REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

CANCER DETECTION IN FINE NEEDLE ASPIRATION SAMPLES USING BACK PROPAGATIONAL ARTIFICIAL NEURAL NETWORK

Zaid Abbas Kurdi SARRAY

Master's Thesis

Supervisor

Asst. Prof. Dr. Zaid HAMODAT

Istanbul, 2022

CANCER DETECTION IN FINE NEEDLE ASPIRATION SAMPLES USING BACK PROPAGATIONAL ARTIFICIAL NEURAL NETWORK

Zaid Abbas Kurdi SARRAY

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled CANCER DETECTION IN FINE NEEDLE ASPIRATION SAMPLES USING BACK PROPAGATIONAL ARTIFICIAL NEURAL NETWORK prepared by ZAID ABBAS KURDI SARRAY and submitted on 14/12/2022 has been accepted unanimously for the degree of Master of Science in Electrical and Computer Engineering.

Asst. Prof. Dr. Zaid HAMODAT

the Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Zaid HAMODAT	Department of Electrical and Electronics Engineering, Altınbaş University	___________________
Asst. Prof. Dr. Mesut ÇEVİK	Department of Electrical and Electronics Engineering, Altınbaş University	___________________
Asst. Prof. Dr. Abbas UĞURENVER	Department of Electrical and Electronics Engineering, İstanbul Aydın University	___________________

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: ___/___/___

I hereby declare that all information and data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Zaid Abbas Kurdi SARRAY

Signature

DEDICATION

Thank you to my supervisor, Asst. Prof. Dr. Zaid HAMODAT for providing guidance and feedback throughout this project. Thanks also to my wife for putting up with me being sat in the office for hours on end, and for providing guidance and a sounding board when required.

ABSTRACT

CANCER DETECTION IN FINE NEEDLE ASPIRATION SAMPLES USING BACK PROPAGATIONAL ARTIFICIAL NEURAL NETWORK

Sarray, Zaid Abbas Kurdi

M.Sc., Electrical and Computer Engineering, Altınbaş University

Supervisor: Asst. Prof. Dr. Zaid HAMODAT

Date: 12/2022

Pages: 79

Breast cancer represents one of the major public health challenges, due to the fact that it is the most common cancer in women and the leading cause of cancer death worldwide. Indeed, nearly one in seven women will be affected by this pathology during her lifetime, the risk increasing with age The research carried out within the framework of this thesis concerns the study of the malignant nature of breast masses, the study of the different stages of a computer-assisted diagnosis system, the comparison of the different decision-making aid strategies in breast imaging and the possible contribution and disadvantages of deep learning. the proposal of an approach to reduce the size of a parsimonious dictionary learned on a mammographic image database for the reduction of the noise present in the images, the design of an unsupervised segmentation method that is robust, efficient, fast and requires only one parameter to be specified the development of an original technique for extracting spicules from the image the construction of two models offering diagnostic aid on both the benign/malignant character of a mass based on its shape, as well as the dense/fatty nature of the centre of the mass.

Keywords: CNN, Segmentation, ANN, Classification.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

AI	:	Artificial Intelligence
FSK	:	Frequency Shift Keying
IoT	:	Internet of Things
PAN	:	Personal Area Network
OFDM	:	Orthogonal Frequency Division Multiplexing
CAD	:	Computer Aided Design
DM	:	Digital Mammography
MRI	:	Magnetic Reasoning Images
CNN	:	Convolutional Neural Network
FS	:	Feature Selection
RGB	:	Red, Green, and Blue
HOG	:	Histogram Oriented Gradients
TDMA	:	Time Division Multiple Access

1. INTRODUCTION

1.1 BACKGROUND

Breast cancer represents one of the major public health challenges, due to the fact that it is the most common cancer in women and the leading cause of cancer death worldwide. Indeed, nearly one in seven women will be affected by this pathology during her lifetime, the risk increasing with age. In addition, studies carried out worldwide in 2012, report 522,000 deaths from breast cancer during the same calendar year, which represents an increase of 14% compared to 2008. in 2013 the changing trends in incidence and mortality linked to cancer over the period 1980-2021.

The conclusions indicate that the incidence of breast cancer, which increased significantly between 1980 and 2000, has been gradually decreasing since 2005. Indeed, the incidence rate standardized for the world population has increased by 1.4% per year on average. between 1980 and 2021, but we note a slightly greater decrease of 1.5% per year if we only take into account the period from 2005 to 2021. As for mortality, it remained practically stable until 1995 and this despite the sharp increase in incidence over the same period; before experiencing a significant drop in 2012. Indeed, we observe an average decrease in mortality which went from 0.6% per year between 1980 and 2021 to 1.5% per year between 2005 and 2021.

Conversely in certain regions of the world, the number of deaths linked to breast cancer continues to decrease, despite the appearance of 48,763 new cases in 2012. This is partly explained by the improvement in the management treatment and earlier diagnosis thanks to screening campaigns [4]:

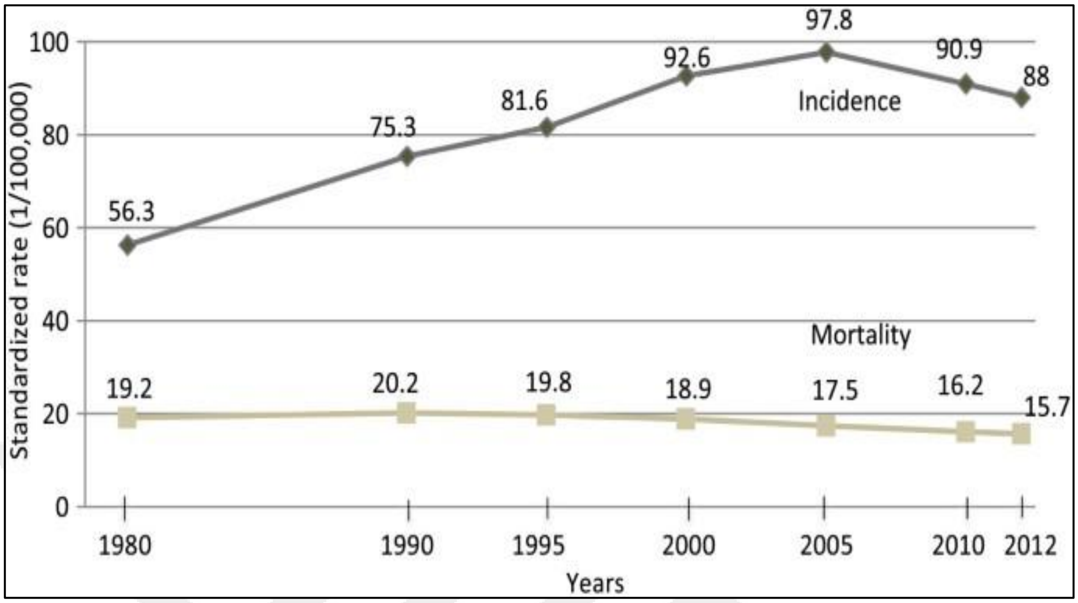


Figure 1.1: Breast cancer statistics from 1980 to 2012 [3].

With the aim of ensuring early detection of the tumor, radiologists have increased the frequency of mammograms – which is the reference examination used in the context of screening – in particular for the age group most affected, to namely 50 to 74 years which represents a considerable amount of data, making the process of image interpretation by radiologists tedious and time-consuming. In view of the increase in the number of mammograms in recent years, various research studies have been carried out either to automatically detect lesions – i.e. micro-calcifications or masses – in the images using computer-aided detection systems or to provide a second opinion about the lesion detected through computer-aided diagnosis systems.

This has led to the design of several mammographic image processing all inspired by the Breast Imaging Reporting and Data System (BI-RADS) classification of the American College of Radiology (ACR) to classify the masses into two categories, namely benign and malignant The case that concerns us in this work is the conception computer-assisted diagnostic systems and is divided into five steps which are pre-processing, segmentation, extraction of spicules and other descriptors, learning a model and classification. The difference between a detection system and a diagnostic system is that the latter integrates the extraction of descriptors and their classification in order to provide diagnostic assistance.

However, the quality and relevance of the second opinion provided by any diagnostic system strongly depends on the quality of the image, the precision of the mass segmentation and, in our case, the detection of the structures of interest, namely the spicules. This is what justifies the fact that we are interested in this thesis at all these stages. Contrary to what has always been done in the state-of-the-art, where researchers have systematically sought to classify masses as benign or malignant, this work mainly emphasizes the distinction between spiculated masses and architectural distortions, in particular because of the difficulty experienced by radiologists in differentiating them on superposition images obtained after mammography, which represents a real challenge in computer vision.

Indeed, according to BIRADS, benign masses have a round or oval shape and a circumscribed or micro lobulated outline, while malignant masses usually have a lobulated or irregular shape and an indistinct or spiculated outline. It can therefore be seen that the morphology of the masses may be sufficient to separate the malignant masses from the benign ones. This is obviously not enough to distinguish the malignant masses between them and requires the combination of several other criteria to hope to achieve this. In this work, we designed a fully automatic computer-assisted diagnostic system dedicated to both the extraction of spiculated masses and architectural distortions. To achieve this, we have proposed a method of pre-processing the images in order to remove noise and improve the contrast between the different anatomical structures; then we implemented an unsupervised mass segmentation approach that is both robust and efficient. Then, we proceeded to extract the spicules from the image through an original approach of segmentation of structures of interest.

The three proposed methods were then compared with the most effective state-of-the-art approaches in order to prove their effectiveness. The different comparison strategies implemented have made it possible to prove that our methods are both faster and more efficient than those proposed in the literature.

To our knowledge, to date, no study exists on the distinction between malignant spiculated masses and architectural distortions through the extraction of spicules; this work can therefore be seen as the first of its kind. Finally, we carried out an unsupervised extraction of the descriptors using a method derived from Deep learning, namely artificial neural networks and more precisely convolutional neural networks. The extracted descriptors are then learned to design a model which we use to provide diagnostic assistance, either on the morphology of the mass or on the dense or fatty nature of the center of the latter.

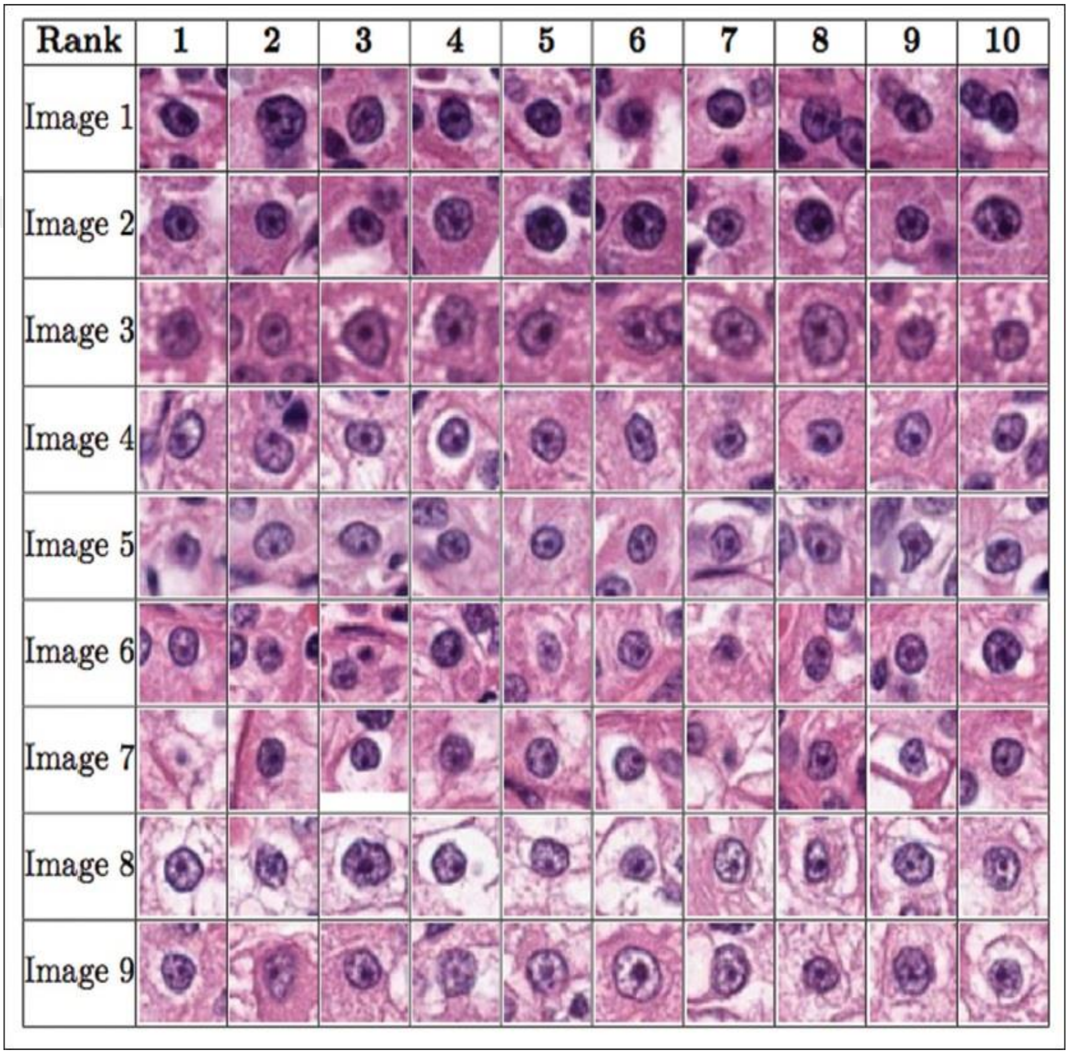


Figure 1.2: Classification of breast cancer images in the BIRADS dataset.

1.2 PLANNING THE PROBLEM

The diagnosis of breast cancer is based on self-examination, routine mammograms and biopsies of tumor masses. The introduction of self-examination and mammograms in women's health programs occurred mutually in the 1980s, dramatically increasing the diagnosis of this disease

[14]. However, the impacts of mammography screening on breast cancer prognosis and mortality are still discussed today [14] a 23-40% drop in mortality rates in participants of mammography programs, other authors believe that this decrease is due to other factors (self-exams, improvement of systemic therapies, improvement in diagnostic imaging technology) [17].

Although a consensus has not yet been reached, it is unanimous among health professionals who are vital for the positive prognosis of breast cancer that a) there is an early and accurate diagnosis; b) the tumor stage is determined [14]. Medical diagnostic aid programs (CAD) applied in medical examinations aim to improve the quality of diagnosis. These tools have shown promising results in several areas [20], and intensive research has been carried out in the area of mammograms [21]. To improve the diagnosis of mammography exams, it is essential to seek a tool that allows the early and accurate detection of anomalies.

The main objective is to reduce the rate of false negatives, preventing lesions from going unnoticed by the physician. This type of tool can perform a previous analysis of images, pointing out regions with possible anomalies for a subsequent professional evaluation, or a second opinion, confirming the diagnosis made by the doctor. [7].

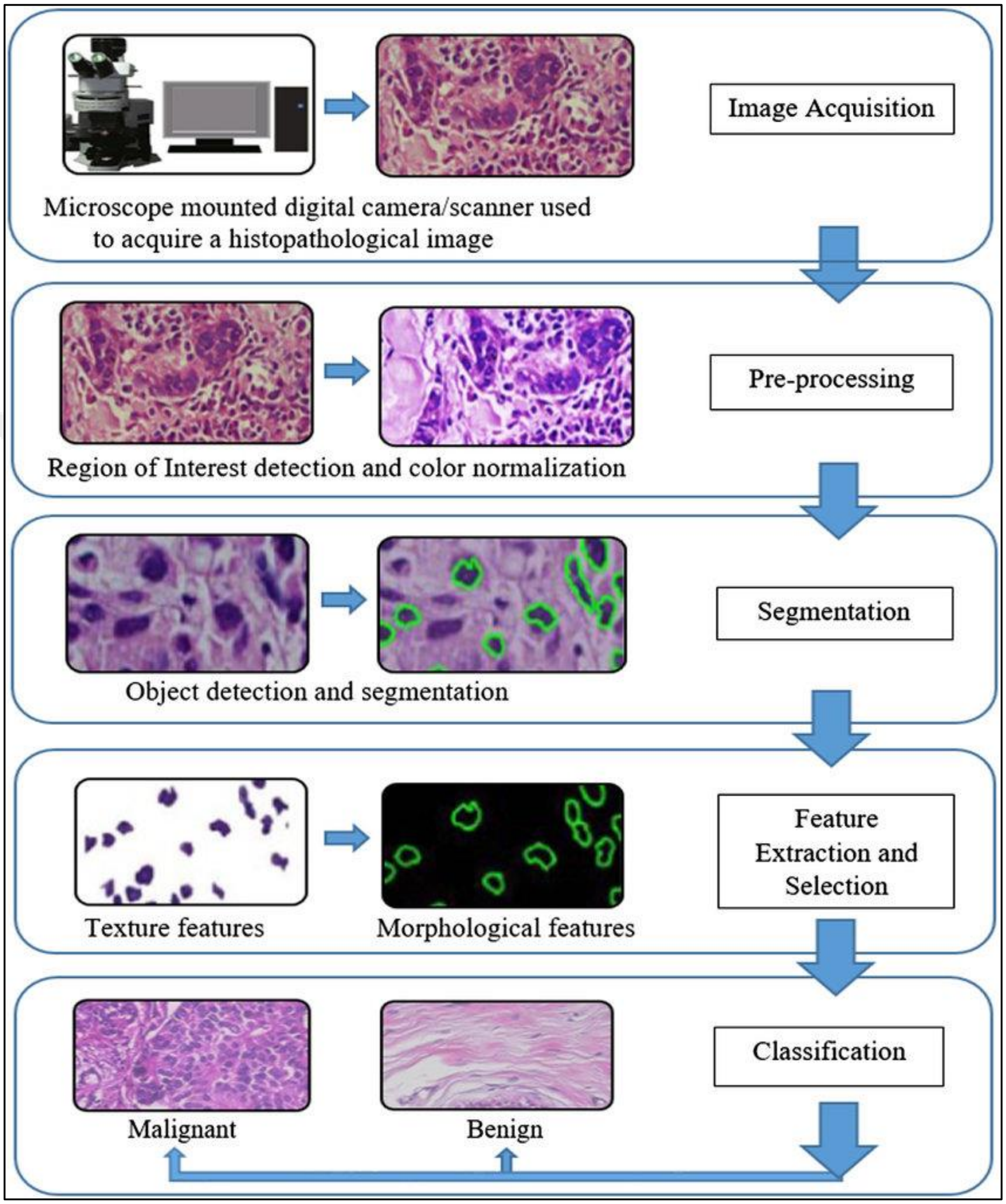


Figure 1.3: Computer aided detection CAD in breast cancer image processing [5].

1.3 THESIS OBJECTIVES

The overall objective of this project is to design a method for Developing a tool to aid the diagnosis of breast cancer using deep convolutional networks. To be able to answer this problem, it is necessary to achieve the following objectives:

i. Study the main characteristics of breast cancer and the most relevant properties of mammography exams.

ii. Apply modern tools of deep learning to evaluate anomalies in mammography exams.

iii. Compare the results obtained with other methods in the literature.

1.4 THESIS CONTRIBUTION

The research carried out within the framework of this thesis concerns the study of the malignant nature of breast masses, the study of the different stages of a computer-assisted diagnosis system, the comparison of the different decision-making aid strategies in breast imaging and the possible contribution and disadvantages of deep learning.

The originality of our work lies fundamentally in:

i. The proposal of an approach to reduce the size of a parsimonious dictionary learned on a mammographic image database for the reduction of the noise present in the images,

ii. The design of an unsupervised segmentation method that is robust, efficient, fast and requires only one parameter to be specified

iii. The development of an original technique for extracting spicules from the image

iv. The construction of two models offering diagnostic aid on both the benign/malignant character of a mass based on its shape, as well as the dense/fatty nature of the center of the mass.

This document is made up of four chapters organized as follows: In the first chapter, the anatomy of the breast was first studied, then breast cancer is presented in order to justify our interest in it. Then, a presentation of the radiological examinations that are screening and diagnosis of breast cancer highlighted the need to automate the interpretation of mammograms by radiologists. We then discussed the medical imaging tools used in screening and diagnosis while emphasizing the advantages and disadvantages of each of them. Then, we expose the link which exists between the composition of the anatomical structures of the breast and their appearance on an image after the realization of the mammography.

Next, we present the features ie. the shape, outline and density of masses in order to differentiate not only benign masses from malignant masses but also and above all to identify the differences between malignant spiculated masses and architectural distortions.

Subsequently, we define computer-assisted detection and diagnosis systems as well as the different stages that constitute them. This allowed us to focus on the problem of this thesis, namely the distinction between spiculated masses and architectural distortions. Finally, the different image bases that exist and that are used to design detection or diagnostic systems have been widely discussed.

1.5 THESIS ORGANIZATION

The dissertation is structured in five chapters:

Chapter 2 of the thesis presents a critical review of the literature on the various concepts useful during this work. The topics of the breast cancer, deep learning and feature extraction will be discussed. Then, chapter 3 will present the methodology adopted in our work to set up and validate the proposed solution. Chapter 4 presents the article in which the developed solution is presented as well as the criticism of its performance. The following chapter evaluates the success of answering the research question. The conclusion of the work as well as the discussion of future work are presented in chapter 5.

2. LITERATURE REVIEW

2.1 CHAPTER INTRODUCTION

In this chapter, we will first review the work related to the deep learning networks, their architecture and the weaknesses that emerge. Then we will see the different current strategies based on Artificial Intelligence to solve these problems. Finally, we will discuss the different training sets in the field of breast cancer detection. Several tactics for detecting breast cancer cells are addressed in further detail in the following critical evaluation of the literature. Additionally, a range of strategies for achieving improved results are discussed in depth.

2.2 RELATED WORKS

According to projections, there will be 2.26 million new cases of breast cancer in women by 2020 [1], making it the most prevalent form of cancer in women. There were an estimated 0.68 million cases of breast cancer that resulted in the death of women, ranking it fourth among the 35 other possible cancer sites [1].

Imaging techniques such as mammography, digital breast tomosynthesis, breast ultrasound, and magnetic resonance imaging are currently used for breast cancer screening (MRI). Two types of mammography exist: digital mammography (DM) and screen film mammography (SFM). Both of these types of mammography utilize radiation to produce a two-dimensional image of the breast tissue. Mammography has also reduced the risk of cancer-related mortality by facilitating the early detection of breast cancer in its earliest stages [2-4].

Imaging technology has developed to the point where patients can now receive digital breast tomosynthesis (DBT). [3-7] The DBT resolved issues with mammography and improved the image acquisition process. Initially, a multi-slice, two-dimensional image of the breast is captured, and then the results of these images are combined to create a three-dimensional image. Due to the fact that these 3D images (volumes) are constructed from a collection of 2D images, they are only quasi-3D [6-7].

In addition, an x-ray tube capable of capturing images in slices and rotating between 15 and 60 degrees is utilized. Automated Breast Ultrasound, also known as ABUS, is a high-frequency imaging technique that examines both breasts. The transverse plane generates two-dimensional images, which are then used to generate a three-dimensional volume [8].

The use of powerful magnets and radio waves generated by a computer are required for magnetic resonance imaging (MRI). The ability to obtain a "second opinion" from CAD systems enables radiologists to make diagnostic decisions with greater confidence [9-11].

The requirements listed in [12] for a CAD system include enhancing the performance of radiologists, reducing time waste, and achieving seamless workflow integration. The integration of deep learning algorithms into CAD systems aims to achieve a number of objectives, including a reduction in recall rates and an increase in cancer detection rates [13, 14], as well as a reduction in assessment variability between radiologists [13,14].

Currently, deep learning models are superior to those used by radiologists for both the classification and localization of cancer tumors [15, 16].

Access to large, meticulously curated datasets has been facilitated by the development of more efficient algorithms and more potent computers. If you work in the healthcare industry, techniques from the field of machine learning can help you complete your responsibilities. Deep learning, a subfield of machine learning, simplifies learned representations to solve the problem of learning multiple representations [18].

Deep learning, on the other hand, is a technique that employs multiple artificial neural networks to identify patterns of more precisely expressed actual representations. This is performed to enhance precision. In 2012, the use of convolutional neural network architecture led to a 15.3% error rate. This represented an increase in accuracy of 10.9% compared to the second-place entry. Since this breakthrough, interest in deep learning and the application of CNNs in image recognition has increased. CNNs are specialized networks used for learning the spatial hierarchies of features contained in processed data and for processing data in a well-known grid pattern. ing modalities, the overall mortality rate decreased thirty percent, reaching seventy percent. [11].

Deep learning-based computer-aided design (CAD) systems have made significant advances in the medical field to aid in disease diagnosis and prognosis by analyzing data from multiple fields, including radiology (11), pathology (14), cardiology (15), pharmacology (16), oncology (17,18), and genomics. These disciplines consist of radiology (11), pathology (14), cardiology (15), and genomics (16). (19). In recent years, machine learning techniques have enabled the development of more sophisticated cancer detection techniques. DL is one of these techniques; it has been shown to accurately predict cancer and provide a prognosis [20], making it one of the most popular options.

Compared to mammograms, ultrasounds, and DBT, DL has a higher diagnostic accuracy [21] for detecting breast cancer. For the foreseeable future, DL will continue to be the primary clinical treatment for BrC. Various studies on deep learning and BrC [22–24] have been published over the past few years. Deep learning-based algorithms are able to handle the complexities and difficulties of automatic BrC diagnosis. Although a large number of review studies on the BrC classification have already been conducted, very few of these studies have been able to point future researchers in the right direction.

These studies have not investigated every aspect of deep learning; only a few topics fall into this category. The vast majority of BrC review studies focused on either traditional machine learning (ML) algorithms or generic artificial neural networks in order to make a diagnosis (ANNs). They were incapable of addressing deep-learning architectures such as generative adversarial networks (GAN) and extreme learning machines (ELMs). A number of innovative imaging modalities for BrC have been developed as a result of recent research. In contrast, previous review studies have, for the most part, ignored emerging imaging modalities such as infrared thermal imaging, digital breast tomosynthesis, and computed tomography.

2.3 CHAPTER CONCLUSION

Deep learning, which has demonstrated significant growth in supervised learning, has been of great assistance in enhancing radiology's workflow and decision-making. This is possible because of deep learning. This chapter provided an overview of the theory and applications of deep learning, as well as a discussion of the obstacles deep learning must overcome. In addition, it was hoped that this paper would shed light on the current state of deep learning for breast cancer diagnosis so that future directions could be easily discerned. There are a few review articles on digital breast tomosynthesis [26], but they cannot cover all imaging modalities used in BrC classification.

In addition, their analysis of deep learning-based techniques was incomprehensible because they did not provide a comprehensive summary of the advantages and disadvantages of previous research. Within the scope of this study, six factors must be considered: the model architecture of BrC diagnosis; datasets and image preprocessing; breast cancer imaging; performance measurements; and research directions to overcome the limitations described earlier. Consequently, this study provides a solid foundation for a future study that will conduct a comprehensive and in-depth investigation of the current state of the art in deep learning basedBrC image classification using CAD systems the authors in the literature mentioned some common disadvantages in their work which include:

i. Individually Taken Sample Sizes

Small datasets pose a difficulty for the training of deep learning models because it is necessary to familiarize the model with all possible cases to reduce the number of classification errors. In addition, a lack of standardized datasets makes it difficult to compare and reproduce past studies. The ideal training dataset for deep learning models should include normal mammograms, BI-RADS 1-6, mass types, calcium deposition, asymmetries and architectural distortions, and multiple cases within a single mammogram. Even though these techniques have been utilized in the past, large datasets that have been meticulously curated are still required for a well-trained model to use data augmentation techniques to increase the sample size from 5,040 to 40,320 2D slices. This is due to the fact that the greater the sample size, the more precise the model will be.

ii. Contradiction between Classes

Dealing with classes with variable proportions of total training data [46] is an additional challenge for deep learning. Class imbalance can lead to biases in the classifier, which in turn can result in predictions that are skewed toward either the positive or negative class, depending on how the data is divided between the two classes. When there is a class imbalance in the metrics used to evaluate the model, such as accuracy, the model's performance can also be negatively impacted. This issue can be resolved, however, by employing strategies at both the data and classifier levels. We employ random undersampling and random oversampling at the data level, and cost-sensitive learning and thresholding at the classifier level. Random undersampling and random oversampling are used.

iii. Memory limitations

Computational costs are the next items on the agenda. When working with training data that has a large feature space, such as when working with high-resolution or high-dimensional images, memory constraints become an issue. Due to the complexity of the task, computational and memory costs skyrocket when building a 3D CNN from scratch utilizing DBT data with large feature spaces, for example. D. Variability in System Settings Has an Effect on Image Quality The settings on the medical screening device and the manufacturer's specifications impact image quality, but image quality also impacts model performance. It is possible that the presence of these breast cancer screening images will hinder the model's ability to train, predict, or locate lesions [17].

Table 2.1: Summary of the literature review.

Author	Year	Findings
Jaya Sherma, R. P. Tewari, and J. K. Rai	2014	Choosing the best preprocessing method for a batch of mammography images based on their peak SNR
Sharma and Bobbin Preet's	2016	"Classification of Mammogram Images by Using CNN Classifier" The study's authors propose the abbreviation CNN for computer-aided design system. This research compares CNN to Logistic Regression.
Nathan Jacobs, Jinze Liu, and Erik Y. Han	2017	Co-wrote a piece named The goal of the research is to construct and improve machine learning models for mammography entire picture categorization, according to the study's authors.
J. C. Tobias Charistian Cahoon, Melanie A.Sutton	2020	These frameworks are increasingly used in classification and pattern recognition algorithms, including cancer detection. Medical diagnosis requires machine learning to examine and decide.
Kui YI, Yue-Hua DING	2019	The mammography picture is evaluated by identifying patterns and structures in the matrix. No longer must it separate distinct objects to estimate tissue size distribution.
Hua, K.-L., et al	2015	Image processing strategies for detecting malignant cells are examined. More research is required to understand malignancy. Texture analysis can distinguish benign from malignant mammographic masses.
Zhang, Y., et al. DeepSplice	2016	These features are extracted from the breast's sensitive zone for classification. Maximum Likelihood (MLE) estimates the tumor size.

3. MATERIALS AND METHODS

3.1 BREAST CANCER

Breast cancer affects 1 in 8 women in the arc of life. It's the tumour more frequent in the sex female and represents 29% of all tumours that they affect women. It is the leading cause of cancer mortality in the sex female, with a death rate of 16 % of all deaths from oncological causes [2-I].

The chance of healing is highly dependent on the degree of development of the pathology in the moment when it comes diagnosed: if the disease he comes identified in her phase initial, when the size of the wound is still reduced, the chance of recovery is high. It is estimated which, in these cases, five-year survival in the women treated is 98% [2-I]. But it completes it no symptoms, if not in phase advanced, makes the diagnosis extremely difficult. Becomes so necessary adopt a prevention and diagnosis policy early with controls periodicals. In case of grade tumor advanced therapies adopted they are always of type demolition, such as the removal intervention of the udder, followed by heavy radiotherapy and chemotherapy treatments, with dramatic consequences psychophysical for the patient. A tumor at the stage initial, however, often it only needs some kind of interventions conservative, such as removing the lump or a small one part of the udder. So here it is than the diagnosis early acquires an importance basic also from a therapeutic point of view.

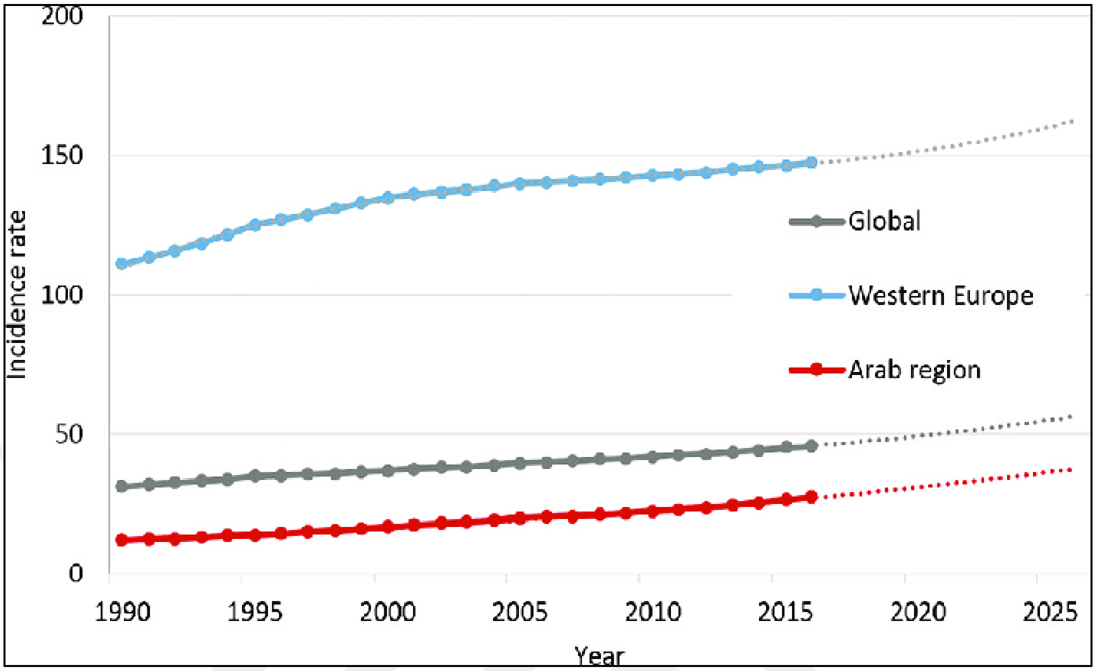


Figure 3.1: Breast cancer rate by 2025.

3.2 DIGITAL MAMMOGRAPHY

digital is an X - ray technique that allows you to have an internal view of the udder with good resolution spatial and high contrast. An X- ray beam, emitted by an apparatus radiological said mammogram, crosses the breast and comes so absorbed different according to the type of fabric that meets. The rays remaining, they come recorded by a sensor digital which converts them into potential electric. That resulting is an image grayscale digital, which can to be transferred numerically on a computer and displayed on a screen, representing the internal structure of a breast The internal structure of the breast can to vary considerably from subject to subject and consequently his too. If the breast is formed mainly from fabric fibrous mammography will result bright (in this case yes speaks of udder dense), if instead the breast is predominantly formed from fabric mammogram fat will appear darker. Yes, it comes then to create a match proportional between the intensity of color and the density of the fabric corresponding (Figure 3.1).

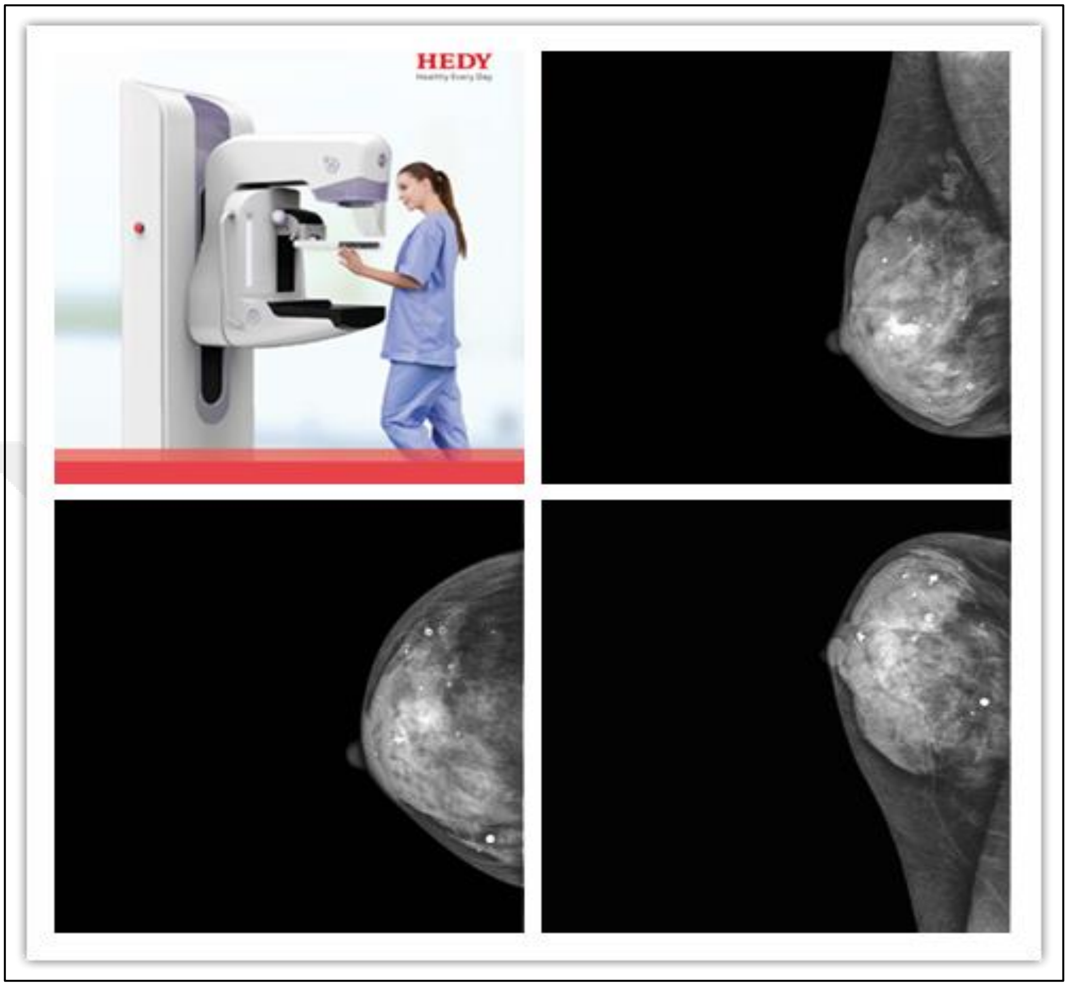


Figure 3.2: Digital mammography.

A basic screening consists of two screenings for each udder: one long the crania - caudal axis (CC) and an oblique one medio -lateral (MLO), inclined at 45 ° The technique traditional for the detection of breast cancer is the analysis, by a radiologist, of the image produced on screen. The injuries present in the otherwise have a degree of absorption characteristic, for this reason in greatest part of the cases for the radiologist it is relatively easy to identify them. However, despite the research have improved notably the techniques in the course over the years, yes esteem that on the wholeness of the injuries neoplastic mummeries do not come diagnosed between 10% and 30% of cases [32].

The factors that contribute to an error rate Like this high Is different: first of all, nature itself of the mass, characterized from one great variability of characteristics morphological; also, this detection is made difficult by the similarity with the fabric host Big problems yes they typically have in patients ' breasts young people, in the Which yes features a fabric very dense fibrous, and therefore low contrast between the various areas of the fabrics.

The last one stage of the trial is the verdict of the radiologist same . Here, in addition to those above, take over the imperfection factors human as, for example, fatigue of the eyes due to work too much intense, the fact that yes lend greater pay attention to parts of the image in Which yes thinks are more likely find masses or the intrinsic difficulty of interpretation of the pictures.

3.3 COMPUTER AIDED DETECTION

CAD View the fundamental importance of a diagnosis early and difficulties found by the radiologist, over time they are States introduced mechanisms that can somehow facilitate it, reducing the possibility of error.

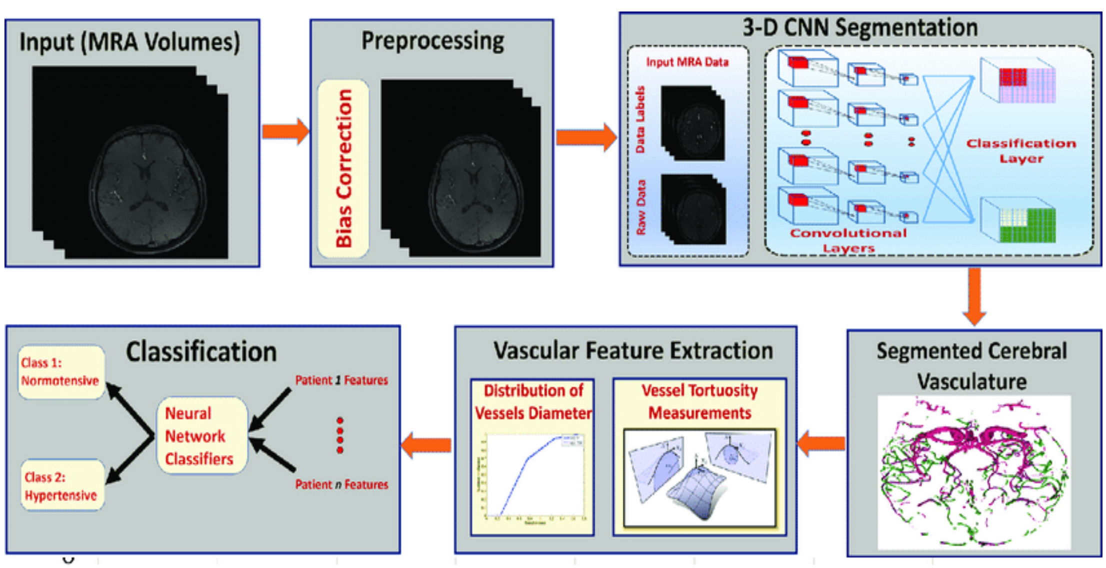


Figure 3.3: Example of computer aided detection.

At the beginning it was in the 70s introduced computer use in the mammogram for improvement image (contrast increase, filters, etc. ...). However, the lack of an image acquisition system digital yielded the scanning phase is necessary: this introduces noise and artifacts in images and results in a loss of information.

Then they were developed techniques of analysis in Which they came two radiologists employed: both they did their diagnosis independently and then the results. Subsequently, the advent of the mammography and the success of the double analysis method they have led to one of his evolutions: the replacement of one of the two radiologists with a detection system automatic said precisely CAD. It is important underline that the CAD sentence is not to be understood exhaustive, but only indicative to the radiologist of the regions suspicious. In this way the specialist will have a chance to do one more careful assessment of the aforementioned regions. Most part of the CAD systems is composed of two levels of classification:

i. detection, responsible identification of the regions suspicious present on mammogram (Region of Interest - ROI) and then elimination preventive areas not at risk.

ii. the classification (classification) of the ROIs in masses and tissue healthy. In practice both of them the levels perform a classification operation.

The difference is in the done that the detection classifies the regions as suspicious and unsuspicious, discarding the latter while the classification analyzes only the “surviving” regions at the first level and classifies them into true and false masses alarms. One feature essential of the detection algorithms will be so a high sensitivity, i.e. the ability to find suspicious areas : each mass lost in this phase , in fact , fleeing to the analysis of the levels subsequent , will come lost forever. Usually, the price to pay for sensitivity high consists of a high number of fakes positive.

3.4 PERFORMANCE EVALUATION

in this paragraph we introduce some elements of analysis accuracy of the classification systems dichotomous. The problems with this like all those cases in which the elements to be classified they are divided into two classes and the belonging of an element to a class excludes membership to the other. In particular they belong this category the medical diagnostic systems, which must distinguish between positive and negative.

The efficiency of a system diagnostic can be evaluated considering the following parameters:

i. TPF: fraction of true positive (true positive fraction);

ii. TNF: fraction of true negative (true negative fraction) These values, indicated respectively as sensitivity and precision, they measure how many positive and how many negatives they come recognized correctly with respect to the relative one's total: Yes, they can to introduce two other fractions:

iii. FNF: fraction of fakes negative (false negative fraction); - FPF: fraction of fakes positive (false positive fraction); These indices result the complementary of the previous in that indicate the fractions of the positive and gods' negatives classified mistakenly. We therefore have: Through the pairs (TPF, TNF) or (TPF, FPF) it is therefore possibly knowing the values of all indices defined for the diagnosis system studied.

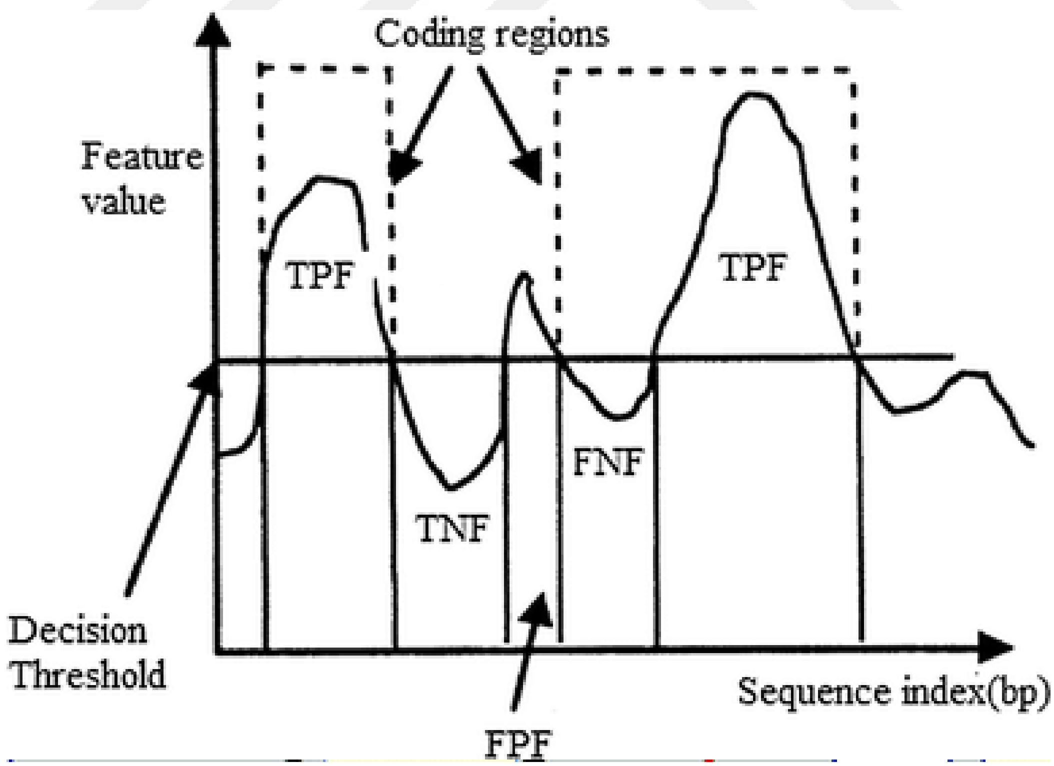


Figure 3.4: ROC curve with true positives and true negatives, false positives and false negatives shown on the curve.

ROC and FROC curves There are two characteristics fundamental for I diagnostic systems that sensitivity and precision, or I They complementary, they do not allow to highlight. The first is capacity intrinsic to distinguish the positive come on negative and it depends from the overlap of the probability distributions of the two classes. These indicate, for each value of the variable characteristic used in classification, probability what an element belongs to one or the other class The second it depends on the fact that's the distinction happens applying a threshold at the variable feature and just applying this threshold is possible obtain sensitivity and precision. they are representing the two distributions. ROC The threshold defines a certain working point of the classification system. To get an analysis overall system is possible evaluate the pairs (TPF, FPF) by values different of the threshold in a fairly range ample to cover distributions of the two classes.

The graph that yes gets placing the FPF values on abscissa and I respective values of TPF on the ordinates is the so -called Receiver Operating Characteristic (ROC) curve. This allows you to view performance of the system, as the operating point varies. The diagonal that unites the points (0,0) and (1,1) indicate a diagnosis system whose responses Is independent give it entrances elements; such a system in no way distinguishes the two classes. One system optimal, that is that distinguish totally the elements of the two classes, it will be characterized from the curve formed come on segments that join the point (0,0) with the point (0,1) and the latter with the point (1,1). Systems for I Which there is an overlap between distributions Is characterized by intermediate curves between the cases limit described and performance improve as you approach of the curves in point (0,1). Some examples Is represented in Figure 3.5.

The ROC curve allows you to view the performance of the system in his various work points but provides also an index quantity overall for this performance. This yes gets evaluating the area under the curve (), which can be interpreted as: - the average value of the sensitivity for a precision randomly chosen between 0 and 1 or - the mean value of the precision for a sensitivity randomly chosen between 0 and 1.

3.5 MACHINE LEARNING

Introduction Limitation efficiency and robustness with which the brain human represents information has been the central challenge for decades intelligence research artificial (artificial intelligence - AI). The obstacle principal that yes meets in this challenge, in particular in the area of pattern recognition, it has to do with high dimensionality of the data: in this context, in fact, the complexity of learning grows exponentially with dimensionality of the data. This it means than the number of examples necessary to estimate, with a given level of accuracy, a function arbitrary, it grows exponentially with respect to the number of input variables (features) of the function. With a number set of training examples, therefore, power predictive decreases as it increases of the dimensionality. Richard Bellman, one of the first courses to face up to this problem, the defined it with the expression “curse of dimensionality” [15].

The approach traditionally more widespread upon overcoming this obstacle consists in the divide the pattern recognition system into two modules the feature extractor, which pre - processes the input patterns to decrease the dimensionality, and the classifier (not to be confused with the concept of classifier introduced for CAD systems). This approach however has a great one limit: the accuracy of the recognition is highly dependent by the skill of the creator in choose the whole appropriate features and therefore the type of problem. It is born hence the need to find a valid mechanism more general. Someone discoveries in the scope of the neuroscience [15, 16] have provided explanations on the principles that regulate the representation of the information in the brain of the mammals, leading to new ideas for building systems that imitate these capacity . One of the discoveries main it concerns the neocortex, which is responsible for many cognitive skills: it does not pre -process explicitly the signals sensory but, rather, it makes them flow along a complex hierarchy of modules that, over time, learn to represent observations based on regularities that exhibit.

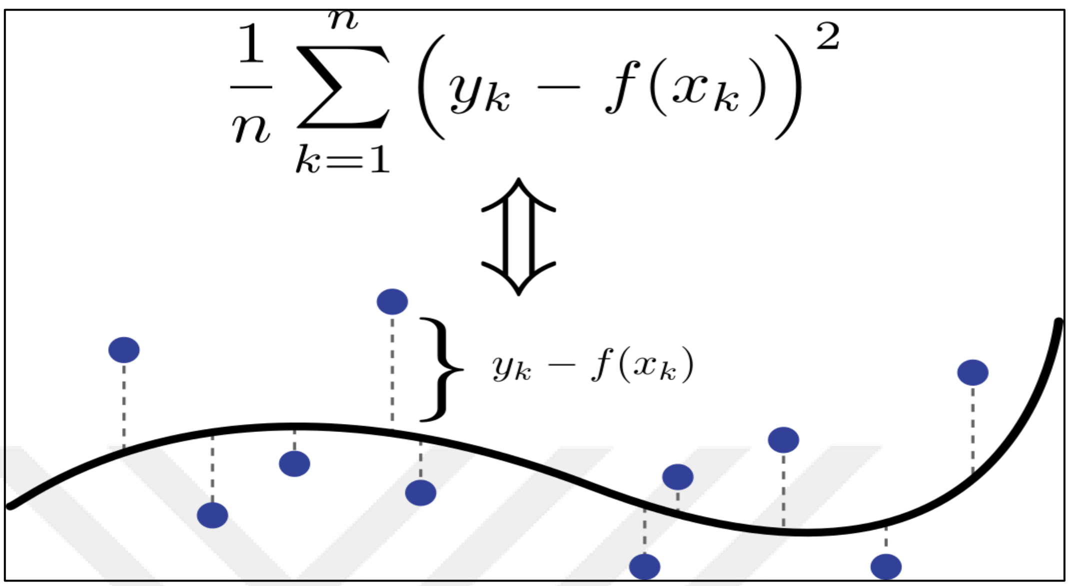


Figure 3.5: Early machine learning model.

This discovery motivated the birth of deep learning (DL), yield possible from the combination of several factors such as the availability of low- cost machines with unit's arithmetic fast, the rise of the size of databases for broad problems market and interest. The DL is a new one subcategory of machine learning (ML) that yes concentrates on model development computational for representation of the information structured with features similar to those of the neocortex.

Inspired by the results of the searches in the scope of the neuroscience, the DL basically has the purpose of bringing machine learning back to one of his targets originating: intelligence artificial. To do this it uses a set of algorithms that shape independently the data through a hierarchy organized on different levels (layers) of representation and abstraction: I concept at the levels the higher they are defined starting from those at the levels more bass through a series of nonlinear transformations.

The brain of the mammals, in fact, is organized in an architecture deep where a certain input data he comes represented by more levels of abstraction, each corresponding to a different area of the bark. The beings humans describe such concepts in a manner hierarchical, with more levels of abstraction.

The brain it seems also process information through more phases of transformation and representation. This is particularly evident in the system visual primate, with his sequence of processing steps: detection of the edges, perception of the forms, from primitive to gradual ones more complex.

3.5.1 Supervised Learning

The expression “Artificial Intelligence " was minted in 1956 by the American mathematician John McCarthy, during a historian seminary disciplinary held in New Hampshire: with this term yes intends generally the ability of a computer to perform functions and reasoning typical of the mind human. In his appearance purely IT, it understands the theory and techniques for developing algorithms that allow machines to show a skill and / or activity smart, at least in specific domains. Machine Learning (ML) is one of the areas fundamental intelligence artificial and yes occupies of the realization of systems and algorithms that yes based on observations as data for the new synthesis knowledge.

Learning can take place capturing features of interest from examples, structures data or sensors, to analyze them and evaluate their relationships between the variables observe. For example, an ML system can be trained on emails to distinguish between spam and non-spam messages. Learning supervised The field of learning is very broad and later we will limit ourselves to considering the so -called learning supervised: this term yes reports generally to machine learning techniques that aim to train a system computer scientist to solve of the tasks based on a number of examples ideals. The system learn that is to approximate an unknown function from a set of training examples formed by pairs input-output:

$$J = \sum_{i=1}^{n} \frac{(mX_i + c - Y_i)^2}{n} \tag{3.1}$$

for each input yes communicates to the system the output desired. Learning consists in capacity of the generalization system to new one’s unknown examples. Formally yes defines learning supervised the approximation of functions of unknown form, starting from a set of pairs of values, considering that these values are obtained from where with yes, they intend possible noise contributions. It goes underlined than in this case the shape of the function is unknown. There are two categories main learning supervised:

i. Classification, if the value hired from the function Is discrete and represent the classes to which they belong of the vectors;
ii. Regression, if the function takes values continue. Within the scope of the classification, the case particular in which yes distinguish only between two classes covers an importance remarkable.

An example of this type is represented by the problem addressed in this thesis, in which the distinction happens between masses and tissues healthy. Exist different learning algorithms supervised but they all share a characteristic common: training, training, takes place minimizing the so -called function cost, loss function, that represents the error accomplished in esteem of the values associated with the input data, via the function.

The possible multiple choices for functions cost but most part of the systems uses the reject function quadratic: ∑ where˜ indicates I values taken out of the system when come introduce yourself the incoming. The choice of the function to be minimized implies the principle that is at the basis of learning.

Use a function which, like that of waste quadratic, is a computation direct of the error on all examples of the training set, means Suppose that minimizing the error on these data yes minimize also the error incurred by the machine when is called to generalize, that is, to determine the value of the function for points do not present in the training set. In practice this assumption, especially for a small set of data, is not valid , because this type of minimization leads adaptation too much specific of the machine aid used for training , an effect called , in literature , overfitting. This leads to a worsening of the performance in the generalization phase.

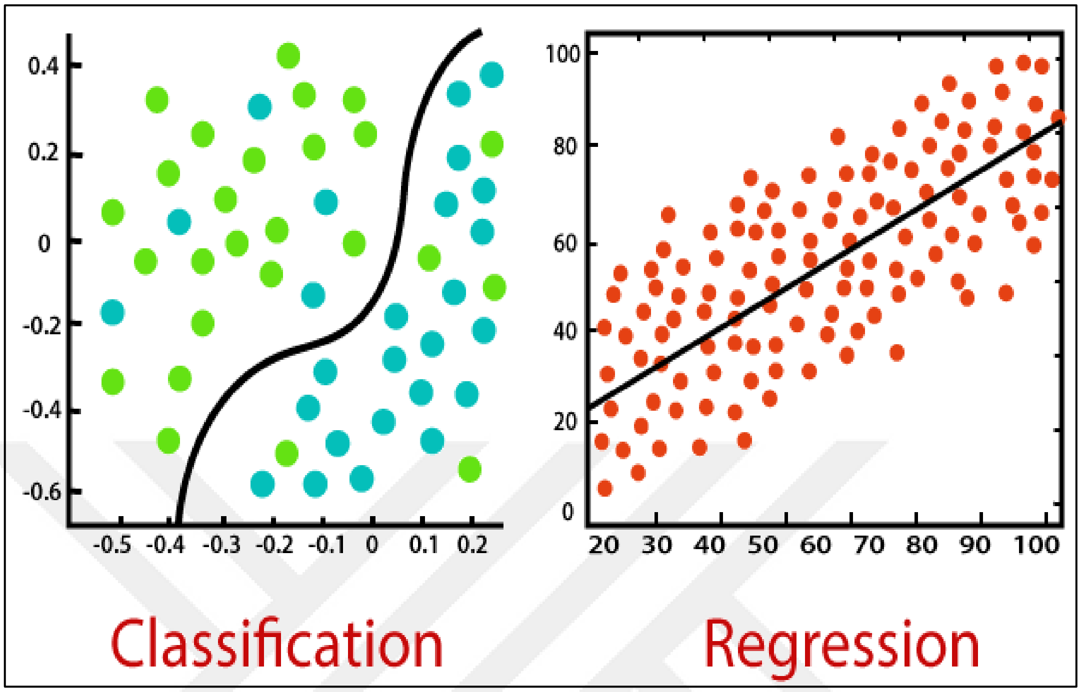


Figure 3.6: Regression vs classification operations in machine learning.

To remedy these problems is possible look for an estimate different of the error in generalization. This yes can obtain checking efficiency through a dataset not used in the training phase, evaluating the loss function for this together. This approach, called validation, allows also to optimize implementation parameters not directly belonging the learning algorithm, for example, in the case of image recognition, size of the input images. Also optimizing the system through the validation data, yes can anyway run into an adaptation too much specific. For this reason, a third together, called test, comes used as a final check. If the error obtained it is similar to that of validation the overfitting did not occur.

Cross Validation The training, validation, test schema is the approach traditional to the problem optimization of a learning system supervised.However, for many problems, a database is not available enough wide to allow a division into three set of dimensions sufficient. One technique developed to manage situations of this type is cross validation. It allows you to use all data available to optimize the system. The whole of the data he comes divided into N partitions, fold, which have approximately the same dimension, to then train the system with N-1, using the remainder for validation.

This method he comes repeated N times excluding each time a set different. The results they are then merged as if the validation had been performed on everything the whole of the data. In this way all examples available Is used for learning. Despite this technique increases remarkably the calculation times, as training it goes performed N times, various Education theorists and simulations they have tried a good one effectiveness of this system in obtaining good estimates for the generalization efficiency.

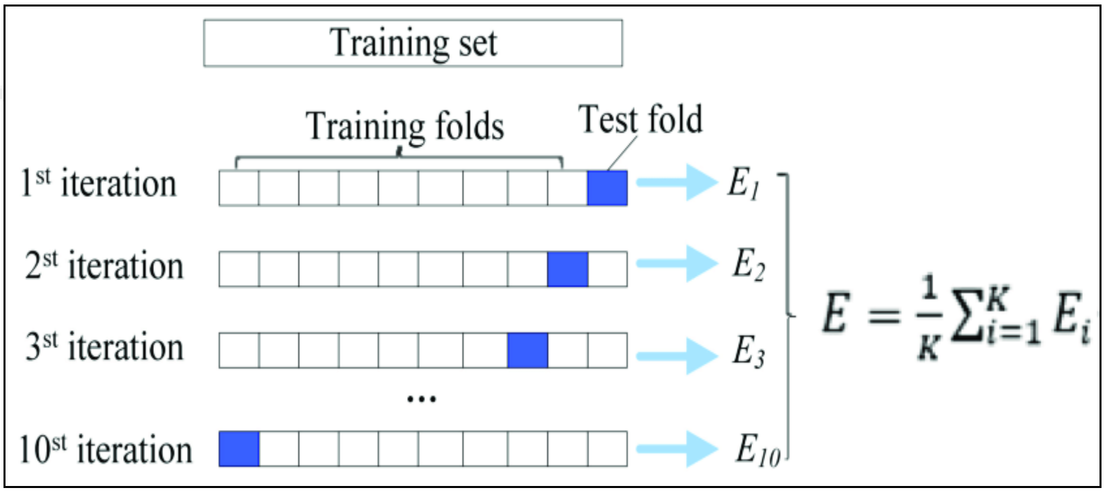


Figure 3.7: Cross validation with E number of training iterations and K folds.

Gradient-based learning and gradient back-propagation the problem general of the minimization of a function with respect to a set of parameters is a root of many problems computer scientists. The loss function can to be minimized making an estimate impact that they have on its small variations of the values of the parameters. This can to be measured via the gradient of the loss function with respect to the parameters. This is the basic idea of many valued parameter algorithms continue called gradient-based learning algorithms. In general, the parameter set is a vector with values real, with respect to which the loss function is continuous and differentiable. Give these conditions, the minimization procedure used more often it is the gradient descent algorithm where it is iteratively " adjusted " as follows:

$$\frac{\frac{1}{n}\|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}\|^2}{\left[\frac{1}{n}\mathrm{Tr}\left(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^T\right)\right]^2} \tag{3.2}$$

where it controls the descent rate along the gradient and, in cases simpler, it is a constant. Therefore, given a point in the space of the parameters, to minimize, yes moves in direction opposite to that of the gradient. The utility amazing of the techniques based on the gradient descent for complex machine learning tasks remained unknown until at the discovery of the back-propagation algorithm (which we will see in details in the next paragraph) by Hinton, Rinehart and Williams [30], which allows the computation of the gradient of non - linear systems composed of levels multiples of processing. The idea underlying the backpropagation is that the gradient can be calculated efficiently through propagation of the mistakes from the output to the input.

3.5.2 Neural Networks

Among the learning algorithms supervised, those derivatives give her networks neural they have played a role basic in field development. The origin conceptual of the networks neural yes find in modelling communication system abstract between neurons. The neuron A neuron is a cell composed of a central part, called soma, from which yes branch off a main nerve fiber, the axon, and a series of ramifications, I dendrites. The neurons Is connected between They through the synapses, which Is the contact points between the part terminal the axon and I dendrites or the soma of other neurons.

The communication between neurons happens in the following way . When a neuron is active sends a pulse electric through his axon; this impulse he comes transmitted, through the synapses, to neurons connected. The transmission happens through the exchange of substances chemicals, calls neurotransmitters, from the axon to the receptors present on the soma and on the dendrites.

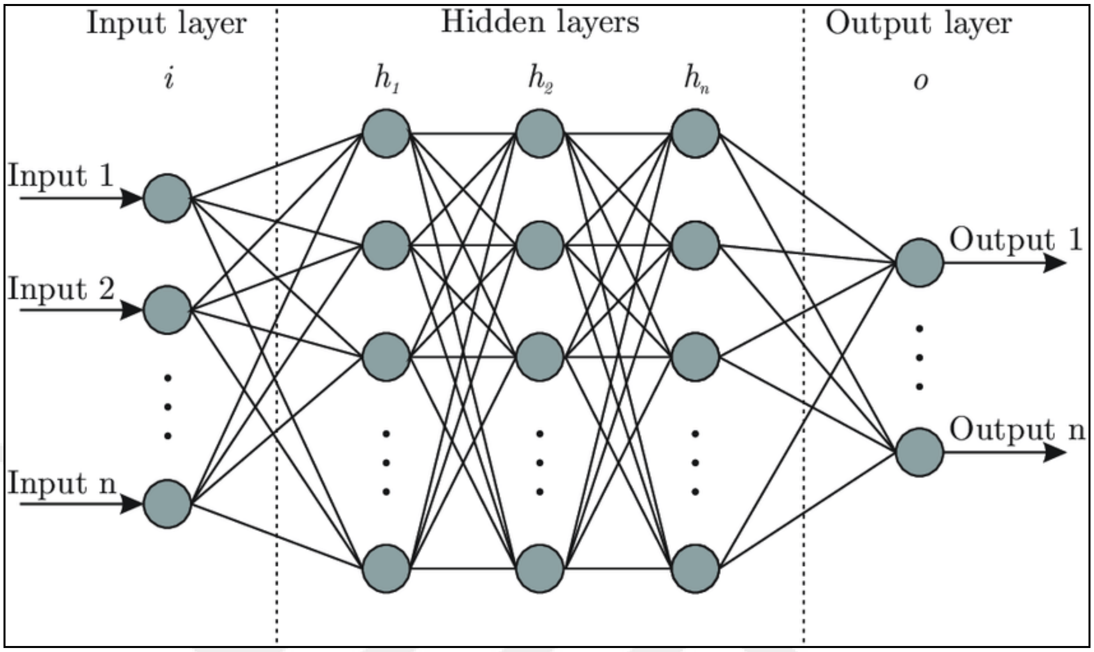


Figure 3.8: Artificial neural network architecture.

According to the type of stimulus and the configuration of the receptors, an impulse changes the potential of the neurons connected. Every neuron has its own activation potential, which can be reached only when yes come to create certain activation configurations of the neurons to it connected. Once this potential is reached, the neuron yes activates and in turn transmits the impulse to the neurons connected at the part terminal of its own axon. The ability to learn is in possibility to modify the characteristics of the receptors in order to modify the configurations that involve activation of the neuron.

Networks neural artificial Starting from this basic knowledge of how the system works nervous, The States built models simplified both in order to study the functioning of the They fees biological that to check if such models are equipped with learning skills. The element fundamental of all these models is the neuron artificial, first introduced by McCulloch and Pitts in 1943 It is a computing unit with N inputs and one output.

Each of the inputs represents a termination synaptic and is therefore connected at the exit of other neurons artificial. At each input is associated also a synaptic weight. The output values can be two values discrete, usually 0 and 1 or -1 and 1 that indicate the status active or inactive neuron; or of continuous intervals, generally [0,1] or [-1,1], which indicate the degree of activation of the neuron. The calculation of the output value he comes carried out determining first of all the potential P, through a sum of exits of the neurons connected, weighed through the weights synaptic:

$$W_{new} = W_{old} - \eta * \frac{\partial E(X, W)}{\partial W} \tag{3.3}$$

The exit yes gets directly from the potential through a transfer function: In case the output must to be Discrete yes can use threshold functions such as the sign function or the Heaviside step ϑ. In the other cases yes can choose different types of functions. In the majority part of the cases yes use the tangent hyperbolic or function sigmoid in that they allow to regularize respectively the sign function and the ϑ function, being both derivable and having the same behavior of these functions ai extremes. Connecting these units according to various topologies is possible obtain different types of networks. Learning happens modifying the weights synaptic in such a way as to minimize the error during the training phase, to do this it is necessary define a law of learning that die the new values of the weights starting from the value of the loss function. These methodologies they will come exposed, by chance particular, in next paragraph, where we will describe a very common type of network suitable to learning supervised. Multi-Layer Perceptron - MLP The multi-layer perceptron is a type of neural network artificial studied specifically for learning supervised. The model consists of at least three layers of neurons artificial connected in one direction, feed forward. Every neuron, as well ai inputs relating to outputs of the neurons previous, possesses a further input: the bias. To it he comes given a value fixed equal to 1. This involves a term constant in the sum. Often, in the calculation of the number of the layers, the first does not come considered Why his elements do not possess weights synaptic: indeed, this layer yes just submit the values of each input to neurons of the layer next, running simply a transformation via the transfer function. for example, is considered two or three - layered, counting respectively the number of weights or neurons.

It is preferable to count the layers of weights, as it is precisely the modification of these last that makes the network able to learn. The first and the last layer must have a number of units even respectively at the dimension of the entrance and exit spaces: these are the terminations of the “box black " which represents the function to approximate. For the two- class classification, only one output is sufficient, the value of which is passed across the threshold, will indicate the belonging of an incoming vector to one of the two classes. In case yes have more than two classes yes uses a number of outputs equal to the number of the classes themselves, too subject to threshold. The output value of the network yes calculates in the following way. At each neuron of a layer yes apply the sum and the transfer function:

$$W_{new} = W_{old} - \alpha \underbrace{\frac{dJ}{dW}}_{\text{gradient}} \tag{3.4}$$

The values calculated Is the inputs of the layer next, whose outputs yes, they calculate at the same way. We proceed Like this until you reach at the last layer whose outputs represent the exit of the whole network. We will see that’s the law of learning requires that the transfer functions of the neurons are derivable. Later, for simplicity, we will limit ourselves to the case detail of a net with two layers of weights that use as a transfer function that sigmoidal. We introduce the following notation: we consider I inputs, J elements internal, K go out and use I, j and k respectively to indicate the indices referred ai elements of the layers. represents the synaptic weight of the j- the neuron of the layer internal reported to the itch entrance and similarly represents the weight of the Kasimov output neuron referred to the j- the neuron internal. The bias values are Le aforementioned laws of learning they have the same shape. That is that differentiates them are the definitions of and. The second turns out dependent on the first and, in this sense, yes speaks of propagation backwards of the error, hence the name of the algorithm. Turns out necessary at least one layer internal so that yes may introduce non-linearity, which turns out the element basic of the flexibility of the networks.

$$\sum_{i=1}^{m} w_i x_i + bias = w_1 x_1 + w_2 x_2 + w_3 x_3 + bias \tag{3.5}$$

A classic example of the ineffectiveness of a single stratum is the inability to get a network of this guy that may calculate the logical XOR. The internal architecture of the network has a notable impact on performance. The multiplicity of configurations possible involves a number of solutions suboptimal. Also, the same learning algorithm can stop reaching minimum local. A problem that generally afflicts all learning systems supervised is the curse of dimensionality, a progressive decay of the performance as it increases of the dimension of the entrance space. This happens because the number of examples necessary to get a sampling enough of the entrance space increases exponentially with the number of the size. The problem turns out particularly significant in networks neural.

3.5.3 Support Vector Machine-SVM

Support Vector Machines (SVM) are a set of learning methods supervised for regression and pattern classification, developed in the 90s by Vladimir Vatnik and his team at the Bell AT&T laboratories. They belong at the family of the classifiers linear generalized and are also known as maximum classifiers margin, since at the at the same time they minimize the error empirical classification and maximize margin geometric. Support Vector Machines were invented by Vladimir Vatnik and I his colleagues from AT&T Bell Laboratories and first introduced in 1992 at the Computational Learning Theory conference. The roots of this approach, the method of Support Vectors to build the separation hyperplane optimal for the classification of patterns, they go back to the 1964 work of Vatnik and Chernenko’s [18].

Today the SVM are become one of the standard tools in the machine learning community for classification,regression and estimation problems of the density. The method more intuitive to understand the application of the SVM consists in studying the problem of the separation of two classes from a geometric point of view . Let 's consider an entrance space for simplicity two-dimensional.

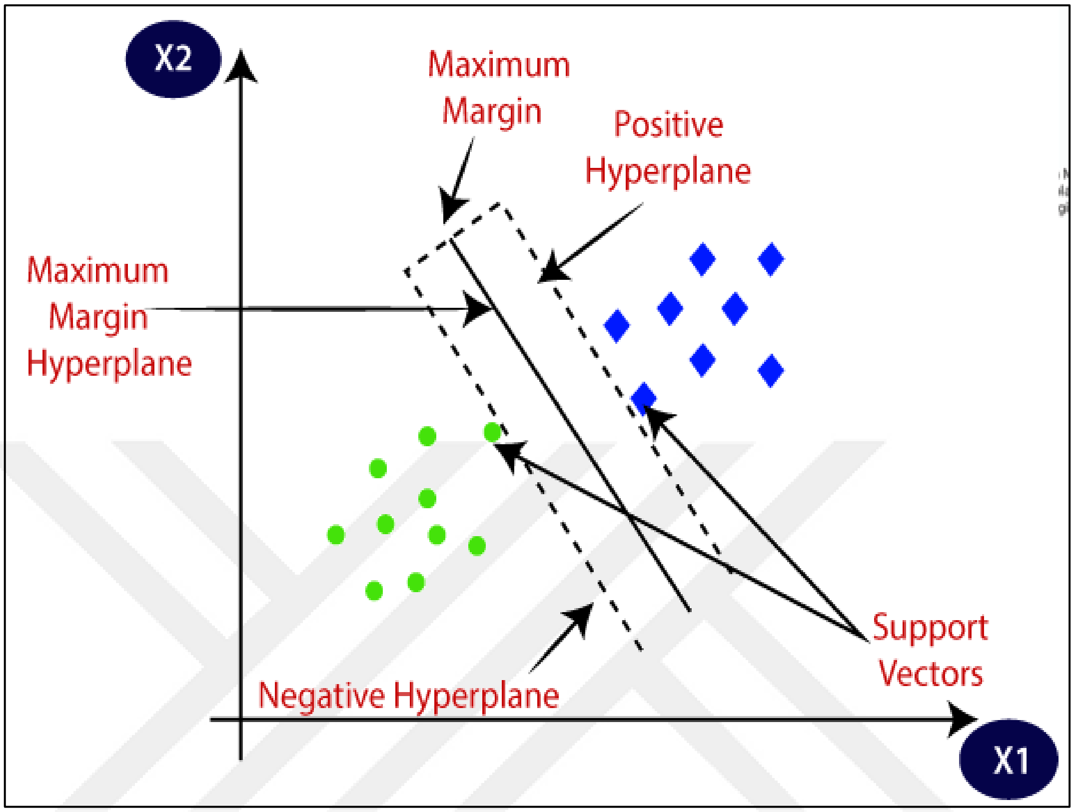


Figure 3.9: SV classification operation.

The elements of the two classes they will be scattered in the plan and, according to their characteristics, they will occupy different regions. In case it turns out possible to separate such classes via a straight line or, in the case of a number higher in size, a hyperplane, all oral SVM allow you to choose uniquely that optimal It is obtained that this hyperplane optimal is that for which it results maximum margin between the two classes, i.e., the minimum distance of points of a class from the hyperplane.

The separation hyperplane divides the space into two parts. The system classifies a new input element as belonging to one or the other class based on which of the two regions contains the point of the space identified by this element.Can to happen that separation complete of the two classes using a hyperplane not either possible. Whatever hyperplane yes choose there will be some points of a class that yes, they find from the part of the space associated to the other class.

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \; \phi(x_i) \cdot \phi(x_j)$$
$$where \; 0 \leq \alpha_i \leq C \; \forall \; i \tag{3.6}$$

This problem yes succeeds to face up to modifying the algorithm: a cost is introduced associate ai classification errors in optimization. In the cases of separation complete yes near the separation hyperplane directly in entrance space. In general,however, the separation through a hyperplane it is inefficient. we see how the distribution is too much complex to be separated in this way, too considering the mistakes previously introduced.

They turn out much more effective curves such as those drawn. The complexity of the distribution can to be reduced by " mapping " the elements in a space of dimensions major, said of features, through a non- linear function ϕ,non-linearity of the function allows you to redistribute the elements to create a separation efficient using a hyperplane. It is therefore possible apply the method indicated for SVM in this space.

In general, knowledge of ϕ is a problem computational too much complex; however yes can demonstrate than to apply the algorithm is uniquely necessarily know the product scalar of two vectors in space features in operation of the They fees carriers in entrance space.

$$\begin{aligned} \text{maximize } f(c_1 \ldots c_n) &= \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)) y_j c_j \\ &= \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i k(\mathbf{x}_i, \mathbf{x}_j) y_j c_j \end{aligned} \tag{3.7}$$

The SVM algorithm owns someone property that make it preferable to other algorithms for learning supervised. Let 's take networks in particular multi - layered neural. We have seen that a feature of these networks is the possibility that optimization stops in minimums local. This there is no problem in the case of SVMs as the optimization function is convex, quadratic and constrained linear. This involves the totality of the minimum. From the Karsh - Kuhn-Tucker theorem it results that the solution of the problem depends only on a subset of training vectors, the support vectors.

That is, the algorithm allows you to choose automatically Which information to use and which one'sdiscard, defining so architecture optimal. Just in case of the networks multi - layered neural,however, the architecture is to be determined through a validation process. The ability to use kernel functions allows to render the algorithm suitable for separating very complex vector distributions. If indeed this was not possible, the non- linearity of the transfer functions of the neurons in networks neural would make it these much more flexible compared to SVMs. Finally, various Education they have established which, compared to networks neural, the problem of the curse of dimensionality results greatly reduced.

3.5.4 Deep Learning

The concept of depth (hence deep learning), referred to a data modelling tool can be illustrated through an example. The operations run to arrive from the input to the output can be represented via a flow chart composed of various nodes which represent an operation elementary and give a result addicted to those of neighbouring nodes. For example, the flowchart for the expression, in consists of two input nodes and, one node for the division which has as input e (i.e., as children), a node for the square (which takes as input), a node for addition (input e) and, finally, the output node that calculates the sine, with only one input from the node addition. Depth, defined formally as path length longer that connect input and output, it is a property important of such graphs.

A depth equal to 2 is usually enough to represent any function with precision arbitrary. This However involves a price to pay: the number of nodes required by the graph can become very big.So, for learning complicated functions, which can represent high levels of abstraction, it can be necessary resort to deep architectures (composed of levels multiples of non - linear operations), as in networks neural layers with many hidden layers. This however, it is not the only one reason for the study of learning algorithms for deep architectures. In fact : - The brain has an architecture deep - The men organize their ideas hierarchically - Gil men learn first concepts simple and then compose them to represent concepts more abstract - Non-deep architectures can to be exponentially inefficient - Gil engineers they break down the solutions into levels multiples of abstraction and elaboration Driven by these reasons the deep learning methods aim , through the use of deep architectures , the learning of features hierarchies , with the features at the levels taller formed by composing those on the levels more bass . Taking inspiration to architecture deep brain, I researchers in the field of networks neural they have tried for decades to train multi - layered networks, but until 2006 there are none States successful attempts: I results were positive up to two or three layers (i.e., one or two hidden layers), but increasing the number yes observed a deterioration of the results. In 2006 Hinton et al. At the University of Toronto have given thrust to the scene decisive introducing the Deep Belief Networks (DBN), i.e., networks which, using an unsupervised learning algorithm (Restricted Boltzmann Machine - RBM), train individually every layer. Since 2006, deep architectures have been successfully applied not only in classification tasks but also in regression, reduction of the dimensionality, texture models, motion modeling, object segmentation, recovery of the information, robotics, language processing natural.

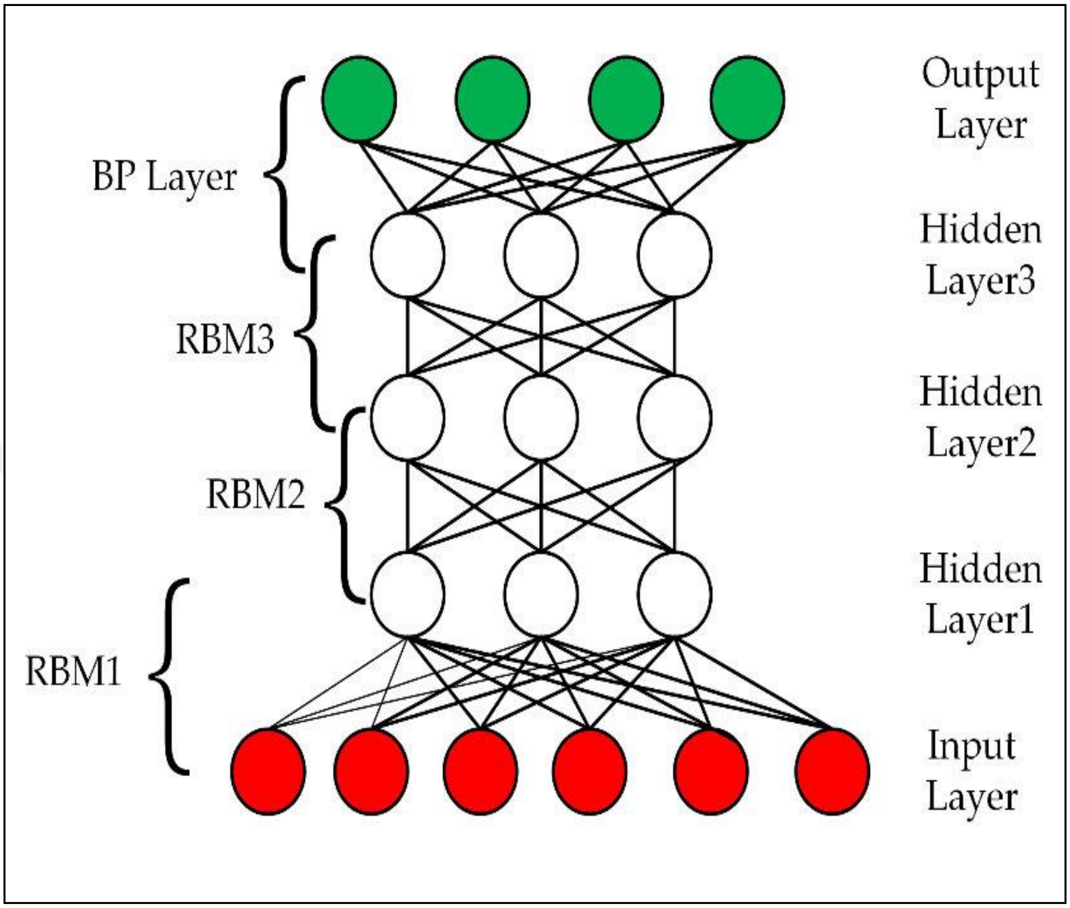


Figure 3.10: Architecture of deep belief network.

3.5.5 Convolutional Neural Networks

The ability of the networks multi -layered neural in learning, starting from adults sets of examples, functions complex, non -linear and high dimensionality, makes them perfect candidates for I image recognition tasks. Images digital and filters an image digital can to be considered as a matrix A of dimension M × N values real or discreet. Every value of the matrix is called pixel and I his indices Is also called coordinates: each pixel A (m, n) represents the intensity in position indicated give it indices. It defines itself filter a transformation applied to an image. We can define various types of filters (kernels) based on which region the starting image is necessary to obtain the value of a pixel:

i. Global: to determine the value the output image is required know each pixel of the input image. The filter is therefore representable via a function.

ii. Local: the value of one pixel of the output image depends on the value of the pixels of a sub -region input image. If that sub -region has side equal to L, the value of each pixel of the output image Sara so obtained from a function

iii. Punctual: the value of a pixel depends only on another pixel, usually that which has the same coordinates as the pixel to be calculated. The function in this case becomes.

About filters global and local, a particular type of operation that yes can run to switch between input to output is represented from the convolution.

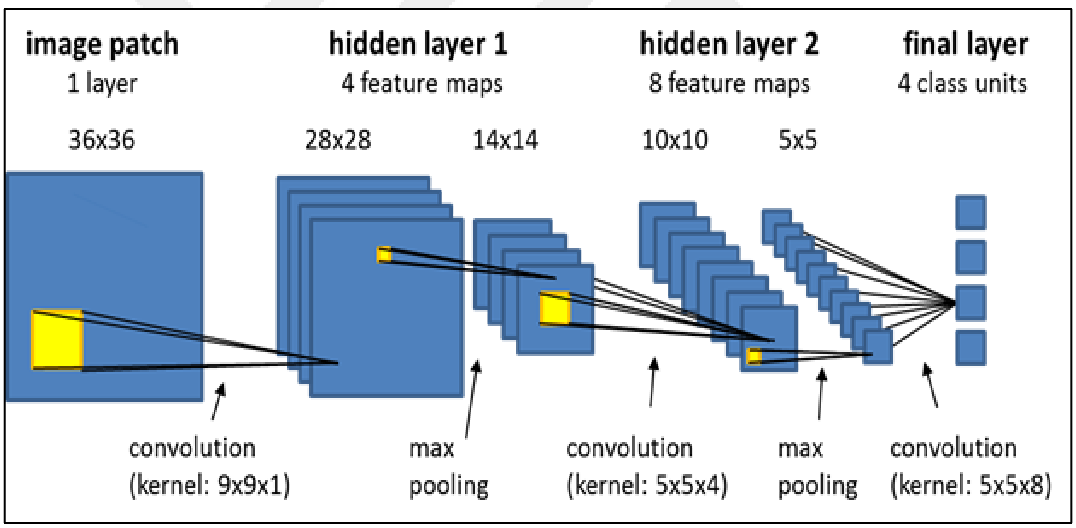
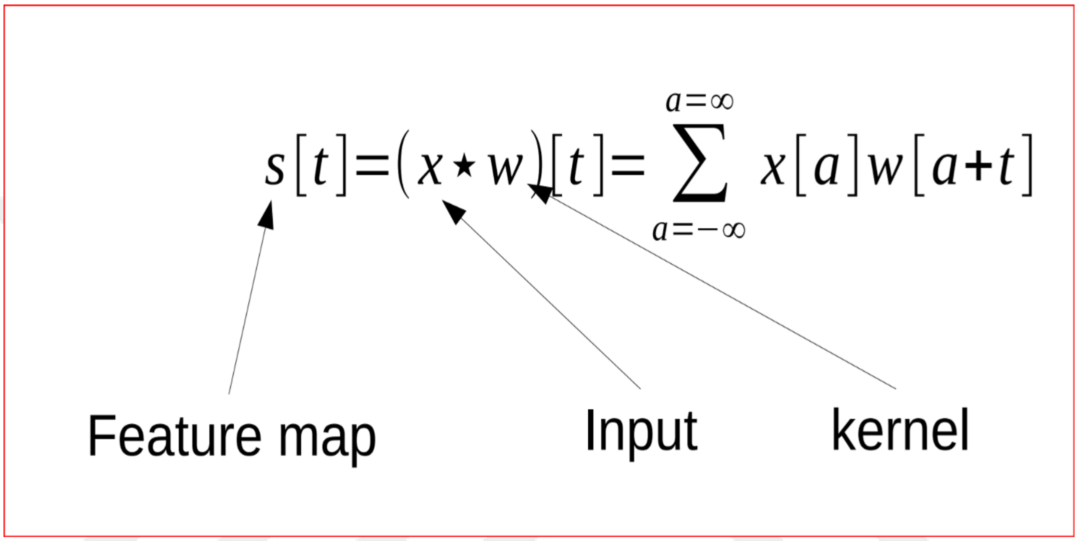


Figure 3.11: Architecture of convolutional neural network.

The convolution, in case yes operate on images digital (convolution discreet), yes can define as: considering the input image, the image filtered and defining the convolution matrix. We get a local filter by choosing for an array of dimensions even at the sub-region which affects the value of a pixel. Each pixel is thus the result of a weighted sum via the matrix of the values of the sub-region which has the pixel coordinates as its center.

It is important note which, why the filters are well defined, it is necessary to consider the extremes the starting image: in fact, the subregioncentered on a point of the edge image touch undefined points. There are two possibilities, to be chosen according to use that yes must make of the image filtered: - extend the image, getting an image filtered having the same size input image - do not consider the undefined frame, get an output image smaller: if the picture input has dimensions and the convolution matrix, the image filtered will have dimensions:



We will see later those convolutional neural networks exploit this second option. Convolutional Neural Networks Online theoretical, a neural network traditional completely connected with dimensions sufficiently large can learn to recognize raw (raw) images without feature extraction. However, with an approach of this like, they would arise of the problems linked to the number of parameters to train (and therefore to the number of training examples needed), which would become unmanageable by many hardware systems. Also, another one defect of the networks neural ordinary is related to the fact that the input has the form of a vector, so only one dimension.

For this reason, if so wants to classify a two - dimensional image, this it will first be " unrolled " by concatenating the rows or columns into a single onevector. In this way the topology input it will come almost completely ignored.

The images However they have a strong local structure:neighboring pixels they usually have a strong correlation space. The Convolutional Neural Networks (CNN), a particular type of networks neural belonging to the deep learning industry, they avoid these problems thank you the use of particular architecturalideas. CNN was created by Yann Lacuna, who he was inspired by the 1980 Fukushima Noncognition.

The difference substantial between convolutional neural networks and networks neural ordinary consists in the done that the former operates directly on images while the latter on features extracted from them. The input of a CNN therefore, unlike that of a neural network ordinary, it will be two- dimensional and the features will be the pixels themselves. it comes represented the basic structure of a CNN.

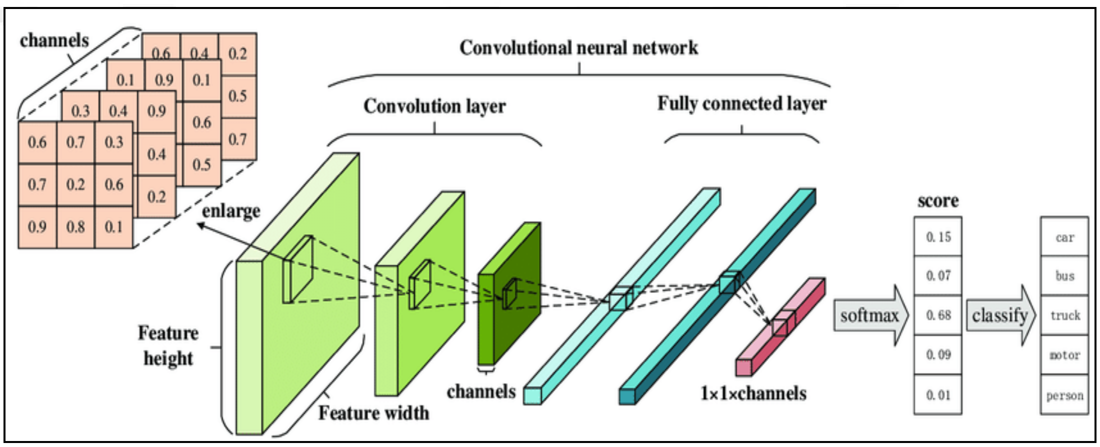


Figure 3.12: Input image into the hidden layers of CNN.

Let's see how they correspond to an input image, inside some hidden, different layers groups of images called feature maps: the feature maps of a layer they are the result of transformations performed on images of the layer previous one through convolution filters or subsampling, too called kernels. The path from the input the output is characterized by an alternation of layers of convolution and subsample and ends with a neural network traditional. In the convolution layers input images are convoluted with several filters. Each kernel therefore produces images, one for each image coming from layer previous.

These images they come added together and the result is one of the feature maps produced come on different filters:these, once passed through a function sigmoid, they will represent the input for the layer next. The units in a convolution layer they will be so organized in planes, the feature maps themselves, inside of the Which all units they share themselves weights, data come on values of the elements of the corresponding kernel. Such weights, initialized randomly, are fix during training of the network via the back-propagation algorithm.

This is how CNN learns independently to extract the most significant features. The units in the same feature map, sharing themselvesweights, they perform the same operation on all the input images. Kernels can be considered as a kind of fields receptive local, which act as detectors elementary features. Thus, forcing the units of a feature map to have themselves values of the weights, they will look for the same " structures " throughout the image kernels are with convoluted. Unit belonging to different feature maps they have instead weights different and seek so different structures.

By means of this structure we have, in each layer, the ability to extract different types of features from each position of the input images. In summary, each unity of a convolution layer receives information as input coming from a set of units located in a small " neighborhood " of the layer previous, whose dimensions are given by size of the kernels. Through the fields receptive locals, the neurons they extract from the scene elementary features (such as edges oriented, angles or endpoints).

These features are then combined from successive layers to form more features complex. The idea of connecting units to fields receptive locals in the input goes back to the Perceptron die early 1960s, coinciding with Hubel and Wiesel 's discovery of neurons with locally sensitive and orientation-selective characteristics in the system visual of the cats. The States identified two basic types of cells: simple cells (S) and complex cells (C). S cells respond to stimuli similar to the edges, inside their fields receptive. C cells are locally invariants exactly position of the stimulus.

A property Interesting of the convolution layers consists in the done that if the input image comes translated, the output of the feature map will be translated of itself quantity but will remain unchanged elsewhere. This property is the basis of the robustness of convolutional networks with respect to translations and distortions input. Once the features come found, the " knowledge " of They position not only becomes less important to identify the pattern (only the position with respect to the other features is relevant), but it is potentially “harmful” since the goal is to look for the same facilities throughout the image.

An easy way to reduce the accuracy with which the position of the different features is registered on a feature map is to reduce the resolution space of itself. This can be done through the so - called subsampling layers, which usually perform local averaging, reducing resolution of the feature map and therefore the sensitivity to translations and distortions. The subsampling operation consists in partition image in a set of non - overlapping rectangles followed from the replacement of each sub - region with the corresponding mean.

For example, if the units in a subsampling layer have size, each of them will return the average of 4 points. Feature maps in a subsampling layer will then have half of the lines and columns of those in the layer previous. Unlike that that happens in the convolution layers, the number of images produced by a subsampling layer is always the same as that of images entering. Finished the layers of convolution and subsampling, the feature maps of final layer are " unrolled " in vectors and entrusted to a neural network traditional that performs the final classification.

4. PROPOSED METHOD

4.1 METHOD OUTLINE

Breast cancer is a disease caused by genetic mutations that lead to abnormal and disordered growth of breast tissue cells. Despite affecting men and women, it is 100 times more common in female patients, being the most frequent form of cancer in this population [1]. In Brazil, breast cancer represents 29% of new cancer diagnoses registered in women annually, with an estimated 59,700 new cases for 2019 [2].

In 2015, the post-diagnosis mortality rate in Brazil was estimated at 13.68 deaths/100,000 women [2]. It is unanimous among health professionals that it is vital for the positive prognosis of breast cancer to have an early and accurate diagnosis [3], with a drastic reduction of deaths for patients with healthy lifestyle habits and who undergo annual exams [2]. Convolutional networks have been heavily used to solve computer vision tasks such as image classification, object detection and semantic segmentation [4].

In recent years, the state of the art in these applications has been marked by the use of convolutional networks, such as Alex Net [5], Vignette [6] and Reset [7], significantly increasing the maximum sensitivity level reached by the algorithms. The use of convolutional networks has also advanced the development of tools to aid medical diagnosis (CAD). These tools have shown promising results in several areas, such as Alzheimer's [8, 9], lung diseases [10, 11] and in mammograms [12].

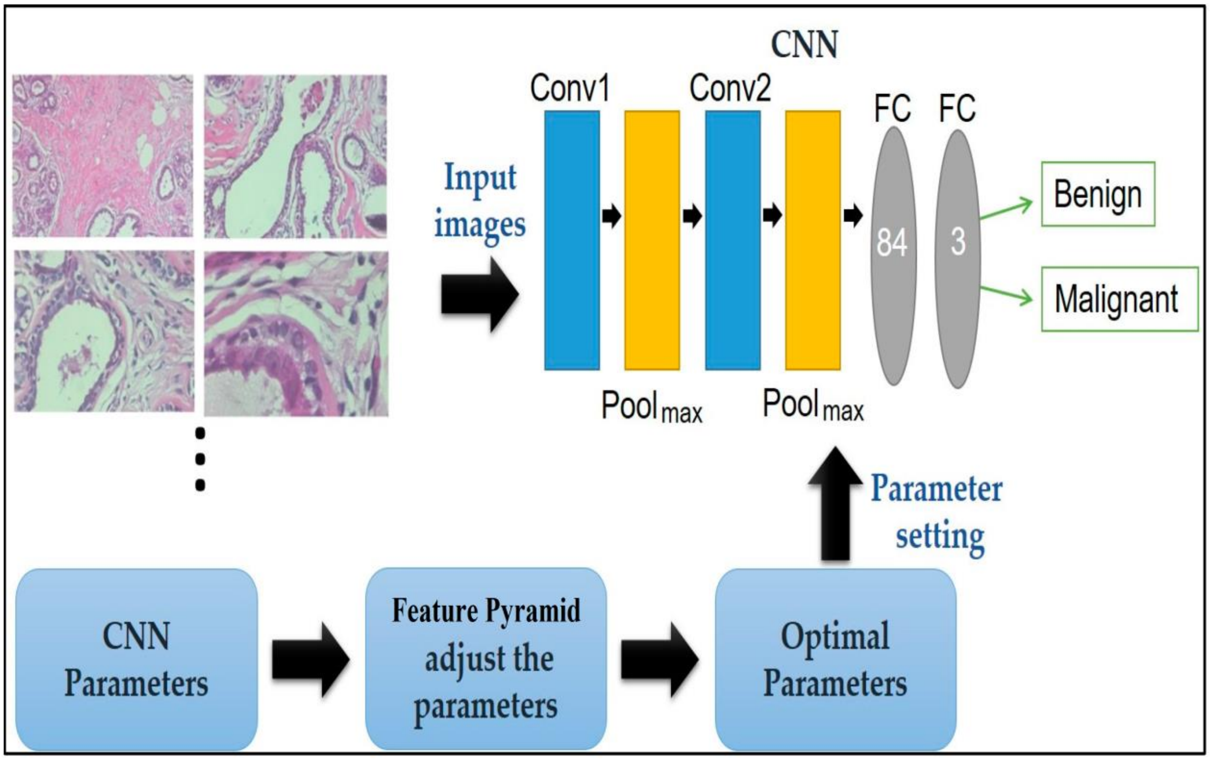


Figure: 4.1: Proposed framework for breast cancer detection.

4.1.1 Proposed Network

This article presents an algorithm for the detection of masses in mammography images, in order to aid medical diagnosis. Our algorithm is inspired by the work of Jung et al. [13], especially in the use of the CNN structure in the detection of objects, however, it differs from this in the following aspects:

i. We use pre- and post-processing steps with image division into regions of fixed size, allowing training with high-resolution images on hardware with limited memory.

ii. Model training and evaluation are performed with public dataset [14].

4.1.2 Feature Pyramid Networks

One of the challenges presents in the area of object detection is the recognition of objects at different scales. The same object can come in different sizes, depending on the relationship between the object and the camera, which makes it difficult to develop algorithms capable of detecting objects in any situation. This problem is even more pronounced in situations where you work with objects of different classes, and different classes naturally contain different sizes. The resulting resolution in a convolutional network may be adequate for evaluating some classes of objects, but inadequate for others. It is therefore interesting to develop detection techniques that allow working with images and objects subject to scale variations.

Feature Image Pyramid [36] is the basis of current solutions, being widely used both with detectors based on descriptors and convolutional networks. This method consists of making predictions for different scales of the original image, where each level of the pyramid is analyzed independently of the others. This ensures that each level of the pyramid maintains a high level of information, but at different scales. The main disadvantage of this method is the need to perform all the processing for each level of the pyramid, significantly increasing memory usage and processing time.

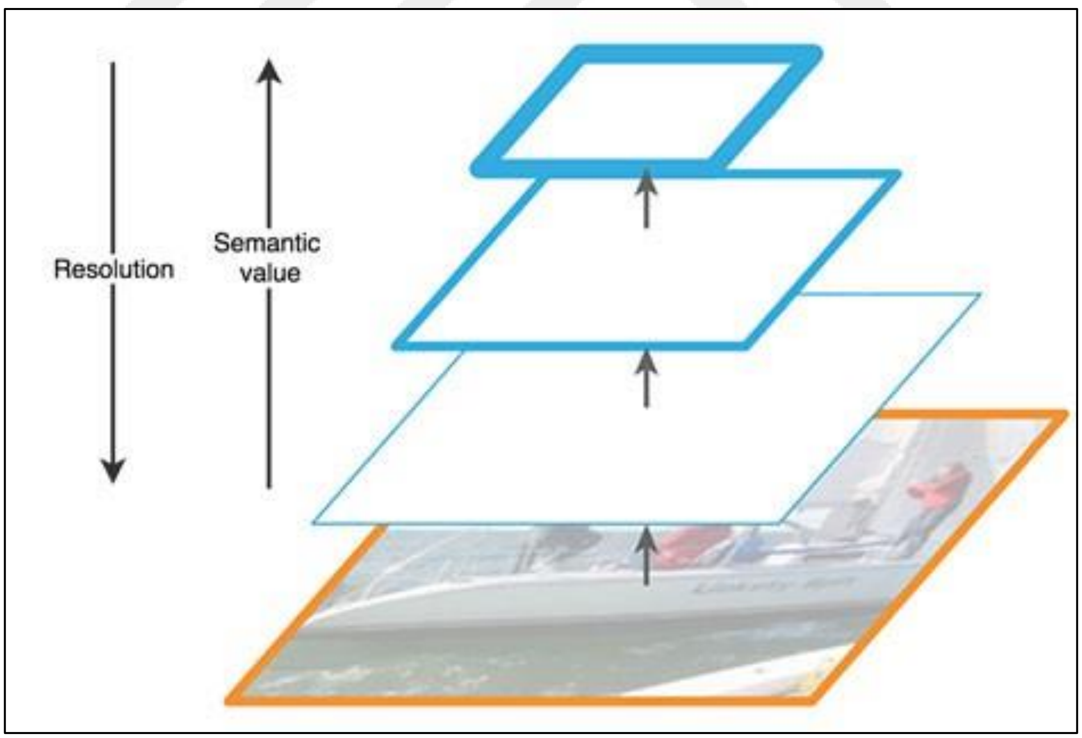


Figure: 4.2: Feature pyramid network.

in convolutional networks, where detection is done only in the last activation layer of the convolutional network. Originally used in networks such as [37, 6], this method obtained interesting results, but the inclusion of multi-scale techniques results in more accurate structures. Pyramidal Feature Hierarchy promises to solve the multiscale and computational cost problem, using layers already naturally produced by the convolutional network for prediction. Used by networks such as SSD [38], this technique mainly fails in the detection of small objects, since the layers with higher resolution are not processed enough to allow the detection of objects. Finally, we have the Feature Pyramid Network (FPN) [36] structure that allows us to produce layers with different scales, all of which contain high-level information capable of enabling the detection of objects.

4.2 BREAST CANCER SEGMENTATION

Segmentation is a ubiquitous problem in image processing and computer vision whose goal is to isolate an object from the background of the image. As part of the design of a Cad system, segmentation plays a central role in the sense that the other steps, namely the description of the mass, the detection of structures of interest such as spicules and the classification of the different lesions strongly depend on it. Thus, the more precise the segmentation, the better the description of the nature of the contour, the texture and the shape of the mass, which will result in a good classification of the tumor. The particularity of segmentation lies in the fact that there is not one method suitable for all applications.

As such, it has been the subject of several publications, in which researchers have tried to provide various and varied solutions that often take into account the nature of the images processed. In general, segmentation methods are grouped into two categories: supervised and unsupervised approaches. Unsupervised approaches partition the image into several distinct regions that satisfy properties related to pixel gray level and texture.

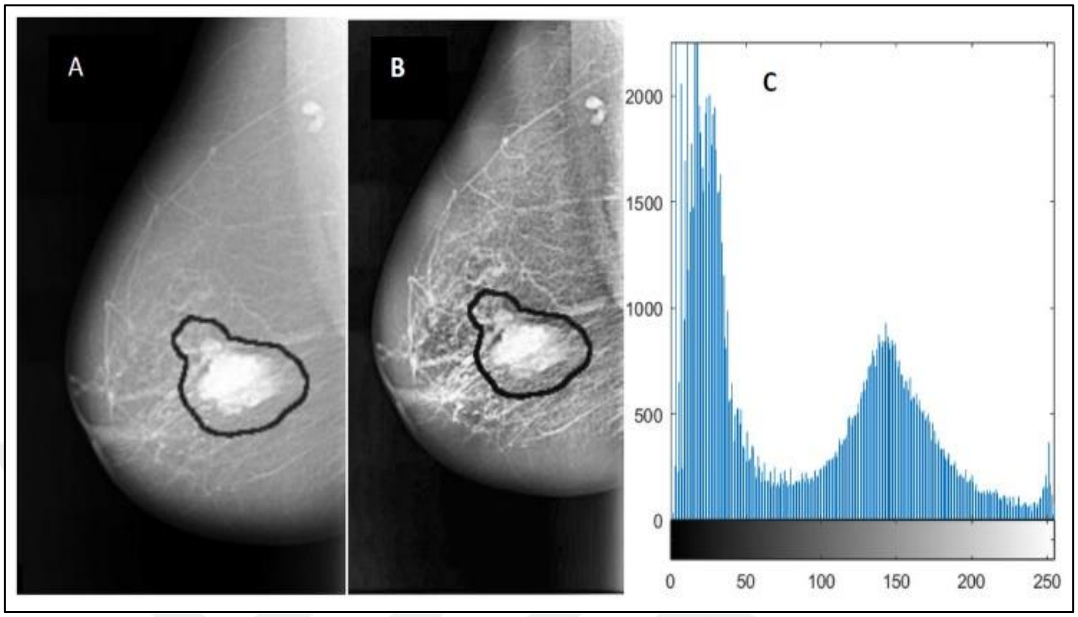


Figure: 4.3: Segmented Breast cancer.

Among the continuous formulations, most effective are those which minimize functionals of active curves by the means of the level sets through a descent of gradient. However, since the gradient descent converges to a local minimum, the results obtained are not always convincing. Added to this is the fact that these methods require a fairly large calculation time without giving the assurance of having a correct result, i.e., tend towards a local minimum close to the global optimum. To solve this problem, a discrete formulation – which conceives the image as functions disc.

4.3 EXPERIMENTAL RESULTS

This chapter will present the database used for training and testing the proposed approaches and will also present all the results obtained at different stages of the process, showing the evolution obtained from the implementation of some techniques presented previously. Finally, an analysis of these results and a comparison with the results obtained by other similar works will be made in order to validate the proposed approaches.

4.3.1Database

In order to train the classifiers and validate the results, a database consolidated in the literature is required. Therefore, we analyzed the databases most used in studies with similar objectives, which were the MIAS (Mammographic Image Analysis Society) and DDSM

(Digital Database for Screening Mammography) The MIAS base was chosen because the DDSM base has more than a thousand images, with larger dimensions, which would require a higher computational cost, which the resources of this work were not able to meet.

Table 4.1: MIAS dataset structure.

Column	Detail
1st	MIAS database reference number
2nd	Character of background tissue: F – Fatty G – Fatty-glandular D – Dense-glandular
3rd	Class of abnormality present: 1 CALC – Calcification 2 CIRC – Well-defined/circumscribed masses 3 SPIC – Spiculated masses 4 MISC – Other, ill-defined masses 5 ARCH – Architectural distortion 6 ASYM – Asymmetry 7 NORM E Normal
4th	Severity of abnormality; B – Benign M E Malignant
5th, 6th	X y image-coordinates of centre of abnormality.
7th	Approximate radius (in pixels) of a circle enclosing the abnormality

This database also provides an annotation with some information about each image, such as the breast tissue, the presence of cancer, type and class of cancer, and its position on the image, indicating its center and radius. Of the 322 images, 209 are from patients who do not have cancer, and 113 have cancer. Of the 113 images with cancer, 62 are benign tumors and 51 are malignant.

Table 4.2: Structure of the DDSM dataset.

Potential Advantages of DM	Comments
Eliminates artifacts and noise related to film processing plus promises an ability to deliver a consistent image quality	Also artifacts with DM, though these may be less serious, e.g. Boyle [A] - motion artifact seen on slot-scanning DM, using a phantom; brief motion caused degradation of only a small part of the image; continuous motion produced smearing with DM and FSM -- they consider these results reassuring
Wider dynamic range	
Greater contrast resolution	
Able to process image to better depict abnormal or suspicious findings	
Potentially lower radiation dose for those being examined	Overall safety advantage over FSM (may not be significant)
Possibly shorter examination times, with reduced discomfort for those being examined	
Elimination of film library costs	Transfer and storage of digital mammograms is not cost-free
Elimination of lost films	Potential for loss, even with electronic archiving
More suitable for computer aided detection and diagnosis (CAD) programs	CAD programs still a developing area; their place in routine health care does not appear to be established as yet
Expected useful role in stereotactic biopsy procedures	
May be used in teleradiology	Staffing and other factors would need resolution before tele-mammography would be practical

4.3.2 Implementation Details

To implement the CNNs, the MatConvNettoolboxdeveloped for MATLAB, which implements CNN networks for computer vision applications in a simple and efficient way, was used. Many pre-trained CNNs for image classification, segmentation, face recognition, and text detection are made available for the toolbox

In addition, the configuration of the machine used in the experiments was: (i) Windows 10 operating system; (ii) Intel Core i7-5500U processor, 2.40GHz with 2 physical cores; (iii) 12 GB DDR4 RAM memory; (iv) 1000 GB storage unit (hard disk); (v) AMD Radeon HD R7 M265 video card, with 2 GB dedicated memory.

Next, the values of the hyperparameters used in the neural networks for each proposed approach are presented.

i. Semi-automatic approach. For training the initial CNN, a learning rate of 0, 0001 was used and 30 iterations were performed. For training the fully connected layers, a learning rate of 0, 01 and 100 iterations were used.

ii. Automatic approach. For training the initial CNN, a learning rate of 0, 0001 was used and 30 iterations were performed. For training the fully connected layers, a learning rate of 0, 0001 and 1000 iterations were used.

iii. Regarding the use of the MIAS database.

iv. Semi-automatic approach. The database was divided into training, validation, and testing, in the proportions of 70%, 20%, and 10%, respectively.

v. Automatic approach. The data set was initially reduced by randomly eliminating some normal ROIs in order to make the quantities more proportional, since the number of normal ROIs was much higher than the number of cancer ones, which could impair the training.

Finally, this group of data was divided into training, validation and testing in respective proportions of 70%, 20% and 10%.

Table 4.3: CNN layers and size.

Layer	Filter	Stride	Output map size	Activation
Convolution1	3×3	1	$256 \times 256\times 32$	ReLU
Pooling1	2×2	2	$128 \times 128\times 32$	-
Convolution2	3×3	1	$128 \times 128\times 64$	ReLU
Pooling2	2×2	2	$64 \times 64\times 64$	-
Convolution3	3×3	1	$64 \times 64\times 128$	ReLU
Pooling3	2×2	2	$32 \times 32\times 128$	-
Convolution4	3×3	1	$32 \times 32\times 256$	ReLU
Pooling4	2×2	2	$16 \times 16\times 256$	-
FullConnected1	-	-	$1 \times 1 times\ 512$	ReLU
FullConnected2	-	-	$1 times 1\times 2$	Softmax

4.3.3 Metrics of Evaluation

Specifically, the problem of cancer detection and classification using mammography images consists of two dependent binary classification tasks. The two binary tasks to be performed are (I) cancer detection, which is the classification of the ROIs extracted from the mammography image as normal or cancerous. (ii) cancer classification, where each ROI already detected as cancerous is classified as benign or malignant.

Considering that both cancer detection and classification are modeled via a binary classification problem, the evaluation used here was done according to the following metrics: accuracy (*Ac*), which is the proportion of total hits over total data; sensitivity (*Se*), which defines the classifier's ability to retrieve the largest number of true positives correctly; and precision (*Es*), which is the classifier's ability to detect true negatives.

The relationships that define these metrics are presented in Equations (4.1) where:VP are true positives,VN are true negatives,FP are false positives, andFN are false negatives.

After finding which models had the lowest validation loss, we selected those models to classify the test set samples, and then examined how effectively those models could categorize the data. A confusion matrix is utilized for displaying and assessing the outcomes of classifications done by each model. The values assigned to each class in a multiclass confusion matrix are as follows: TP The actual class of the goods was precisely predicted to be their intended class. A "true negative" occurs when the class of an object can be reliably predicted to belong to something else. True Positives are referred to as items for which the class prediction is accurate (TP). A result has a false negative (FN) if it incorrectly predicts a class other than the one that really exists. In the last phase, we will construct common metrics for each class of anomaly detection and then use the matrix values to choose which neural network is best suited for each class.

The term "accuracy" refers to the ratio of accurate predictions to the total number of predictions. Precision ratio refers to the proportion of genuine positives relative to the overall number of positive outcomes. Consider the proportion of samples with accurate predictions, the harmonic means of recall and precision for the F1 score, as well as the F1 score [33]:

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

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

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

$$F_1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

4.3.4 Evaluation of The Semi-Automatic Approach

The semi-automatic approach presented in Chapter 3 was evaluated solely on the cancer detection problem, showing the evolution obtained from the different techniques introduced in the work.

i. Mean standard Deviation

The mean must be determined over all images, taking their height and breadth into consideration, but not across channels. When working with 3-by-3-pixel-colored images, an output tensor should be present. The formula used to compute the standard deviation is as follows [34]:

$$\sigma = \sqrt{E[X^2] - (E[X])^2}$$

ii. AFC measure test

The Aggregated Channel Features (ACF) detector introduced in [15] was constructed by connecting three channels of the LUV color space with a normalized gradient channel and six channels of the histogram-oriented gradient channel (HoG). An ACF detector's proposed region extraction technique is responsible for differentiating between positive and negative regions. Training data consisting of the bounding box of the ground truth area are used to generate positive proposal regions. In contrast, the sliding window will automatically delete the region of the negative proposition from all picture areas, excluding the bounding box of the ground truth area. I(x,y) represents a mxn-pixel RGB image with three channels. The following formula is used to convert the image to the LUV color space prior to calculating the gradient magnitude of the data [34].

$$M(i,j) = \sqrt{\left(\frac{\partial I(i,j)}{\partial x}\right)^2 + \left(\frac{\partial I(i,j)}{\partial y}\right)^2}$$

iii. Skewness

The removal of outliers is a crucial component of the data cleansing process, which follows the exploratory data analysis step. However, when removing outliers using approaches such as the standard deviation method and the interquartile range method, you should proceed with caution. Given that the median of this skewed data set is zero, you will likely exclude the bulk of integers that are not outliers. This is due to the median being 0 The calculation of outliers was improperly performed: Using the Standard Deviation Method, the proportion is calculated to be 96.5 percent. (This technique treats the majority of data as outliers) [34]:

$$\text{Skewness} = \frac{[(N-1)(N-2)]}{N} \sum_{j=1}^{N} \frac{(x_j - \bar{\bar{x}})^3}{\sigma_x^3}$$

4.3.5 Evaluation of The Automatic Approach

The evaluation of the automatic approach presented in Chapter 3 was performed on three criteria: its ability to generate valid ROIs, its performance on the cancer detection problem, showing its ability to classify ROIs as normal or cancerous, and finally its performance on the cancer classification problem, showing its ability to classify cancerous ROIs as benign or malignant tumors.

i. Segmentation evaluation for ROI generation

The main objective of this step was to generate the largest number of ROIs, implying the need to perform an over-segmentation in order to obtain the highest possible sensitivity.

To validate the results, a pixel-by-pixel evaluation of each image was performed. As this step used an unsupervised technique, the results were obtained over the entire database.

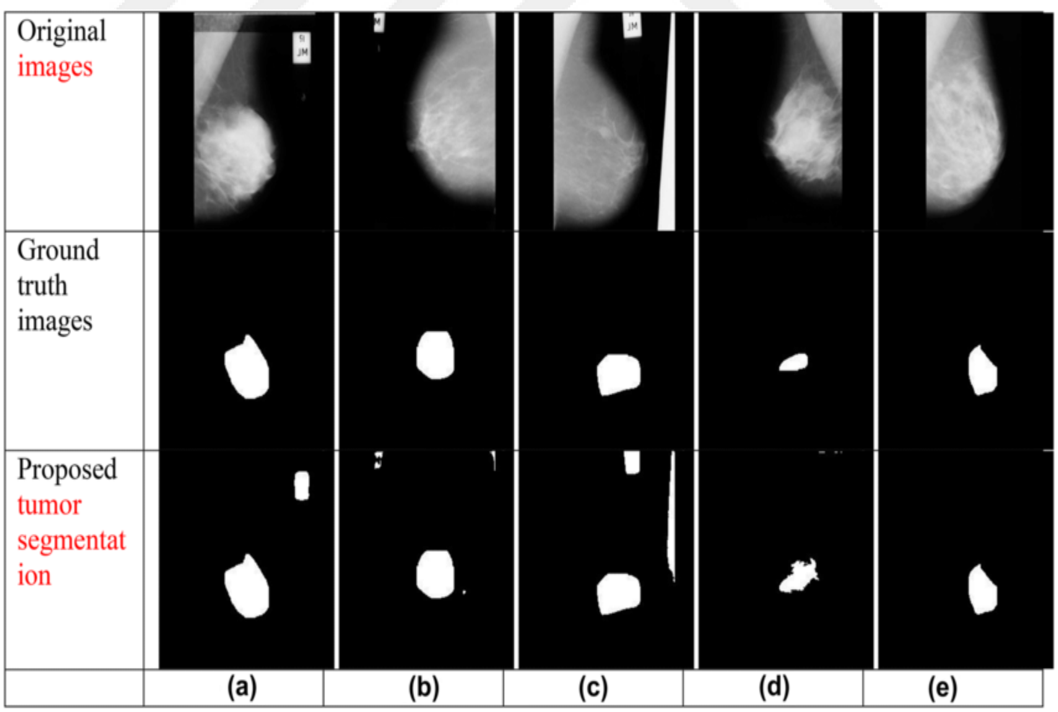


Figure: 4.4: original image after segmentation compared with the ground truth (GT).

ii. Evaluation of Classification in Cancer Detection

As presented in Chapter 3, in classifying the automatically detected ROI as normal or cancerous, three different CNN networks were tested. It can be seen that the CNN AlexNet network performs best, with an accuracy of 91, 89%, followed by the CNN VGGNet network, which also obtained results close to AlexNet, with an accuracy of 95, 93% in its two configurations. The simple CNN obtained much lower results than the others, showing a higher efficiency of the pre-trained models that are already widely used in the literature, besides being much more complex structures.

Table 4.4: Comparison of CNN alexnet vs VGGnet and inception.

Comparison					
Network	Year	Salient Feature	top5 accuracy	Parameters	FLOP
AlexNet	2012	Deeper	84.70%	62M	1.5B
VGGNet	2014	Fixed-size kernels	92.30%	138M	19.6B
Inception	2014	Wider - Parallel kernels	93.30%	6.4M	2B
ResNet-152	2015	Shortcut connections	95.51%	60.3M	11B

The results obtained by this work are considered good, if compared to other similar works present in the literature These results do not reach the best results in the literature, but it is important to emphasize that the other works used the regions demarcated by the database, unlike the complete methodology developed in this work, which performs the automatic segmentation of the regions of interest, increasing the complexity of the problem.

Analysing the wrongly classified ROI, it is possible to identify some difficulties of the implemented methodology. One of the main limitations of this methodology, and the main cause of errors in the classification of ROI as cancerous or not, is the lack of precision in the segmentation step, because in most cases the real ROI is not found in the segmentation step.

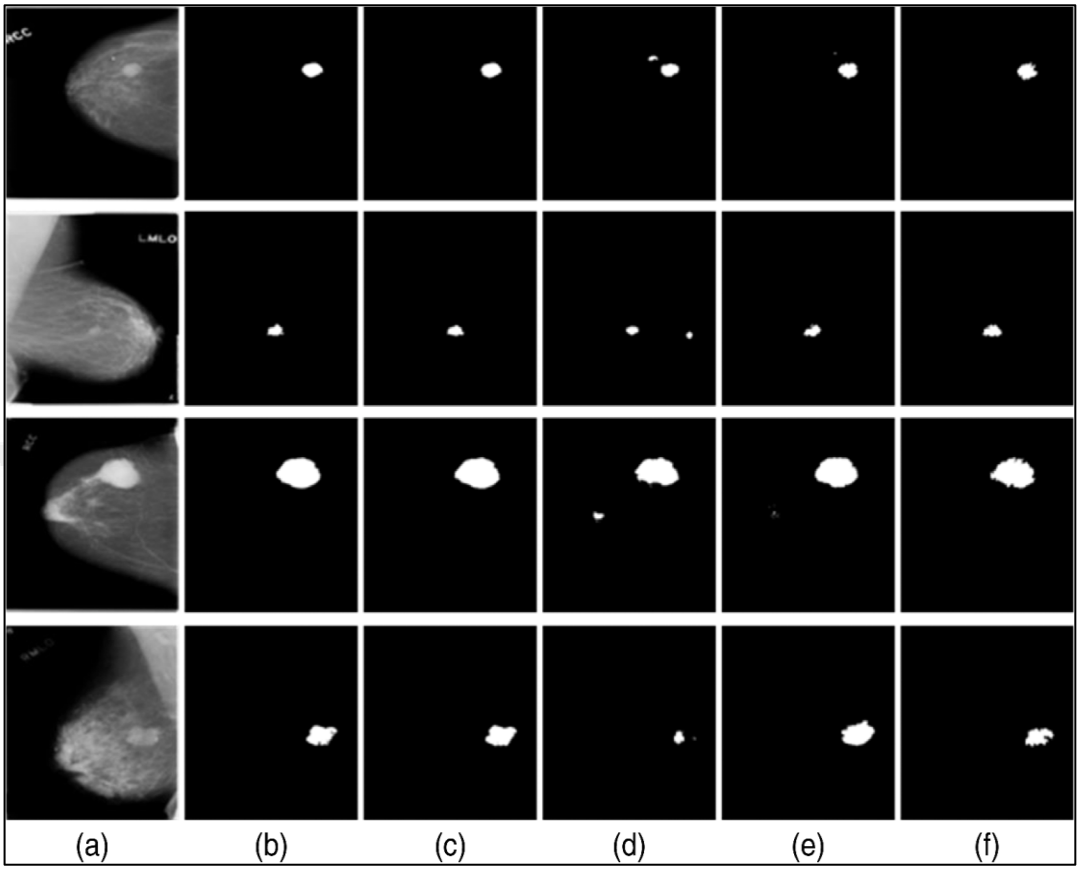


Figure: 4.5: Final segment output from the dataset.

Centralized in the detected ROI, and may be displaced, reduced or enlarged. In Figure 41 it is possible to observe an example of failure in the detection caused mainly because the segmented ROI are not centered in the tumor, besides being affected by another difficulty even more difficult to be circumvented, which is the identification of tumors in images of denser breasts, where the tumor ends up mixing with the breast region and thus hinders the identification of the tumor contour which is of utmost importance in this step of detection. In Figure 42 it is possible to observe an example of an image with denser regions in the breast that get confused with tumors, being wrongly detected.

iii. Cancer Classification

As presented in Chapter 3, in classifying as benign or malignant the ROIs detected as cancerous, three distinct CNN networks were tested. It can be seen that, the results obtained by the pre-trained convolutional neural networks again are superior to the simple CNN. However, this time, VGG obtained equivalent results to AlexNet, where both achieved an accuracy of 82.14%. Analyzing the wrongly classified ROI in this step, it was more difficult to find a main cause for these errors, because the characteristics of benign and malignant tumors do not have such a clear separation, making this task very difficult. In Figure 43 it is possible to observe an example of two images of very similar tumors, where both were classified by the system as benign, but one of them is malignant and was misclassified. In Figure 44 the opposite occurred, the two tumors present in the two images were classified as malignant, but one of them is benign.

Along the obtained results, we might conclude that the program has the ability to detect the existing attack stream through the statistical measurements implemented throughout the instruction codes. Now, through applying CNN algorithm on the same packet stream, the MSE, success rate, as well as the losses rate have been obtained and presented for both trained and tested CNN sets. The tried program code will illustrate the classification activity of the Convolutional Neural Network (CNN) algorithm upon the information packet samples (N=2000 samples) to be contrasted with the preparation original examination as far as the successions rates also the MSE standards. The obtained results of the above code2 program have been extricated and exhibited in Figures 4.8-4.11down.

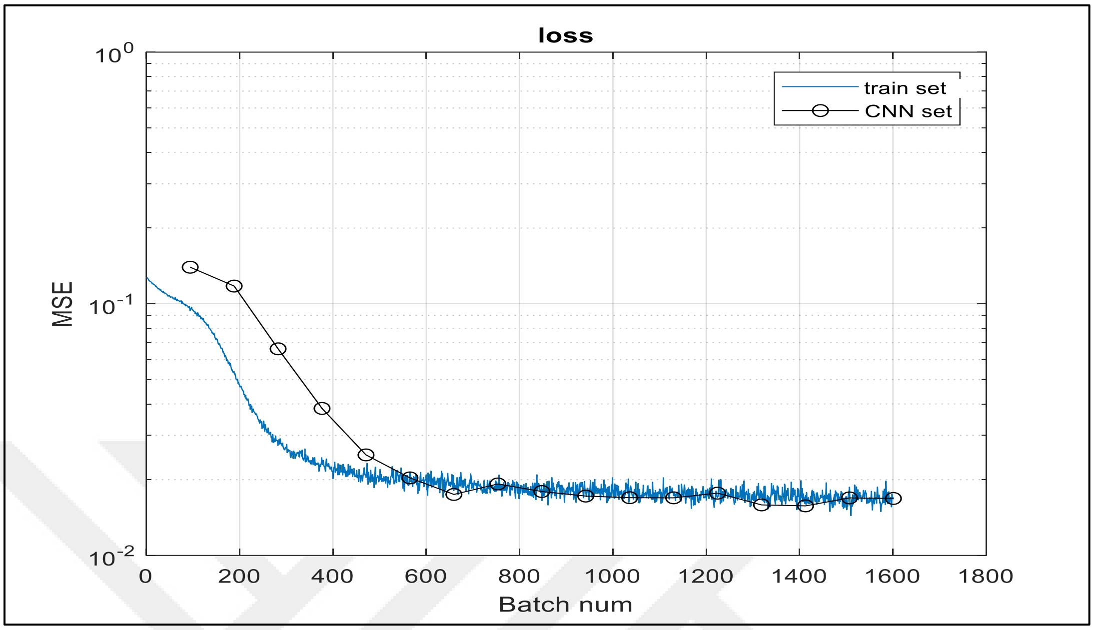


Figure: 4.6: MSE of tested and CNN sets comparison.

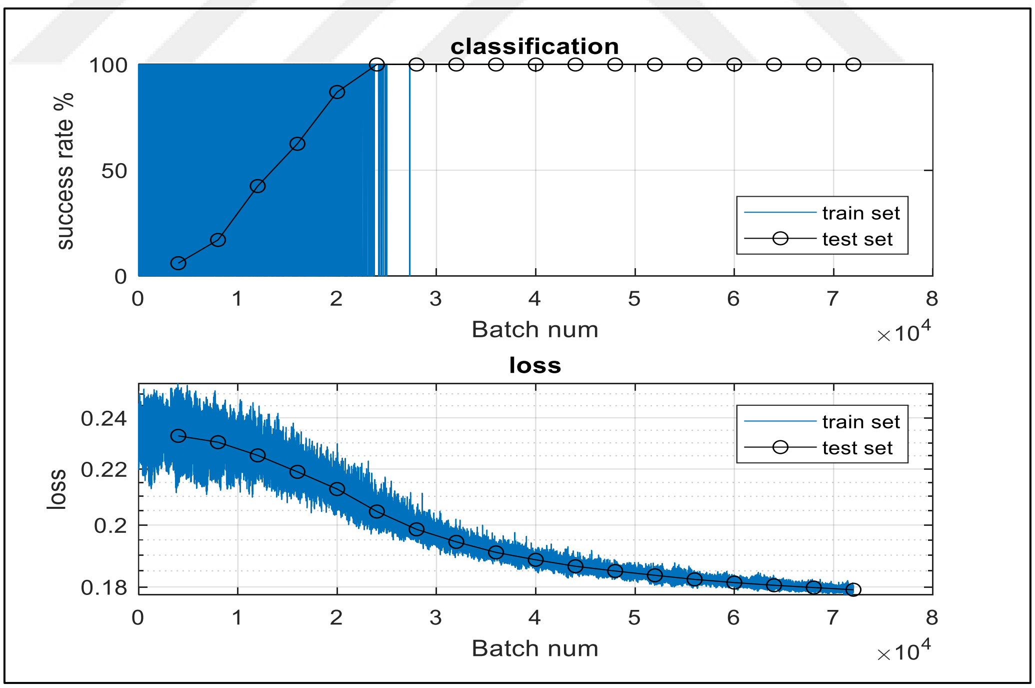


Figure: 4.7: Success rate and loss utilizing CNN algorithm test as compared against the tested info (Batch number=8*10^4).

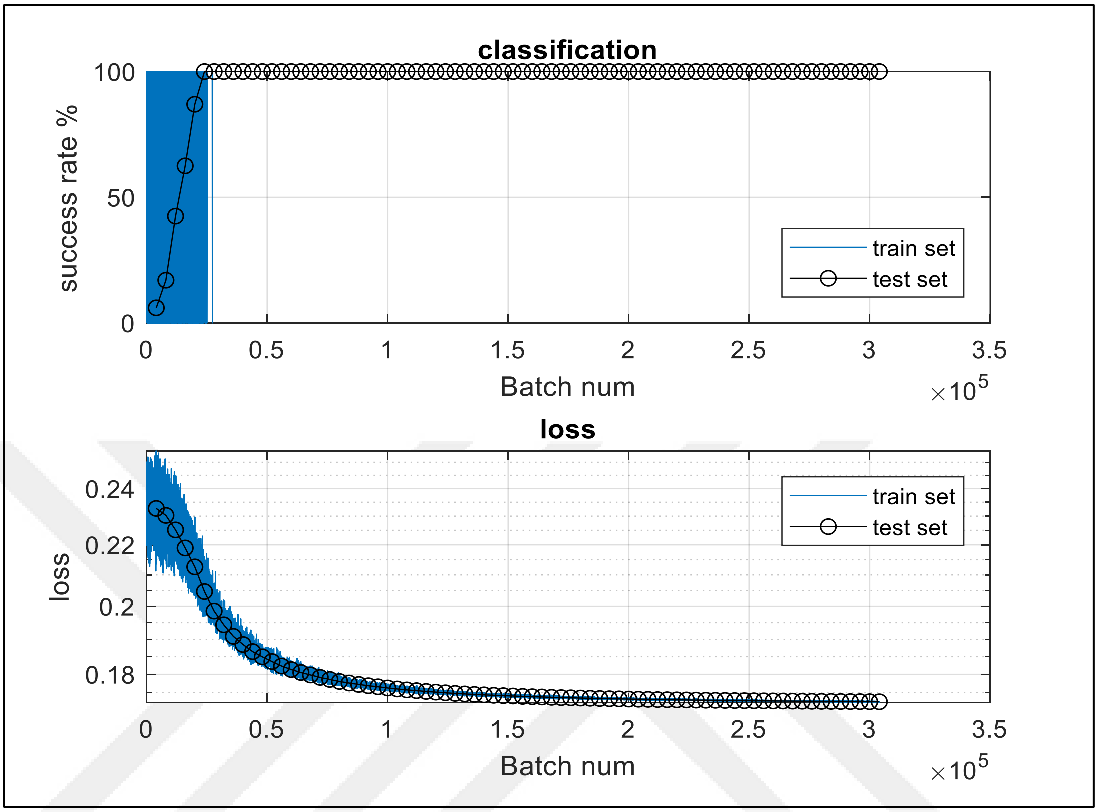


Figure: 4.8: Success rate and loss utilizing CNN algorithm test as compared against the tested info (Batch num=3.5*10^5).

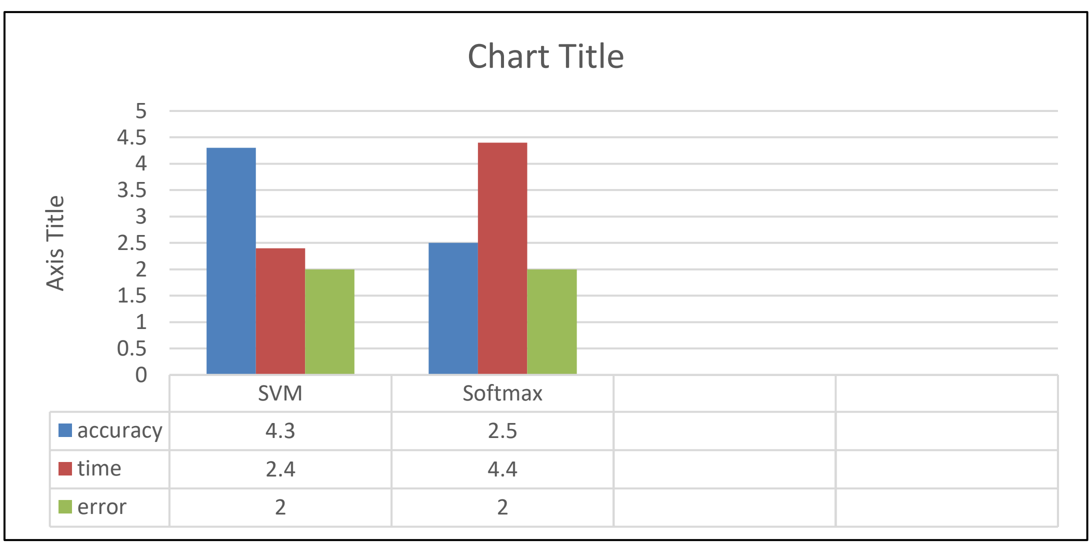


Figure: 4.9: Comparison of CNN SoftMax Function with SVM Classification.

Table 4.5: Learning rate and accuracy of the training with data increase.

Learning rate	Batch size (train)	Loss (train)	Loss (val)	Accuracy (val)
0.01	32	0.2827	0.4419	79.55
0.001	32	0.3113	0.4420	80.20
0.0001	32	0.3224	0.4340	81.60
0.00001	32	0.3327	0.4545	82.10
0.01	**64**	**0.0034**	**0.0002**	**99.80**
0.001	64	0.4550	0.4763	80
0.0001	64	0.6728	0.5859	69.00
0.00001	64	0.5850	0.5956	70.39

4.4 RESULTS DISCUSSION AND SUMMARY

In this chapter the results of the two proposed approaches were presented, where it was possible to observe a good performance of the semi-automatic approach due to a greater simplicity of the problem. In the automatic approach, different CNNs were evaluated in order to achieve better results. Finally, the automatic approach achieved reasonable results, given the complexity of the problem, since other works that presented superior results used the semi-automatic approach.

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This work sought to develop an effective methodology for the objective of performing image classification of digital mammograms. For this purpose, two main methodologies were developed, being a simpler one, developed at an initial stage, and a more complete one with a greater future applicability.

The first methodology, simpler, sought to classify the images only in cancerous or normal, and used the information provided by the database to identify the regions of interest. This way, due to a greater simplicity of the problem, it was able to achieve great results, reaching 99.41% of accuracy, 98.57% of sensitivity and 100% of precision, values above other works with similar objectives. Also in this methodology, techniques that proved to be very effective for this application were evaluated.

Already in the pre-processing stage, the *feature pyramid*techniquewas implemented, which achieved an increase of approximately 4% accuracy in the final result. And later, the *data augmentation* technique was applied to the CNN training, which was responsible for an increase of more than 2% in accuracy. Despite the significant results obtained in the first methodology, it becomes inapplicable since it uses information provided manually in the database to identify the region of the tumors. Therefore, to eliminate this limitation, a more complete methodology was developed, which sought to perform the image segmentation, obtaining candidate regions so that this way there can be applicability for the work developed.

In the first stage, an over-segmentation was performed focusing on sensitivity, since it would be performed a stage of classification of the regions between normal and cancerous. The result obtained was considered good, since it reached a sensitivity of 94.73%, indicating that the over-segmentation method recovered almost all the regions with cancer. The analysis of the distribution of gray levels and texture features in the cancer regions had a great influence on the good results achieved. In the classification step between normal and cancerous, 3 different CNNs were evaluated.

The results obtained were considerable given the difficulty of the problem. The best results were obtained by CNN AlexNet, with an accuracy of 91.89%, sensitivity of 88.52%, and precision of 96%. The other CNNs achieved the following results: 85.58%, 87.75% and 83.87% for the 19-layer VGG, and 81.08%, 80.70% and 81.48% for the simple CNN, these being values for accuracy, sensitivity and precision, respectively. In the classification step between benign and malignant, the same method was used. topology developed for the classification step between normal and cancerous.

The results were lower than the previous step due to a greater difficulty, but the results obtained are considerable, with the best ones being obtained again by the pre-trained networks: AlexNet with an accuracy of 82.14%, sensitivity of 72.41% and precision of 92.59%, and VGG with 16 layers achieving the same accuracy of 82.14%, sensitivity of 81.48% and precision of 82.75%. Therefore, in general, the results obtained were very good given the difficulties encountered, such as a base with little data, with only 330 images, not being feasible to use larger bases due to lack of computational capacity. Another limitation is the information used for classification, which are only images in gray levels, limiting the ability to solve the problem, perhaps more information about the patient could facilitate the task.

5.2 FUTURE WORK

Although good results have been obtained, the methodology developed can still be improved, evaluating the use of other techniques or even improving the techniques used. This way, there are many possibilities for future works that seek to evolve this methodology. In the pre-processing and segmentation stage, other more efficient techniques can be evaluated to avoid the high cost required for the over-segmentation technique presented in this work. In the feature extraction and classification stage, other CNNs can be evaluated, such as GoogleNet, for example. Another possibility for future work is the implementation of the methodology in a real system, performing a period of testing with the help of experts to validate the results.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2016,” CA Cancer J Clin, vol. 66, pp.7-30, 2016.

[2] Poorolajal, J., et al., Breast cancer screening (BCS) chart: a basic and preliminary model for making screening mammography more productive and efficient. J Public Health (Oxf), 2017: p. 1-8.

[3] Z. Jiao, X. GAO, Y. Wang, and J. Li, “A deep feature based frameworkfor breast masses classification,” Journal of Neurocomputing, vol. 197, pp. 221-231, 2016.

[4] J. Arevalo, F. A. González, R. Ramos-Pollán, J. Oliveira, and M. A.Lopez, “Representation learning for mammography mass lesion classification with convolutional neural networks,” Comput. Methods Program Biomed, vol. 127, pp. 248-257, 2016.

[5] Deng, L., Yu, D.: Deep Learning: Methods and Applications. Now publishers, Boston (2014).

[6] Krizhevsky, A., I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems. 2012.

[7] Zhang, Y., et al. DeepSplice: Deep classification of novel splice junctions revealed by RNA-seq. in Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on. 2016. IEEE

[8] Hua, K.-L., et al., Computer-aided classification of lung nodules on computed tomography images via deep learning technique. OncoTargets and therapy, 2015.

[9] Pravin S. Mane, Indra Gupta, M. K. Vasantha, “Implementation of RISC Processor on FPGA”, Electrical Engineering Department, Indian Institute of Technology Roorkee, 2006 IEEE.

[10] Kui YI, Yue-Hua DING, “32-bit RISC CPU Based on MIPS”, International Joint Conference on Artificial Intelligence 2019.

[11] Mrs. Rupali S. Balpande, Mrs. Rashmi S. Keote, “Design of FPGA based Instruction Fetch and Decode Module of 32-bit RISC (MIPS) Processor”, International Conference on Communication Systems and Network echnologies, 2011.

[12] Mr. S. P. Ritpurkar, Prof. M. N. Thakare, Prof. G. D. Korde, “Synthesis and Simulation of a 32Bit MIPS RISC Processor using VHDL”, IEEE International Conference on Advances in Engineering & Technology Research, 2014.

[13] David A. Patterson, John L. Hennessy, “Computer Organization and Design”, the Hardware and Software Interface, Third Edition.

[14] J. C. Tobias Charistian Cahoon, Melanie A.Sutton, “Three-class mam mogram classification based on descriptive cnn features,” 2020.

[15] S. Sharma, M. Kharbanda, and G. Kaushal, “Brain tumor and breast cancer detection using medical images,” International Journal of Engi-neering Technology Science and Research, vol. 2, 2015.

[16] P. Hankare, K. Shah, D. Nair, and D. Nair, “Breast cancer detection using thermography,” Int. Res. J. Eng. Technol, vol. 4, 2016.

[17] S. Hwang and H.-E. Kim, “Self-transfer learning for fully weakly supervised object localization,” arXiv preprint arXiv: 1602.01625, 2016.

[18] Z. Mohammadzadeh, R. Safdari, M. Ghazisaeidi, S. Davoodi, and Z. Azadmanjir, “Advances in optimal detection of cancer by image processing; experience with lung and breast cancers,” Asian Pacific journal of cancer prevention: APJCP, vol. 16, no. 14, 2015.

[19] Ding, J.; Cheng, H.-D.; Huang, J.; Liu, J.; Zhang, Y. Breast ultrasound image classification based on multiple-instance learning. J. Digit. Imaging 2012, 25, 620–627. [CrossRef] [PubMed]

[20] Bing, L.; Wang, W. Sparse representation based multi-instance learning for breast ultrasound image classification. Comput. Math. Methods Med. 2017, 2017, 7894705. [CrossRef]

[21] Prabhakar, T.; Poonguzhali, S. Automatic detection and classification of benign and malignant lesions in breast ultrasound images using texture morphological and fractal features. In Proceedings of the 2017 10th Biomedical Engineering International Conference (BMEiCON), Hokkaido, Japan, 31 August–2 September 2017; IEEE: New York, NY, USA, 2017; pp. 1–5.

[22] Zhang, Q.; Suo, J.; Chang, W.; Shi, J.; Chen, M. Dual-modal computer-assisted evaluation of axillary lymph node metastasis in breast cancer patients on both real-time elastography and B-mode ultrasound. Eur. J. Radiol. 2017, 95, 66–74. [CrossRef]

[23] GAO, Y.; Geras, K.J.; Lewin, A.A.; Moy, L. New frontiers: An update on computer-aided diagnosis for breast imaging in the age of artificial intelligence. Am. J. Roentgenol. 2019, 212, 300–307. [CrossRef]

[24] Geras, K.J.; Mann, R.M.; Moy, L. Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives. Radiology 2019, 293, 246–259. [CrossRef]

[25] Fujioka, T.; Mori, M.; Kubota, K.; Oyama, J.; Yamaga, E.; Yashima, Y.; Katsuta, L.; Nomura, K.; Nara, M.; Oda, G. The utility of deep learning in breast ultrasonic imaging: A review. Diagnostics 2020, 10, 1055. [CrossRef] [PubMed]

[26] Zahoor, S.; Lali, I.U.; Javed, K.; Mehmood, W. Breast cancer detection and classification using traditional computer vision techniques: A comprehensive review. Curr. Med. Imaging 2020, 16, 1187–1200. [CrossRef] [PubMed]

[27] Kadry, S.; Rajinikanth, V.; Taniar, D.; Damaševiˇcius, R.; Valencia, X.P.B. Automated segmentation of leukocyte from hematological images—A study using various CNN schemes. J. Supercomput. 2021, 1–21. [CrossRef]

[28] Abayomi-Alli, O.O.; Damaševiˇcius, R.; Misra, S.; Maskeliunas, R.; Abayomi-Alli, A. Malignant skin melanoma detection using ¯ image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. Turk. J. Electr. Eng. Comput. Sci. 2021, 29, 2600–2614. [CrossRef]

[29] Maqsood, S.; Damaševiˇcius, R.; Maskeliunas, R. Hemorrhage detection based on 3d cnn deep learning framework and feature ¯ fusion for evaluating retinal abnormality in diabetic patients. Sensors 2021, 21, 3865. [CrossRef]

[30] Hussain, N.; Kadry, S.; Tariq, U.; Mostafa, R.R.; Choi, J.-I.; Nam, Y. Intelligent Deep Learning and Improved Whale Optimization Algorithm Based Framework for Object Recognition. Hum. Cent. Comput. Inf. Sci. 2021, 11, 34.

[31] Kadry, S.; Parwekar, P.; Damaševiˇcius, R.; Mehmood, A.; Khan, J.A.; Naqvi, S.R.; Khan, M.A. Human gait analysis for osteoarthritis prediction: A framework of deep learning and kernel extreme learning machine. Complex Intell. Syst. 2021, 1–19. [CrossRef]

[32] Dhungel, N.; Carneiro, G.; Bradley, A.P. The automated learning of deep features for breast mass classification from mammograms. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; Springer: Cham, Switzerland, 2016; pp. 106–114.

[33] Alhaisoni, M.; Tariq, U.; Hussain, N.; Majid, A.; Damaševiˇcius, R.; Maskeliunas, R. COVID-19 Case Recognition from Chest CT ¯ Images by Deep Learning, Entropy-Controlled Firefly Optimization, and Parallel Feature Fusion. Sensors 2021, 21, 7286.

[34] Odusami, M.; Maskeliunas, R.; Damaševiˇcius, R.; Krilaviˇcius, T. Analysis of features of alzheimer's disease: Detection of early ¯ stage from functional brain changes in magnetic resonance images using a finetuned resnet18 network. Diagnostics 2021, 11, 1071. [CrossRef]

[35] Nawaz, M.; Nazir, T.; Masood, M.; Mehmood, A.; Mahum, R.; Kadry, S.; Thinnukool,

O. Analysis of Brain MRI Images Using Improved CornerNet Approach. Diagnostics 2021, 11, 1856. [CrossRef]

[36] Farzaneh, N.; Williamson, C.A.; Jiang, C.; Srinivasan, A.; Bapuraj, J.R.; Gryak, J.; Najarian, K.; Soroushmehr, S. Automated segmentation and severity analysis of subdural hematoma for patients with traumatic brain injuries. Diagnostics 2020, 10, 773. [CrossRef] [PubMed]

[37] Meng, L.; Zhang, Q.; Bu, S. Two-Stage Liver and Tumor Segmentation Algorithm Based on Convolutional Neural Network. Diagnostics 2021, 11, 1806. [CrossRef] [PubMed]

[38] Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear recognition based on deep unsupervised active learning. IEEE Sens. J. 2021, 21, 20704–20713. [CrossRef]

[39] Majid, A.; Nam, Y.; Tariq, U.; Roy, S.; Mostafa, R.R.; Sakr, R.H. COVID19 classification using CT images via ensembles of deep learning models. Comput. Mater. Contin. 2021, 69, 319–337. [CrossRef]

[40] Sharif, M.I.; Alhussein, M.; Aurangzeb, K.; Raza, M. A decision support system for multimodal brain tumor classification using deep learning. Complex Intell. Syst. 2021, 1–14. [CrossRef]

[41] Liu, D.; Liu, Y.; Li, S.; Li, W.; Wang, L. Fusion of handcrafted and deep features for medical image classification. J. Phys. Conf. Ser. 2019, 1345, 022052. [CrossRef]

[42] Alinsaif, S.; Lang, J.; Alzheimer's Disease Neuroimaging Initiative. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data. Comput. Biol. Med. 2021, 138, 104879. [CrossRef]

[43] Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; de Albuquerque, V.H.C. Multi-Class Skin Lesion Detection and Classification via Teledermatology. IEEE J. Biomed. Health Inform. 2021, 25, 4267–4275. [CrossRef].

[44] Masud, M.; Rashed, A.E.E.; Hossain, M.S. Convolutional neural network-based models

for diagnosis of breast cancer. Neural Comput. Appl. 2020, 1–12. [CrossRef]

[45] Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Appl. Sci. 2020, 10, 8298. [CrossRef]

[46] Zeebaree, D.Q. A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images. J. Appl. Sci. Technol. Trends 2020, 1, 78–91.

[47] Huang, K.; Zhang, Y.; Cheng, H.; Xing, P. Shape-adaptive convolutional operator for breast ultrasound image segmentation. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

[48] Sadad, T.; Hussain, A.; Munir, A.; Habib, M.; Ali Khan, S.; Hussain, S.; Yang, S.; Alawairdhi, M. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. Appl. Sci. 2020, 10, 1900. [CrossRef]

[49] Mishra, A.K.; Roy, P.; Bandyopadhyay, S.; Das, S.K. Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. Expert Syst. 2021, 38, e12713. [CrossRef]

[50] Hussain, S.; Xi, X.; Ullah, I.; Wu, Y.; Ren, C.; Lianzheng, Z.; Tian, C.; Yin, Y. Contextual level-set method for breast tumor segmentation. IEEE Access 2020, 8, 189343–189353. [CrossRef]

[51] Xiangmin, H.; Jun, W.; Weijun, Z.; Cai, C.; Shihui, Y.; Jun, S. Deep Doubly Supervised Transfer Network for Diagnosis of Breast Cancer with Imbalanced Ultrasound Imaging Modalities. arXiv 2020, arXiv:2007.066342020.

[52] Moon, W.K.; Lee, Y.-W.; Ke, H.-H.; Lee, S.H.; Huang, C.-S.; Chang, R.-F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional

neural networks. Comput. Methods Programs Biomed. 2020, 190, 105361. [CrossRef]

[53] Byra, M.; Jarosik, P.; Szubert, A.; Galperin, M.; Ojeda-Fournier, H.; Olson, L.; O'Boyle, M.; Comstock, C.; Andre, M. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. Biomed. Signal Process. Control 2020, 61, 102027. [CrossRef]

[54] Kadry, S.; Damaševiˇcius, R.; Taniar, D.; Rajinikanth, V.; Lawal, I.A. Extraction of tumour in breast MRI using joint thresholding and segmentation–A study. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; IEEE: New York, NY, USA, 2021; pp. 1–5.

[55] Lahoura, V.; Singh, H.; Aggarwal, A.; Sharma, B.; Mohammed, M.; Damaševiˇcius, R.; Kadry, S.; Cengiz, K. Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. Diagnostics 2021, 11, 241. [CrossRef].

[56] Maqsood, S.; Damasevicius, R.; Shah, F.M. An Efficient Approach for the Detection of Brain Tumor Using Fuzzy Logic and U-NET CNN Classification, International Conference on Computational Science and Its Applications; Springer: Cham, Switzerland, 2021; pp. 105–118.

[57] Rajinikanth, V.; Kadry, S.; Taniar, D.; Damasevicius, R.; Rauf, H.T. Breast-cancer detection using thermal images with marinepredators-algorithm selected features. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Noida, India, 26–27 August 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

[58] Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear recognition based on deep unsupervised active learning. IEEE Sens. J. 2021, 21, 20704–20713. [CrossRef]

[59] Majid, A.; Nam, Y.; Tariq, U.; Roy, S.; Mostafa, R.R.; Sakr, R.H. COVID19

classification using CT images via ensembles of deep learning models. Comput. Mater. Contin. 2021, 69, 319–337. [CrossRef]

[60] Sharif, M.I.; Alhussein, M.; Aurangzeb, K.; Raza, M. A decision support system for multimodal brain tumor classification using deep learning. Complex Intell. Syst. 2021, 1–14. [CrossRef]

[61] Liu, D.; Liu, Y.; Li, S.; Li, W.; Wang, L. Fusion of handcrafted and deep features for medical image classification. J. Phys. Conf. Ser. 2019, 1345, 022052. [CrossRef]

[62] Alinsaif, S.; Lang, J.; Alzheimer's Disease Neuroimaging Initiative. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data. Comput. Biol. Med. 2021, 138, 104879. [CrossRef]

[63] Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; de Albuquerque, V.H.C. Multi-Class Skin Lesion Detection and Classification via Teledermatology. IEEE J. Biomed. Health Inform. 2021, 25, 4267–4275. [CrossRef].

[64] Masud, M.; Rashed, A.E.E.; Hossain, M.S. Convolutional neural network-based models for diagnosis of breast cancer. Neural Comput. Appl. 2020, 1–12. [CrossRef]

[65] Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Appl. Sci. 2020, 10, 8298. [CrossRef]

[66] Zeebaree, D.Q. A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images. J. Appl. Sci. Technol. Trends 2020, 1, 78–91.

[67] Huang, K.; Zhang, Y.; Cheng, H.; Xing, P. Shape-adaptive convolutional operator for breast ultrasound image segmentation. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

[68] Sadad, T.; Hussain, A.; Munir, A.; Habib, M.; Ali Khan, S.; Hussain, S.; Yang, S.;

Alawairdhi, M. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. Appl. Sci. 2020, 10, 1900. [CrossRef]