross-validation performed better than classic SVM.Same accuracy was achieved before and after combining PCA with SVM. The hybrid algorithm outperforms traditional SVM in recall, precision, accuracy, and f1-score.

  As table 3.7 clearly demonstrates, our PCA-ANN model establishes a new benchmark for breast cancer classification on the WDBC dataset, outperforming existing approaches across all key metrics:

  - Accuracy: Highest correct classification rate
  - Precision: Fewest false positives
  - Recall: Best at identifying true cases
  - F1-score: Optimal balance between precision and recall

Clinical Significance:

1. The improved precision reduces unnecessary patient anxiety from false alarms
2. The model achieves this high performance using only 15 principal components, making it computationally efficient for clinical deployment.

This study has certain limitations that should be acknowledged, including:

- The model achieved 100% accuracy; however, this result is likely influenced by the small size of the dataset
- PCA transforms features into principal components, making it more difficult to interpret the clinical significance of inputs

Model	**Accuracy**	**Precision**	**Recall**	**f1-measure**
Hybrid SVM-PCA (2023) [28]	96.5%	90%	90%	90%
Classic ML algorithms (2023) [10]	97.08%	97.28%	97.45%	77.52%
ANN-WDBC (2022) [6]	99.47%	98.71%	99.59%	99.13
Propsed PCA-ANN (2025)	**100%**	**100%**	**100%**	**100%**

Table 3.7: Comparison of various machine learning models for the WDBC dataset

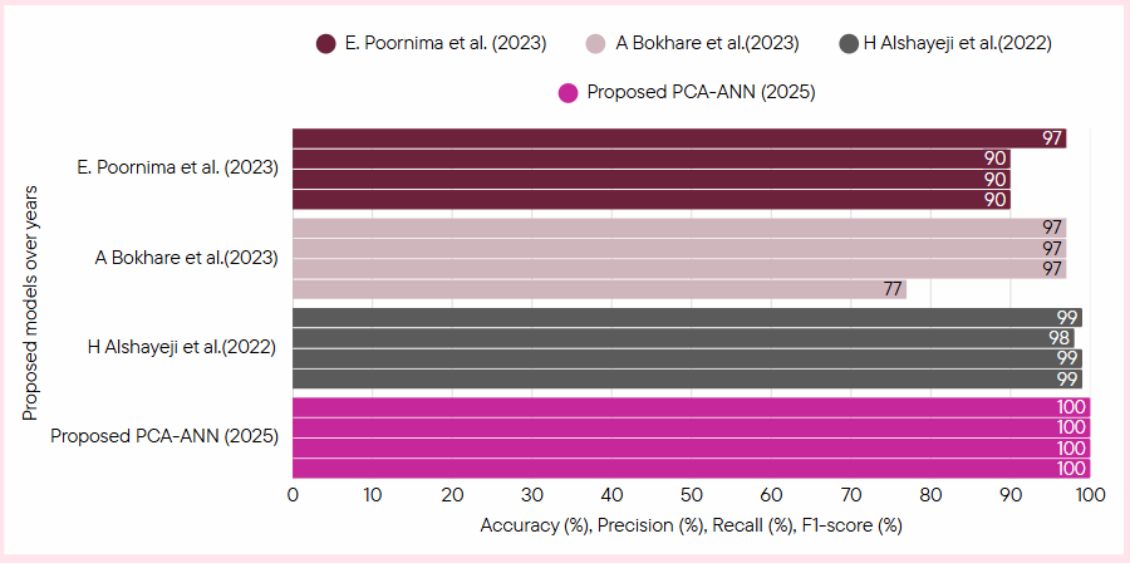

Figure 3.15: Graphic illustrating first model comparision with litterature

- **The shallow CNN model comparision :**

M Yusof et al.(2024), developped and compared the accuracy of classification BC using ResNet50 V2, VGG19 and Convolutional Neural Networks (CNN), on the CBIS-DDSM dataset. They used image enhacement (CLAHE) technique to improve image contrast,for highlighting important details in specific areas of interest. They tried multiple combinations of parameters to get the best performance (batch sizes, learning rates and epochs). The data showed that model performance differs with the number of epochs, each model achieved better accuracy with certain number of epochs. The results showed that VGG model was the best performing one.

H Shekhar (2023), compared the performance of proposed shallow CNN architecture against pre-trained deep CNN architectures trained on mammography images (CBIS-DDSM and INBreast datasets). Mammogram images are pre-processed, and splitted into train/test sets. The paper consists of two different approaches. A very small CNN with only two convolutional layers has been used for simulation in the first part of the first approach. A dropout layer has been added in the second part to mitigate hard overfitting, and in the third part, data augmentation, another regularization technique, is applied to the previous model to further prevent overfitting. The second method involves testing a variety of pre-trained CNN networks using a fine-tuning strategy, including VGG19, ResNet50, MobileNet-v2, Inception-v3, Xception, and Inception-ResNet-v2 respectively. The finetuned Xception model achieved the best results, especially with INBreast dataset.

C Murthy et al.(2024), designed deep learning architecture combining Convolutional Neural Networks (CNNs) for image processing with Stochastic Gradient Descent (SGD) optimization techniques for better classification, their objective was to improve the diagnosis of BC by combining datasets: the WDBC and the CBIS-DDSM data sets. Both data sets were integrated and pre-processed with different techniques, including feature extraction with PCA and feature selection. Prepared data was fed to CNN architecture which contains 2 convoultional layers, 2 average pooling layers and 3 dense layers, trained with RELU activation function and softmax for last classification. Cross validation was used, and model was trained on WDBC dataset.

As demonstrated in table 3.8, our proposed model outperforms first and second existing studies across all key metrics, including accuracy, precision, and recall While our results were surpassed by the hybrid CNN-SGD model.

The CNN-SGD model was specifically developed for multimodal detection, which represents its primary advantage. Our implementation faced certain limitations:

1. Time constraints prevented the development of a comparable multimodal system
2. Multimodal integration remains a key objective for future research

Authors	Model	Accuracy	Precision	Recall
H Shekhar et al. (2023)	First shallow CNN	77.4%	78.8%	77.8%
	Second shallow CNN	80.4%	82.3%	79.8%
	Third shallow CNN	79.0%	80.7%	78.8%
	VGG19	77.9%	79.9%	77.4%
	ResNet50	83.2%	84.6%	83.2%
	Inception v3	87.6%	89.5%	86.5%
	Xception	89.2%	91.3%	87.9%
	Inception-ResNet v2	85.7%	87.8%	84.5%
M Yusof et al. (2024)	CNN	64.0%	–	–
	VGG19	72.0%	72.0%	–
	ResNet50	70.0%	69.0%	–
Proposed shallow-CNN (2025)	**Customized CNN architecture**	**90.0%**	**90.3%**	**90.3%**
C Murthy et al. (2024)	Hybrid CNN-SGD approach	96.0%	95.0%	95.0%

Table 3.8: Comparison of various CNN-based approaches for CBIS-DDSM

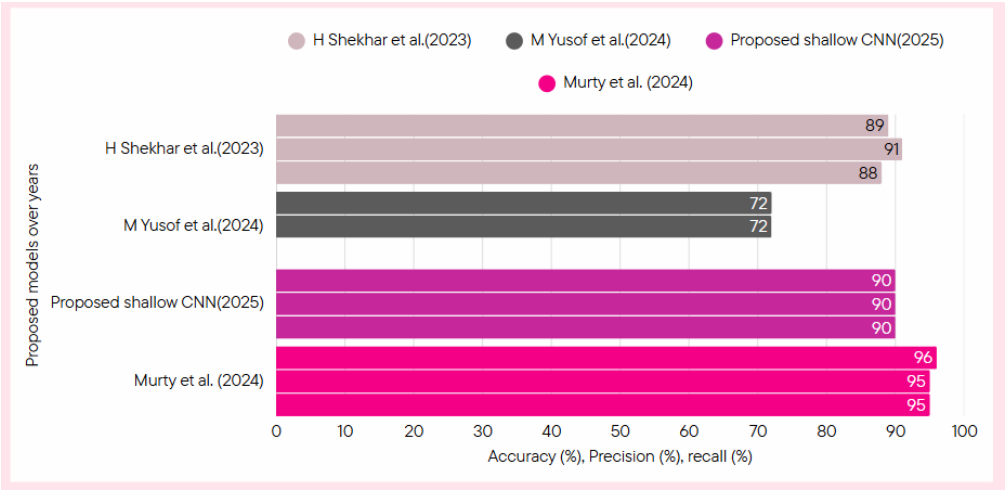

Figure 3.16: Graphic illustrating second model comparision with litterature

3.11 Conclusion

In this chapter, two models were developed for the early detection of BC by classifying tumors as either benign or malignant. Various datasets were utilized, each with its appropriate pre-processing techniques. Different network architectures were designed and trained using different activation functions, loss functions, optimizers, learning rates, and batch sizes. Results of the proposed models were obtained and the performance was then evaluated and compared with existing studies to assess their effectiveness and accuracy. The comparative analysis demonstrates our models diagnostic accuracy improvements over existing solutions while identifying opportunities for future enhancement through multimodal integration.

General Conclusion

The primary objective of this thesis was to support medical professionals in the early detection of breast cancer and to integrate artificial intelligence in the fields of medicine. To this end, we thoroughly investigated two distinct modeling approaches:

- **a classical ANN enhanced with PCA** for tabular data: the key aspect of our approach involves integrating Principal Component Analysis (PCA) for robust feature extraction, which enhances the model's ability to focus on the most relevant information within the dataset, thereby improving overall efficiency and accuracy.
- and **a shallow CNN**, specifically tailored for image-based classification tasks.

Both models were selected for their inherent strengths and proven efficacy. Dealing with different types of data sets allowed us to learn more about preprocessing and handling data.

We investigated several methods of classifying breast cancers, with special attention to building tailored architectures to improve our knowledge of deep learning. Furthermore used to enhance model performance and generalization were data augmentation and other pre- processing methods; stratified cross-valuation strategy guarantees model robustness.

The experimental results definitively validate the effectiveness of both proposed methods. The ANN-PCA model demonstrated its strength in handling structured clinical data by efficiently reducing dimensionality and extracting salient features, leading to high classification accuracy. Simultaneously, the CNN model showcased its powerful ability to automatically learn hierarchical features from imaging data, a crucial aspect for visual diagnostics. While this study primarily focused on individual data modalities (tabular for ANN-PCA and imaging for CNN), the successful outcomes of each model provide a strong foundation for future advancements.

The generalizability of our findings is limited by the relatively small size of the datasets, which may not fully represent the diversity of breast cancer cases in the broader population. This highlights the need for further validation on other datasets to confirm its generalizability.

Future work : Based on the findings of this thesis, several new research directions have emerged that could be explored to further enhance the results achieved. These are outlined below :

1. Development of a thorough multimodal detection method comes first immediately.

2. We will give top priority to verifying our models using bigger, more diversified, and more external datasets to guarantee their resilience and generalizability over more patient populations and other clinical environments.

3. Developing and merging segmentation methods with our CNN model to exactly pinpoint tumors inside images.

4. Development of an understandable user interface for our models is part of our clinical implementation strategy. The ultimate goal is to create a practical, real-world tool that facilitates and streamlines the diagnostic process for medical professionals.

Bibliography

[1] Algeria - global cancer observatory. https://www.google.com/url?sa=iurl=https Accessed: 2025/04/25.

[2] Breast cancer wisconsin (diagnostic) data set. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data. Accessed: 2025/03/15.

[3] Cbis-ddsm: Breast cancer image dataset. https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset/data. Accessed: 2025/04/15.

[4] Dr.eric sebban,les stades et types de cancer du sein. https://www.docteur-eric-sebban.fr/cancer-du-sein/diagnostic-cancer-sein/stades-et-types-de-cancer-du-sein/. Accessed: 2025/05/09.

[5] How to use google colab. https://www.geeksforgeeks.org/how-to-use-google-colab/. Accessed: 2025/05/22.

[6] M. H. Alshayeji, H. Ellethy, S. Abed, and R. Gupta. Computer-aided detection of breast cancer on the wisconsin dataset: An artificial neural networks approach. *Biomedical signal processing and control*, 71:103141, 2022.

[7] R. AOUASSA. Cnn based deep face recognition. 2020.

[8] I. Aziz. Deep learning: An overview of convolutional neural network (cnn). *Faculty of Information Technology and Communication Sciences M. Sc. Thesis*, 2020.

[9] R. BENNOUH and A. Oussama. *A healthcare system using deep learning*. PhD thesis, 2022.

[10] A. Bokhare and P. Jha. Machine learning models applied in analyzing breast cancer classification accuracy. *Int. J. Artif. Intell*, 2252:1371, 2023.

[11] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263, 2024.

[12] E. Calli. *Faster Convolutional Neural Networks*. PhD thesis, Radboud University Nijmegen, The Netherlands, 2017.

[13] D. Caschili. *Optimization of cnn-based object detection algorithms for embedded systems*. PhD thesis, Politecnico di Torino, 2019.

[14] H. S. Das, A. Das, A. Neog, S. Mallik, K. Bora, and Z. Zhao. Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach. *Frontiers in Genetics*, 13:1097207, 2023.

[15] S. Doyle-Lindrud. State of ehealth in cancer care: review of the benefits and limitations of ehealth tools. *Number 3/June 2020*, 24(3):10–15, 2020.

[16] M. Ghorbian and S. Ghorbian. Usefulness of machine learning and deep learning approaches in screening and early detection of breast cancer. *Heliyon*, 9(12), 2023.

[17] S. HAMMOUDI and M. KORCHI. Memoire de fin d'etudes de master en electronique.

[18] M. Kaddes, Y. M. Ayid, A. M. Elshewey, and Y. Fouad. Breast cancer classification based on hybrid cnn with lstm model. *Scientific Reports*, 15(1):4409, 2025.

[19] I. H. A. Kandel. *Deep Learning Techniques for Medical Image Classification*. PhD thesis, Universidade NOVA de Lisboa (Portugal), 2021.

[20] A. La Moglia and K. M. Almustafa. Breast cancer prediction using machine learning classification algorithms. *Intelligence-Based Medicine*, 11:100193, 2025.

[21] B. Mahesh et al. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386, 2020.

[22] B. M. I. D. Mak. Digital solutions for early breast cancer detection. *Lancet Oncol*, 24:936–44, 2023.

[23] A. Mitsuk. Breast cancer information for young women: a project for terveysnetti. 2016.

[24] P. S. C. Murty, C. Anuradha, P. A. Naidu, D. Mandru, M. Ashok, A. Atheeswaran, N. Rajeswaran, and V. Saravanan. Integrative hybrid deep learning for enhanced breast cancer diagnosis: leveraging the wisconsin breast cancer database and the cbis-ddsm dataset. *Scientific Reports*, 14(1):26287, 2024.

[25] A. Patel. Benign vs malignant tumors. *JAMA oncology*, 6(9):1488–1488, 2020.

[26] D. Peña Moliner. Extending object classificatiion convolutional neural networks to custom logo detection. Master's thesis, Universitat Politècnica de Catalunya, 2020.

[27] J.-R. Pérez-Núñez, C. Rodríguez, L.-J. Vásquez-Serpa, and C. Navarro. The challenge of deep learning for the prevention and automatic diagnosis of breast cancer: A systematic review. *Diagnostics*, 14(24):2896, 2024.

[28] E. Poornima, P. Mamatha, M. Krishnamai, C. Sruthi, S. K. Shah, and L. R. Kumar. Automated early phase breast cancer detection using hybrid machine learning algorithms. In *E3S Web of Conferences*, volume 430, page 01035. EDP Sciences, 2023.

[29] C. Quandel. *Deep Convolution Neural Networks for the Analysis of a few Medical Images*. PhD thesis, 2019.

[30] C. Room. Confusion matrix. *Mach. Learn*, 6:27, 2019.

[31] C. H. Salh. *Breast Cancer Diagnosis Using Deep Learning Models*. PhD thesis, Polytechnic University, 2023.

[32] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. Sharma. Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research*, 1(2):109–126, 2010.

[33] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. Sharma. Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research*, 1(2):109–126, 2010.

[34] S. N. Tesema. *Deep Convolutional Neural Network Based Object Detection Inference Acceleration Using FPGA*. PhD thesis, Université Bourgogne Franche-Comté, 2022.

[35] K. P. Trayes and S. E. Cokenakes. Breast cancer treatment. *American family physician*, 104(2):171–178, 2021.

[36] J. van den Broek. Public health benefits and harms of breast cancer screening: Microsimulations informing screening recommendations. 2019.

[37] M. J. Walker. *Breast cancer screening behaviours and outcomes in women with a family history of breast and/or ovarian cancer in Ontario*. University of Toronto (Canada), 2014.

[38] S. Yin. Addressing data imbalance in breast cancer prediction using supervised machine learning. Master's thesis, Purdue University, 2022.

[39] M. F. D. B. M. Yusof and N. H. Zakaria. Breast cancer classification using deep learning.