REPUBLIC OF TÜRKİYE
ALTINBAŞ UNIVERSITY
Institute of Graduate Studies
Electrical and Computer Engineering

BREAST CANCER LOCALIZATION AND DETECTION USING MODIFIED SVM WORKFLOW

Ali Majeed Mohammed MOHAMMED

Master's Thesis

Supervisor
Asst. Prof. Dr. Ayça KURNAZ TÜRKBEN

İstanbul, 2024

BREAST CANCER LOCALIZATION AND DETECTION USING MODIFIED SVM WORKFLOW

Ali Majeed Mohammed MOHAMMED

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2024

This thesis title BREAST CANCER LOCALIZATION AND DETECTION USING MODIFIED SVM WORKFLOW prepared by ALI MAJEED MOHAMMED MOHAMMED and submitted on 11/01 /2024 has been **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering.

_____

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN
Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN	Department of Software Engineering, Altınbaş University	_____
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr. Zeynep ALTAN	Department of Software Engineering, Istanbul Beykent University	_____

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

I hereby declare that all information/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.

Ali Majeed Mohammed MOHAMMED

Signature

ABSTRACT

BREAST CANCER LOCALIZATION AND DETECTION USING MODIFIED SVM WORKFLOW

MOHAMMED, Ali Majeed Mohammed
M.Sc., Electrical and Computer Engineering, Altınbaş University,
Supervisor: Asst. Prof. Dr. Ayça KURNAZ TÜRKBEN
Date: 01 / 2024
Pages: 64

Mammography is the most effective method in the early detection of breast cancer, which can detect up to 90% of cases. mammography was used in only 24% of cases in 2017, which is below the 70% expected by the World Health Organization (WHO). This explains one of the causes of late diagnosis and the increase in mortality. According to WHO data, the average time to start treatment is 120 days after the first appointments. This situation is due to the delay in scheduling consultations, in addition to the fact that many times excessive tests are requested, which slow down the diagnostic process. In this paper, we Propose a medical decision support model in breast cancer to help diagnose more quickly, using prototype selection associated with (Support Vector Classifier (SVC), K-Nearest Neighbor (KNN), Random Forest (RF), Logistic Regression (LR)). In this way, we expect to obtain robust results with agility and efficiency.

Keywords: KNN, SVM, RF, ML, Breast Cancer.

TABLE OF CONTENTS

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

1. INTRODUCTION

1.1 BACKGROUND

Mammography is the most effective method in the early detection of breast cancer, which can detect up to 90% of cases. mammography was used in only 24% of cases in 2017, which is below the 70% expected by the World Health Organization (WHO). This explains one of the causes of late diagnosis and the increase in mortality. According to WHO data, the average time to start treatment is 120 days after the first appointments. This situation is due to the delay in scheduling consultations, in addition to the fact that many times excessive tests are requested, which slow down the diagnostic process. Ebru Bayrak and collaborators [3] looked at breast cancer prevention using machine learning to predict cancer. Making use of techniques to measure ranking performance, and the Wisconsin dataset. The techniques were compared with each other and the best performance was by the Support Vector Machine (SVM) Naresh Khuriwal and collaborators analyzed early stage breast cancer detection using machine learning techniques. The adaptive technique was proposedvoting ensemble, with the aim of explaining how the Artificial Neural Network (ANN) and logistic algorithm give better predictions, after analysis the final result was 98.50% accuracy compared to other machine learning algorithms During the last few years, breast cancer has shown high incidence rates, being considered the second type of cancer with the most occurrences of cases after melanoma skin cancer In the year 2019, around 18,295 deaths were recorded, with 18,068 of these deaths occurring in women and 227 in men. In addition, around 66,280 new cases are estimated for the year 2021.

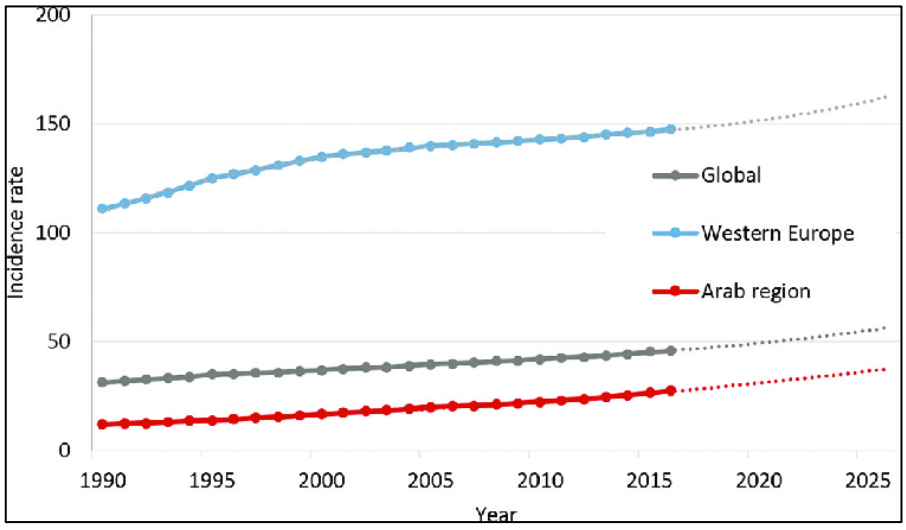

Figure 1.1: Expected Breast Cancer Rates in the World By 2025 [1].

Breast cancer can result in genetic mutations that cause an increase in the hormone estrogen, with the possibility of tumors appearing as well. in regions close to the breast It is also worth noting that some genetic factors, related to lifestyle and the environment, may contribute to the occurrence of this type of cancer Due to the exacerbated advance of breast cancer tumors, it is worth noting that early diagnosis is highly relevant in controlling the evolution of the disease due to the early initiation of treatment, which contributes to increasing the chances of survival of the diagnosed patient Since early diagnosis is of paramount importance, the main procedures used to diagnose this disease are: mammography, clinical examinations, blood tests, X-rays, ultrasound, biopsy and others Among the procedures mentioned above, as highlighted in this work, we have the biopsy, which has played an extremely important role in the detection of breast cancer consisting of the removal of the sample of the tissue that contains the suspected tumor for cell analysis. Generally, this type of procedure is indicated for confirming the existence of cancer in cases of changes in the breast that may appear as nodules. Faced with the advancement of computer vision techniques in relation to the study, analysis and processing of features/resources in images, several studies have developed research and projects for the use of these technologies in terms of classifying these types of images, in order to contribute to the work of pathologists, as it performs the use of color segmentation and recognition of morphological patterns through pixel analysis Among these works, it can be highlighted: an approach to training the ResNet50 network using the K-means algorithm for data clustering, together with color transfer and knowledge transfer, training of VGG networks using color transfer and knowledge transfer, use of Inception-V3 architecture network, with color transfer and SVM algorithm AlexNet network training with color transfer and knowledge transfer both aiming at the classification of histopathological images of breast cancer. In view of this, this work is based on the works mentioned above, in addition to other works that are related to the classification of histopathological images of breast cancer through the use of color and knowledge transfer techniques. use of Inception-V3 architecture network, with color transfer and SVM algorithm AlexNet network training with color transfer and knowledge transfer both aiming at the classification of histopathological images of breast cancer. In view of this, this work is based on the works mentioned above, in addition to other works that are related to the classification of histopathological images of breast cancer through the use of color and knowledge transfer techniques. use of Inception-V3 architecture network, with color transfer

and SVM algorithm AlexNet network training with color transfer and knowledge transfer both aiming at the classification of histopathological images of breast cancer.

1.2 PROBLEM STATEMENT

Breast cancer is the most common among women. According to [2], 2.1 million new diagnoses of breast cancer are expected in 2019. In addition, this type of cancer is the fifth in the world in mortality, with an estimated more than 450,000 deaths per annum. Therefore, quickly identifying, through an early diagnosis, it is of paramount importance to reduce mortality.

The diagnosis of the most varied types of cancer has undergone significant advances over the years, as well as the methods used, from imaging to biopsy, which allows for a more assertive diagnosis. However, early diagnosis of breast cancer is still a challenge.the delay is based, in most cases, on difficult access to health and, in its minority, on “uncertainty”. when the patient is not in the risk group.

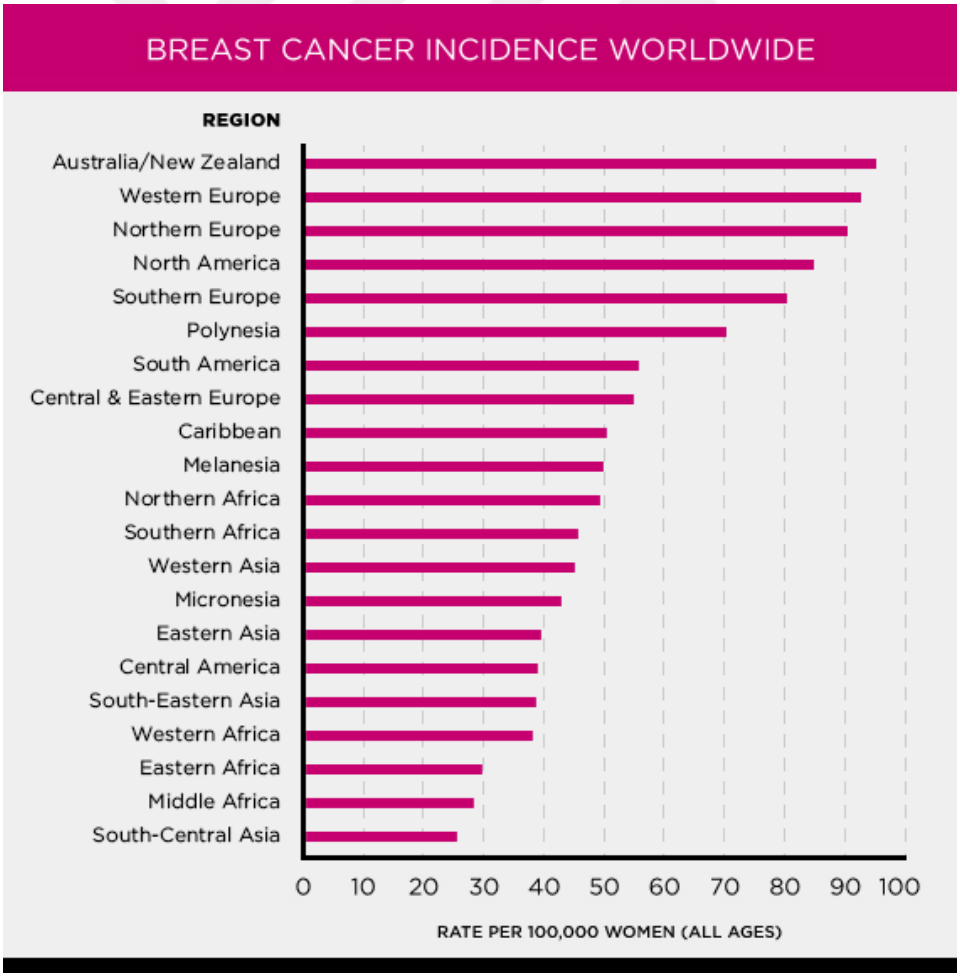

Figure 1.2: Mortality Rates and Ages of Breast Cancer Worldwide.

Control of breast cancer is a priority of health policy and was included as one of the goals of WHO whose objective is to strengthen, integrate and resolve the Unified Health System, through strategies co-responsibility of federal, state and municipal managers. Nowadays, the use of high-resolution mammography equipment equipped with a fine focus for magnification, an adequate film/screen combination and specific processing has enabled the detection of an increasing number of breast lesions, especially small lesions, when they are not yet palpable. According to the literature, mammography has sensitivity between 88% and 93.1% and specificity between 85% and 94.2%, and the use of this test as a screening method reduces mortality by 25%. To ensure the performance of mammography, the image obtained must be of high quality and, therefore, the following are necessary: adequate equipment, correct radiological technique, knowledge, practice and dedication of the professionals involved.

1.3 THESIS OBJECTIVES

Propose a medical decision support model in breast cancer to help diagnose more quickly, using prototype selection associated with (Support Vector Classifier (SVC), K-Nearest Neighbor (KNN), Random Forest (RF), Logistic Regression (LR)). In this way, we expect to obtain robust results with agility and efficiency.

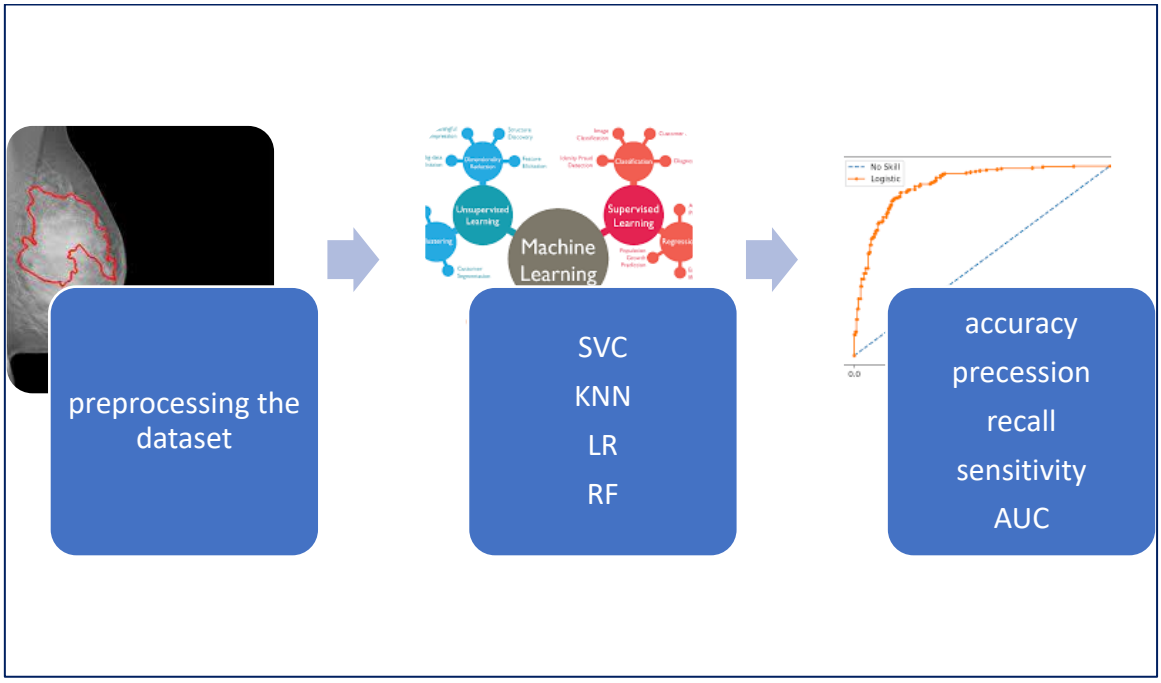

Figure 1.3 Outline of the Proposed Method.

1.4 THESIS MOTIVATION

Even with advancements in early detection technologies, breast cancer remains the most frequent cancer among women. Diagnosis is frequently delayed and discovered at an advanced stage. Because of intensive therapy and a high mortality rate, this causes misery. As a result, any advancement in the development of breast cancer prediction and diagnosis is critical. Machine learning techniques, for example, can have a significant impact on the process of early cancer diagnosis and prediction. The usage of a deep learning model has become consistent in various works, where they are able to interpret photos in a matter of seconds. The Massachusetts Institute of Technology (MIT) claims that this model has been utilized in over 10,000 mammograms and has obtained a 90% agreement rate.

1.5 HYPOTHESIS

Breast cancer detection using prototype selection when compared to machine learning techniques most used in the literature, for example SVC, KNN and logistic algorithm

1.6 METHODOLOGY

The dataset used in this study will be the Wisconsin Breast Cancer (WBC), which will be obtained from the UCI Machine Learning Repository. The dataset includes 699 cases that will be classed as benign or malignant. The project's implementation is divided into three stages. The first part will be to examine the data and perform any necessary pre-processing, and the second will be to group the data according to their characteristics, using the prototype selection method, in which we select what is most representative from the data set while eliminating the others. This allows us to deal with the least amount of data possible, resulting in speedier processing.

2. LITERATURE REVIEW

2.1 CHAPTER INTRODUCTION

The implementation of artificial intelligence can provide significant benefits to both the interpretation of images and the outcomes of diagnostic tests [18]. In the study of artificial intelligence (AI), the subfields of machine learning and deep learning are two areas that need to be researched extensively in order to gain an understanding of the possible medical applications of these areas. "Machine learning," also abbreviated as "ML," is the academic field that focuses on the process of teaching computers new skills without the need for those skills to be specifically taught to the computers themselves. By taking this method, machine learning has the potential to construct models from data and offer predictions on the basis of empirical evidence. Not only is machine learning useful for making predictions, but it may also be put to use to unearth patterns in the data that had been concealed as well as essential information regarding the data's underlying composition. Deep Learning is a subsection of machine learning that focuses on expressing data in the most effective manner possible to make the learning process easier. It is also abbreviated as DL for short. One of the most significant distinctions between DL and ML is that the former can learn from any kind of data, whereas the latter requires features to be defined by humans that precisely reflect the data that is being received in order to learn from it. DL can learn from any data; ML can only learn from data that has been defined by humans.

2.2 RELATED WORKS

According to projections made by the American Cancer Society, there would be 2.26 million new cases of breast cancer identified in women in the year 2023 [1]. This would place breast cancer at the top of the list of the 36 different types of cancer in terms of its prevalence. In addition, out of the 35 different types of cancer, the incidence of new fatalities caused by female breast cancer was the fourth highest among them, coming in at 0.684 million [1]. [Citation needed] Mammography, digital breast tomosynthesis, breast ultrasonography, and magnetic resonance imaging are the four screening methods that are utilized most frequently nowadays for the detection of breast cancer [2]. Screenfilm mammography and Digital Mammography (DM) are the two types of mammography; these are both x-ray imaging procedures that use radiation to create a 2D image of the breast tissue [2, 3]. Mammography can detect breast cancer by detecting changes in the breast tissue that could indicate the

presence of cancer. In the 1940s, mammography was first developed as a diagnostic tool. Mammography has also proved useful in the early diagnosis of breast cancer, which has contributed to the drop in the mortality rate caused by this disease [2, 4]. This has contributed to the decline in the overall mortality rate caused by this disease. Because of the progress that has been made in imaging technology, digital breast tomosynthesis may now be carried out (DBT). The problems that occurred during mammography were resolved thanks to DBT, and the quality of the pictures that were produced as a result was improved [3, 5–7]. This is accomplished by fusing together the information obtained from a variety of distinct two-dimensional breast pictures [2, 5, 6]. On the other hand, these 3D images (volumes) may only be regarded quasi-3D due to the fact that they were reconstructed from a large number of 2D photographs [6, 7]. In addition to this, an x-ray tube that is perpendicular to the chest wall is utilized to acquire image slices along an arc that extends from 15 degrees to 60 degrees [5]. [Citation needed] Because of the high frequency that is used by Automatic Breast UltraSound, the entire breast may be observed in a single pass. This is possible because of the high frequency (ABUS). Before it is possible to build a 3D volume, these photos of cross-sections must first be converted from their initial 2D format into a 3D representation [8]. Imaging of the breast is made possible via magnetic resonance imaging (MRI), which makes use of strong magnets in conjunction with radio waves that are produced by a computer [2]. Because CAD systems function as a "second opinion," radiologists are able to feel a greater sense of certainty in the diagnoses that they have made [9], [11]. According to [12], the CAD solution needs to make radiologists more productive, cut down on the amount of work they have to complete, easily integrate with the processes they already use, and not put the practice at danger in any way, shape, or form. It is anticipated that the introduction of deep learning algorithms into CAD systems will achieve the aforementioned goals, in addition to lowering the recall rate, increasing the sensitivity of cancer diagnosis, and reducing the evaluation variability between radiologists [13, 14]. [13] In the categorization and localization of cancer tumors in medical pictures, models based on deep learning have already outperformed the performance of radiologists [15, 16]. The analyzed results of these models can be employed in the medical field. There is currently a rise in the availability of powerful computers, vast datasets that have been properly managed, and up-to-date algorithms. Deep learning [18] simplifies the representations that are being transmitted in order to get around the difficulty that is associated with learning new

representations. As the study of machine learning advances, it will eventually be possible for it to provide assistance with aspects of medical treatment that were previously inaccessible. [17] Deep learning [18] reduces the complexity of the representations that are being transmitted in order to get around the difficulty that is associated with learning new representations. In addition, deep learning involves the employment of numerous layers of artificial neurons that are connected to one another in order to learn the patterns of simplified representations of the actual data [18, 19]. This is done in order to learn the patterns of deep learning. The architecture of a convolutional neural network had a lower error rate of 15.3% in 2012, which was 10.9% better than the architecture that finished in second place (see entry #20). This accomplishment was the direct cause of an increase in the number of people participating in research on deep learning, and the CNN architecture is still the subject of ongoing research and application for the goal of resolving issues related to image recognition [21]. Convolutional Neural Networks, more commonly known as CNNs, have been developed in order to analyze data sets that contain established grid patterns and to discover the spatial hierarchies of qualities contained within data [18], [22]. Because CNNs are able to teach themselves new characteristics on their own, oncologists are no longer need to manually input characteristics of malignant tumors into the system. The subject of medicine, which provides the ideal setting for the use of deep learning, may prove to be an effective application for this type of learning. The significance of the characteristics that are specifically produced for a given application of machine learning was recently highlighted, and lesion interpretation is not an exception to this rule. The vast majority of the studies use a wide variety of various texture features to illustrate the change in pixel intensities that reflect the structure of the object that is being researched [22]. [Citation needed] Wolfe [23] was the first researcher to establish a link between the texture of breast tissue and the likelihood of getting breast cancer. He was able to accomplish this by recognizing four distinct breast texture patterns. • N1: Lowest Risk, predominantly adipose tissue, no visible channels in the parenchyma; this kind presents the least degree of danger possible. • The danger is modest for a P1 classification since the ducts can only spread to one of the four quadrants of the breast. • The presence of ducts that "severely involve" more than twenty-five percent of the breast tissue is indicative of a high risk, which is indicated by the P2 classification. • Dysplasia carries the highest risk because it is present, in addition to the considerable involvement seen in P2; thus makes DY the condition with the highest risk.

When a healthcare practitioner's interpretation is involved in assigning a breast into a given group, the classification ceases to be objective [24]. Objectivity is lost when a classification relies on the interpretation of a healthcare clinician. Other scholars [25] have arrived at the same conclusion that Wolfe's studies are not the only ones to demonstrate an association between certain characteristics of the parenchyma texture pattern and an increased risk of breast cancer. Because the ties between fatty and glandular tissue that are defined by texture and can be connected with either a healthy or ill condition, the research and manufacture of different texture descriptors was prompted as a result of these links, the descriptors were explored and produced. These identifiers are capable of being classified according to a wide variety of distinct categories, each of which will be examined in further depth in the following paragraphs. There are other alternative groupings of texture features that may be utilized, but the two that we have discussed so far are the ones that are most frequently used in machine learning research on BC identification. Other possible groupings of texture features include: It is necessary to make use of Feature Descriptors [29], which may be located on this page, in order to characterize qualities that are determined by applying the power spectrum. Examples of such properties include the First Moment and the Root Mean Square. The Local Binary Pattern, more commonly referred to as LBP, is employed in texture analysis because of its capacity to provide an exact description of the link between a pixel and the area surrounding it. This ability is the reason the LBP is generally referred to as the LBP. The explanation that follows is one definition of an LBP: An examination will be carried out that covers the surrounding area pixel by pixel, and it will be based on the neighborhood that has been provided. If the value of a pixel is already higher than the value of the center pixel, then the value of that pixel will be dropped by 0; if this is not the case, then the value of that pixel will be increased by 1. After all of those processes have been finished, the pixel values that were produced will be added in a counterclockwise direction in order to make a unique binary identifier that can be translated to a decimal value at a later time. This will be done in order to produce a hexadecimal value. [30]. It was found out by Chen et al. [31] in their research that morphological aspects are commonly used in order to examine the characteristics of tumors. This was a really interesting discovery. One of the key morphological distinctions that may be made between benign and malignant tumors is the relative smoothness and regularity of the benign tumor's margins. This can be done by comparing the margins of the benign tumor to those of the malignant tumor. The writers

create and describe a number of different components, including the ones that are listed below, in order to accomplish this goal. As a result of this, a group of researchers [32] came to the conclusion that mammography should be divided into the following three categories: benign, benign, and malignant. This was made feasible through the utilization of a database that contained information from 322 different mammograms, so thank you for that! (126 normal, 60 benign, and 48 malignant). The remaining mammograms were put to use in the training set, while the test set was made up of a total of 37 benign, 23 noncancerous, and 18 cancerous mammograms. The researchers applied a method that is known as Contrast-Limited Adaptive Histogram Equalization in order to enhance the general quality of the photographs. In order to make the process of analysis more straightforward, a Region of Interest (ROI) was drawn on the initial image and then its contents were retrieved. Following the extraction of features from the GLCM texture, the mammograms were then in a position to be described. It is important to take note that four distinct GLCM were built; one for each of the following degrees: 0 degrees, 45 degrees, 90 degrees, and 135 degrees. Following the removal of the descriptors, the authors attempted to classify the mammograms in the most accurate manner possible so that they could proceed to the following stage. They require a two-way classifier, but the only option they had was a binary one; hence, they decided to make use of a Support Vector Machine (SVM) (SVM). These two-way classifiers demonstrate that the SVM was trained in two distinct ways: first, to differentiate between normal and abnormal tissue; second, to differentiate between benign and malignant tissue within the images that were classified as abnormal; and finally, to differentiate between normal and abnormal tissue in the images that were classified as abnormal. To make things easier to understand, the authors of the study came up with an average value for each quality throughout the entire spectrum of possible orientations and distances. This was done in an effort to simplify the problem. Following the completion of this procedure, a total of fifteen distinguishing characteristics were uncovered. Following the production of an image of the scan, a classification can be carried out in order to determine if the results of a mammography are normal or abnormal. If the result is abnormal, the process continues on to the next step; otherwise, the image is sent to a second SVM, which analyzes it to decide if the lesion is benign or cancerous. The Support Vector Machine (SVM) was put through its paces during both the main and secondary classification stages. When it came to distinguishing normal tissue from abnormal tissue, the classifier had an accuracy of 100%, while its sensitivity and

specificity were 94.4% and 91.3%, respectively, when it came to discriminating malignant tissue from benign tissue. In spite of the fact that the findings of this study are positive, there are a few points regarding it that should be kept in mind: It is hard to assess how effectively the produced model generalizes because the size of the test set that was used to evaluate performance was very small, and because it came from the same distribution as the training data. This makes it impossible to evaluate the performance of the model. Mohanty et al. [33] aimed to design a method that may identify between tumors that are carcinogenic and those that are not cancerous in a manner that is similar to how the previous research was conducted. A dataset that was produced at the University of Florida was utilized by the researchers. This dataset contains digital mammograms that were annotated by a radiologist who specializes in the field with the locations of masses that were seen on the images. These annotations acted as guidance while the region of interest was being extracted from the image. The image had a low-pass filter applied to it so that the essential information could be preserved while at the same time the noise could be eliminated in order to achieve superior image quality. In this particular instance, we have a testing set that consists of photographs of both cancerous and noncancerous tumors, but the training set is comprised of 88 specimens that are identically representative of normal and cancerous tissue (23 and 55, respectively). The initial definition of a region-of-interest (ROI) made it possible to extract 19 features, which were then classed as either GLCM or RLM based on the results of the categorization process. Immediately after the process of feature extraction, a machine learning model was constructed with the help of a Decision Tree. By applying techniques like boosting, winnowing, and pruning, we were able to improve the model's capacity for generalization while simultaneously lowering the likelihood of the model becoming overly accurate. By utilizing 8 GLCM features and 11 RLM features, they were able to accurately diagnose both benign and malignant tumors with an accuracy value that reached up to 96.7%. These findings provide a lot of cause for optimism. In addition, it was discovered that the Area Under the Curve (AUC) of the ROC curve is 0.995, which is a reward in its own right. These findings not only demonstrate the significance of textural features in lesion interpretation, but they also demonstrate how a Decision Tree can be used to create algorithms that accurately classify mammograms by making use of features that can be easily explained. In other words, these findings demonstrate both the significance of textural features and how a Decision Tree can be used. In a nutshell, these findings illustrate not just the value of textural

aspects but also the application of a Decision Tree. The strategy, despite its efficacy, adds more work to the plates of healthcare employees who are already working more than their fair share of shifts because to the time-consuming nature of the processes involved in ROI creation and pre-processing.

2.3 SUMMARY OF RELATED WORKS

2023 will see 2,260,000 breast cancer cases [1]. Breast cancer killed 6844,000 women [1]. Breast cancer screenings include mammography, digital breast tomosynthesis, breast ultrasonography, and magnetic resonance imaging [2]. Screenfilm and DM use radiation to create 2D breast tissue pictures [2, 3]. Mammography early diagnosis lowers breast cancer mortality [2, 4]. Imaging-enabled digital breast tomosynthesis (DBT). DBT-enhanced mammograms [3, 5–7]. 2D breast images become 3D [2, 5, 6]. Many 2D photos rebuilt quasi-3D volumes [6, 7]. An x-ray tube parallel to the chest wall separates images into 15°–60° arcs [5]. High-frequency Automatic Breast UltraSound shows the entire breast (ABUS). 3D cross-sections [8]. Breast MRI combines large magnets and computer-generated radio waves [2]. CAD gives radiologists "second opinions" [9–11]. CAD should improve radiologist efficiency, minimize effort, fit into present methods, and not harm the practice [12]. Deep learning algorithms in CAD systems lower recall rate, increase cancer diagnosis sensitivity, and reduce radiologists' evaluation variability [13], [14]. Deep learning algorithms surpass radiologists in cancer diagnosis and localization [15, 16]. Computers, algorithms, and databases are expanding. Machine learning will make healthcare possible [17]. Deep learning [18] simplifies learning representations. Deep learning simplifies representation patterns utilizing many layers of interconnected artificial neurons [18, 19]. Convolutional neural networks outperformed the runner-up by 10.9% in 2012 with a 15.3% error rate (see item #20). CNN architecture is still employed for picture identification [21]. CNNs learn spatial attribute hierarchies from grid-patterned input [18, 22]. CNNs learn malignant tumor traits for clinicians. Healthcare uses deep learning well. ML requires lesion interpretation. Pixel intensities reflect object organization in most investigations [22]. Wolfe [23] connected breast cancer to four textural patterns. N1—Lowest risk, mostly fatty tissue. P1—Low Risk, ducts reach one breast quadrant. • P2—High Risk, ducts "severely involve" over 25% breast. DY—Highest risk, dysplasia, substantial P2 involvement. Healthcare providers subjectively classify breasts [24]. Wolfe's findings that parenchyma texture pattern factors affect breast cancer risk are supported [25]. Several textural characteristics linked

fatty and glandular tissue to health or illness. Sort these. ML BC detection uses two texture feature types. Feature descriptors [29] define power spectrum properties like First Moment and Root Mean Square. LBP texture analysis accurately describes pixel connections. LBP—neighborhood-based pixel-by-pixel analysis. Pixels larger than the middle pixel gain 1 and lose 0. Pixel values will be added anti-clockwise to provide a decimal-convertible binary identification [30]. Chen et al. [31] say morphology classifies malignancies. Malignant tumor boundaries are rougher. Writers create and describe many things: Tumors are characterized by area, circularity, compactness, eccentricity, elliptic-normalized circumference, and elliptic-area ratio. Larger malignant tumors. Most lesion classification studies predict malignancy. Thus, scientists classified benign, benign, and malignant mammograms [32]. 322 mammograms allowed (126 normal, 60 benign, and 48 malignant). The test set comprised 37 noncancerous, 23 benign, and 18 malignant mammograms. Contrast-Limited Adaptive Histogram Equalization improved scientists' images. Source image analysis needed a Region of Interest (ROI). GLCM mammogram texture analysis. Four GLCM were built for 0, 45, 90, and 135 degrees. Descriptors are used to classify mammograms. The binary two-way classifier needs a Support Vector Machine (SVM) (SVM). The SVM was trained to detect normal and abnormal tissue and then benign and malignant tissue in aberrant images using these two-way classifiers. The authors averaged all attributes across orientations and distances to simplify the issue. This procedure provided 15 attributes. Classifiers evaluate mammograms. Unless it's normal, a second SVM labels the image as benign or malignant. SVM categories were primary and secondary. Classifier's sensitivity was 94.4% and specificity was 91.3%. This study's results are promising, yet the test set is small and from the same distribution as the training data, so we can't judge the model's generalization. Mohanty et al. [33] separated malignant from noncancerous tumors. The scientists used University of Florida radiologist-annotated digital mammograms. Annotations guided ROI extraction. Low-pass filters eliminated noise and maintained features. The training set has 88 normal and cancer images, while the testing set has malignant and benign tumor photos (23 and 55, respectively) (23 and 55, respectively). Decision Tree created a machine learning model utilizing 19 GLCM or RLM features after feature extraction. Boosting, winnowing, and pruning minimized overfitting and improved generalization. 8 GLCM and 11 RLM features accurately diagnosed benign and malignant tumors at 96.7%. Bonus: ROC curve AUC 0.995. These findings demonstrate the importance

of textural cues in lesion interpretation and how a Decision Tree can be used to create algorithms that properly diagnose mammography using easily explained variables. ROI definition and pre-processing require time, adding to healthcare workers' already heavy workloads.

2.4 CHAPTER CONCLUSIONS

Because of this, there is a greater degree of learning that does not require direct participant participation in DL [19,20]. There are many distinct facets of medical imaging that could potentially benefit from the application of artificial intelligence. For instance, artificial intelligence (AI) systems could aid improve imaging systems by enhancing localization and providing a definition of the findings, in addition to optimizing the timing of acquisition. In the meantime, researchers have been looking into the possibility of using automatic lesion detection to diagnose a wide range of medical conditions, including those that affect the breast, the lungs, the thyroid, and the prostate. Some of these conditions include breast cancer, lung cancer, thyroid cancer, and prostate cancer. In addition, magnetic resonance imaging (MRI), ultrasonography, and tomosynthesis have all been applied in the process of determining how effective these AI detection procedures are. Systems that make use of artificial intelligence aren't just useful for automatically discovering lesions; they may also be put to use for analyzing the findings of lesions once they have been found. CAD systems can help radiologists and image interpreters make more accurate diagnoses and better predict the outcomes of any lesions that they find by acting as a second set of eyes and providing a second opinion. This can be accomplished by acting as a second set of eyes and providing a second opinion. This interpretation typically applies to normal or abnormal conditions, but it can be expanded to handle more complex operations. As a result, picture analyzers can save valuable time and improve their overall efficiency. Image registration and volume segmentation are two examples of post-processing and quality analysis that can benefit from the application of AI across a variety of imaging modalities. Both of these post-processing and quality analysis tasks involve registering images and segmenting volumes. Post-processing and quality analysis jobs that require registering images and segmenting volumes are examples of these types of tasks.

3. MATERIALS AND METHODS

3.1 BREAST CANCER

Cancer refers to a group of diseases defined by the uncontrollable proliferation and spread of aberrant cells. If the malignant cells are not removed, the condition will progress, and the affected individual will die sooner or later.

A tumor is a collection of cells that can be either malignant or benign. Breast cancer is one of these malignancies that has become a big public health issue with a real need for intervention and care.

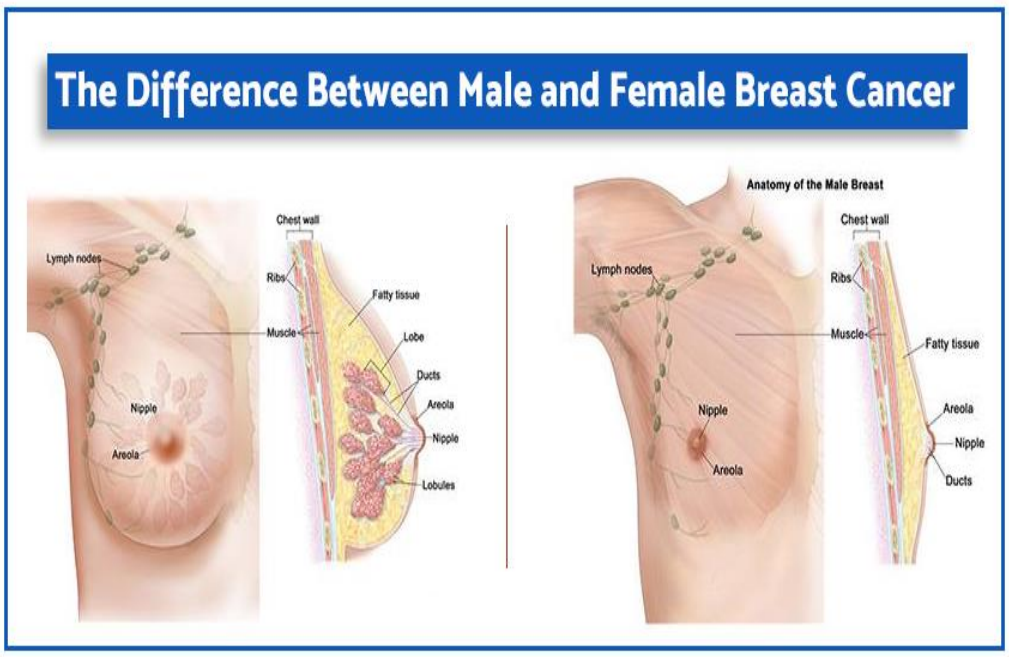

Figure 3.1: Breast Anatomy A) Male And B) Female.

3.1.1 Symptoms

The medical specialist should absolutely examine you if you have such symptoms. Most of the time, affected women discover something abnormal in the chest themselves. In nine out of ten cases, these disorders are not due to cancer.

Any of the following symptoms can be a telltale sign of breast cancer:

a. Hard or solid nodule in breast or armpit.
b. Modification of the skin: in particular redness or orange peel appearance.
c. Discharge from the nipple.
d. Retraction or bulging appearance (thickening) of the skin or nipple.

3.1.2 Processing

If you have discovered breast cancer, the healthcare team decides to start treatment, there are several ways to treat breast cancer.

3.1.3 Causes and Risk Factors

A few factors increase the risk of developing breast cancer. The most important are not editable:

a. Age: The risk increases significantly after age 50, but younger women can also be affected
b. Family history: Women whose sisters, mothers or daughters have breast cancer are at greater risk. Especially if family members are affected before the age of 50.
c. Hereditary predispositions: approximately 5-10% of all breast cancers are triggered by an inherited predisposition. The women concerned are often ill before the age of 50. Natural hormone metabolism: the risk of breast cancer is slightly higher in women whose first period occurred before the age of 12, whose last period occurred after the age of 55, in women who have no children or who gave birth after the age of 30. Lifestyle also plays a role. The following factors may slightly increase the risk:
d. Hormonal treatment for disorders related to "menopause, taking the contraceptive pill, smoking, excessive alcohol consumption, obesity, an unbalanced diet, high in lipids, lack of physical activity".

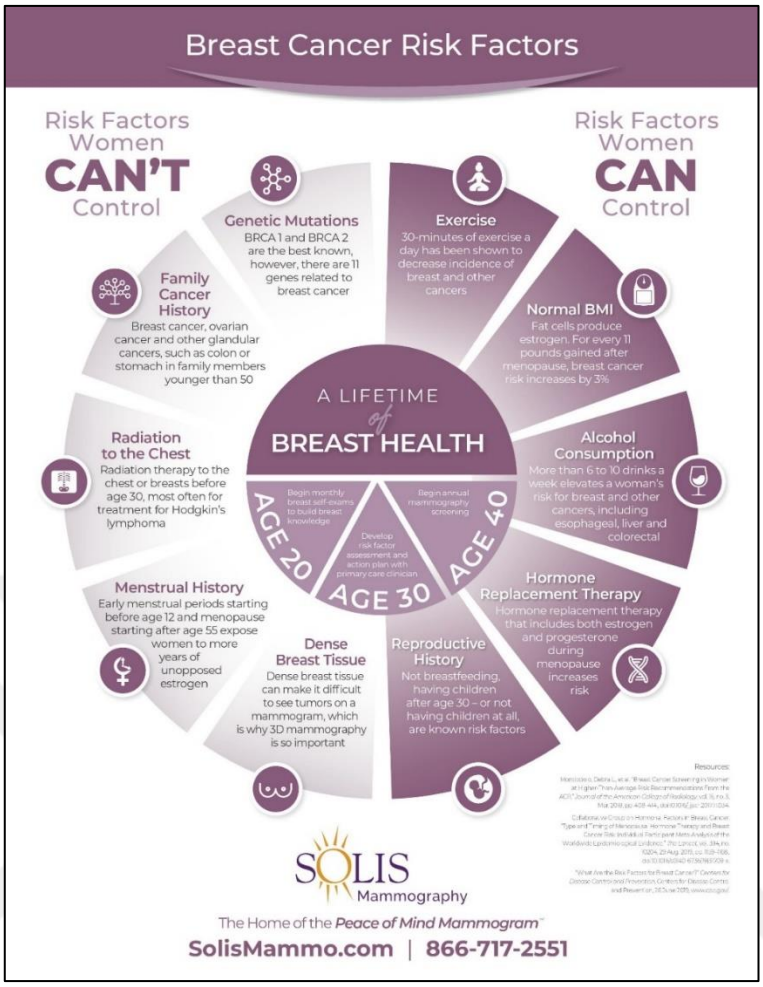

Figure 3.2: Risk Factors For Breast Cancer.

3.1.4 Diagnostic Tests

To diagnose breast cancer, two methods are used in the first place:

Mammography (Breast X-ray): Mammography is a low-dose X-ray of the breast. The resulting image is called a mammogram.

3.2 MACHINE LEARNING

This section gives an overview of the theoretical bases relevant to machine learning. We make a state of the art that gives a general view on classification methods.

Machine learning is the field applied to understand and generate the ability of human learning from an artificial system. This is, very schematically, to design (algorithms and methods) to extract adequate information from data, or to learn behaviors from examples. Thus, the essential goal of ML is to build the relationship between objects and their categories for prediction and knowledge discovery (Mohamed Bouguessa, 2015). We are

now going to describe the main Machine Learning algorithms that will allow us to carry out this learning. We distinguish two parts “supervised learning and unsupervised learning”. In this project, we will only discuss supervised machine learning.

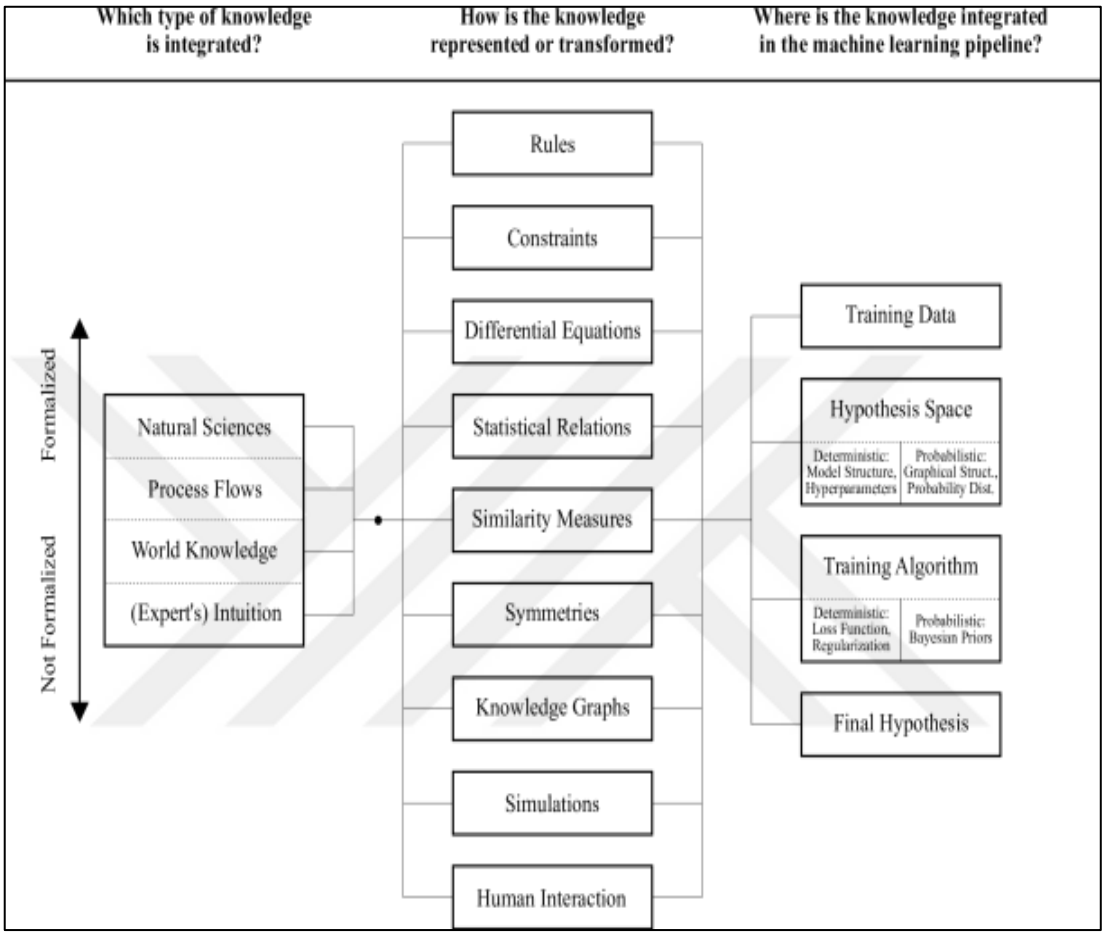

Figure 3.3: Machine Learning Model Taxonomy.

3.2.1 Concepts and Basic Terminologies

Machine learning is one of the fields of study of artificial intelligence. It refers to the capacity of a system to autonomously acquire and integrate knowledge Machine learning refers to the development, analysis and implementation of methods that allow a machine (in the broad sense) to evolve and perform tasks associated with artificial intelligence through a learning process. This learning makes it possible to have a system that is optimized according to the environment, the experiences and the results observed The methods of automatic learning form a class of attractive techniques for the accomplishment mentioned knowledge extraction tasks.

3.2.2 Data Pre-processing

Data preprocessing is one of the most common data mining tasks. It involves a data mining process that prepares data and converts it into a suitable form. Data preprocessing aims to reduce data size, find relationships between data, normalize data, remove outliers, and extract features from data. It includes several technologies such as cleaning, integrating, transforming and reducing data

3.2.3 Normalization

Feature scaling is a technique used to normalize the range of independent variables or data features. In data preprocessing, also known as data normalization, it is usually used in the data preprocessing step**.**

3.2.4 Classification

The training dataset is utilized to improve the boundary conditions that can be used to select each target category. After determining the boundary criteria, the next challenge is to estimate the target category. The entire procedure is known as classification. There are two kinds of classification: supervised classification and unsupervised classification. Predefined information is required for supervised classification, however unsupervised classification, also known as clustering or exploratory data analysis, does not require predefined labeled data.

a. Unsupervised learning:

   In clustering, the category of the object is unknown, however, we know the rules to classify, and we also know the characteristics (independent variables) that can describe the classification of the object. There is no training example to check if the classification is correct. Therefore, objects are simply assigned to groups based on given rules.

b. Supervised Learning:

   Supervised learning is to build an incisive model of the distribution of class labels in terms of predictive features. The resulting classifier is then used to assign class labels to test instances where the values of the predictive features are known, but the value of the class label is unknown supervised learning aims to establish rules of behavior from a database containing examples of already labeled cases. The database is in principle a set of input/output pairs {(X, Y)}. The goal is to learn to predict for any new input X, the output Y, In the supervised learning illustrated in figure 2.2, the learner has two sets of data, a training set and a test set. The idea is that learners "learn" from a set of labeled

examples in the training set so that they can identify unlabeled examples in the test set with the greatest possible accuracy.

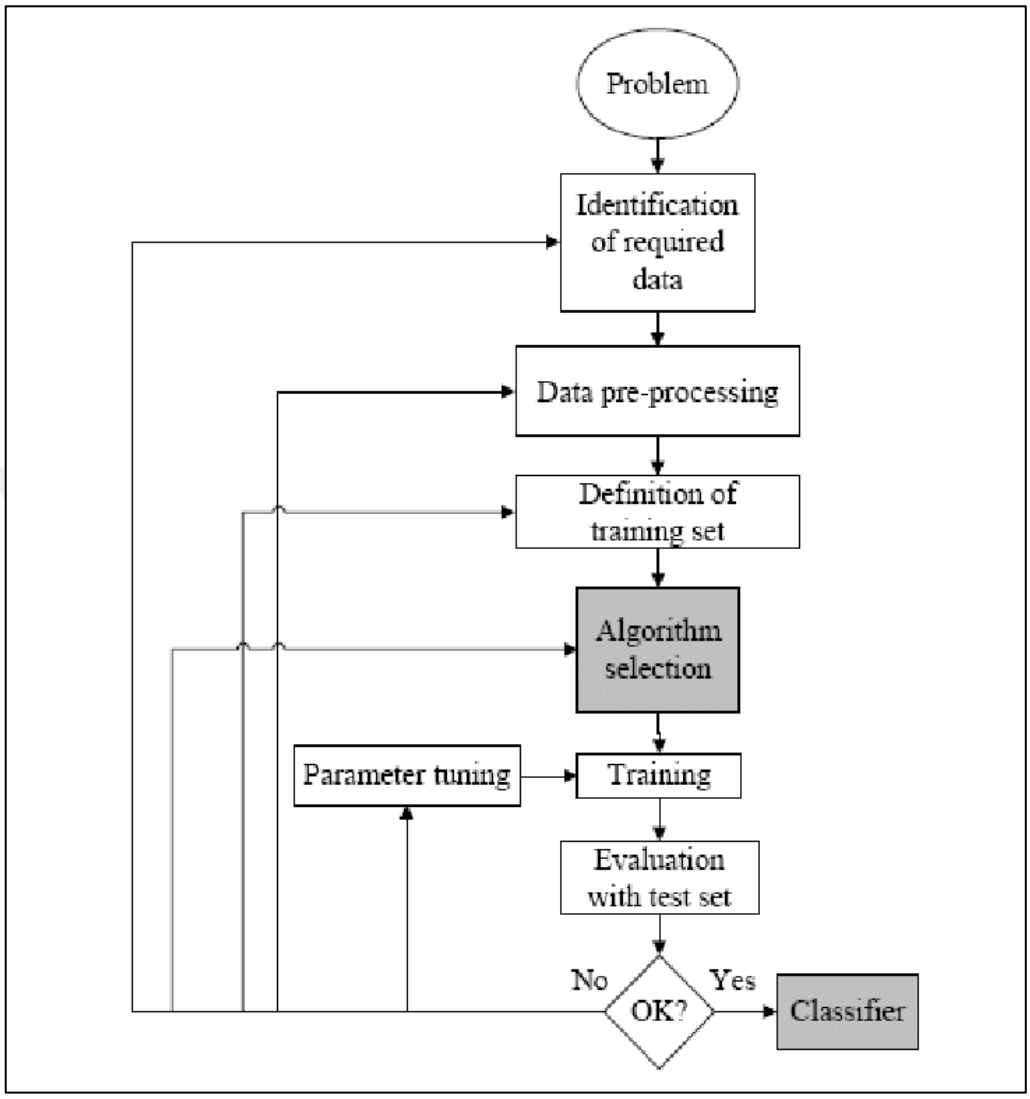

Figure 3.4: Supervised Machine Learning Processes.

c. Supervised learning models:

We use the training dataset to obtain better boundary conditions that can be used to determine each target category. Once the boundary conditions have been determined, the next task is to predict the target category. The whole process is called classification. There are two types of classification, supervised and unsupervised. In supervised classification, available predefined knowledge is needed, while in unsupervised classification, sometimes called clustered or exploratory data analysis, there is no need to have predefined labeled data.

d. Logistic regression:

Regression analysis is widely used for forecasting, understanding the independent variables related to the dependent variable, and exploring the shape of these relationships. In limited circumstances, regression analysis can be used to infer the causal relationship between the independent variable and the dependent variable. The logistic function is an S-shaped curve that can take any real number and map it to a value between 0 and 1, but never exactly to those limits.

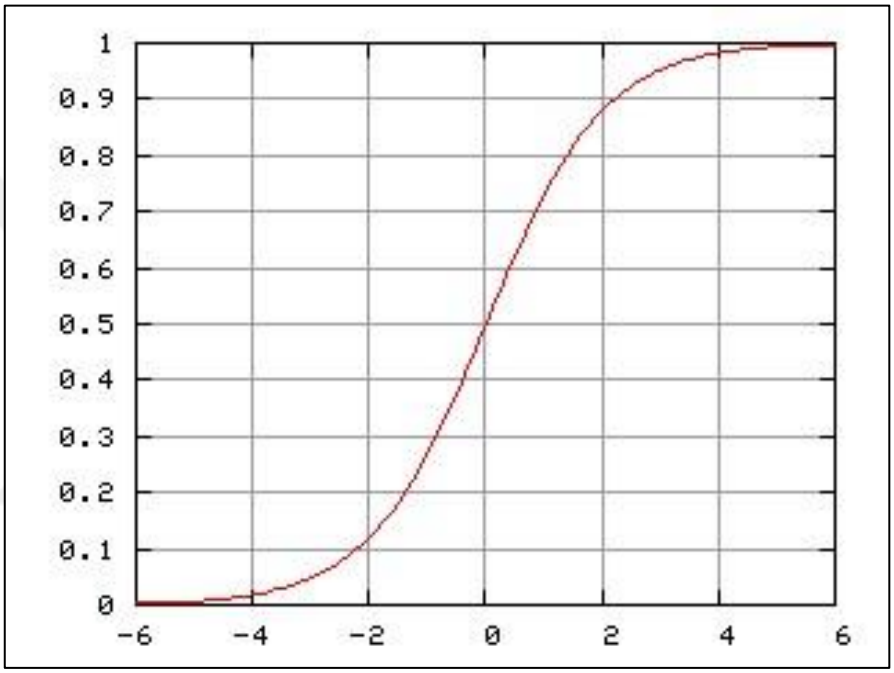

Figure 3.5: The Logistic Function (Sigmoids).

e. Naive Bayes method:

The naive Bayes classification method is a supervised machine learning algorithm that classifies a set of observations according to the rules determined by the algorithm itself. The classification tool must first be trained on a training dataset, which displays the expected category based on the input. In the learning phase, the algorithm develops its classification rules on this dataset and applies them to the classification of the predicted dataset in the second step. The NB classifier signifies that the class of the training data is known and provided. NB classification has achieved outstanding results in many everyday applications, making it the preferred algorithm for classification tools.

The basis of NB classification is Bayes' theorem, which simplifies the assumption of independence between all pairs of variables, which is called naïve of the predictor of the given class.

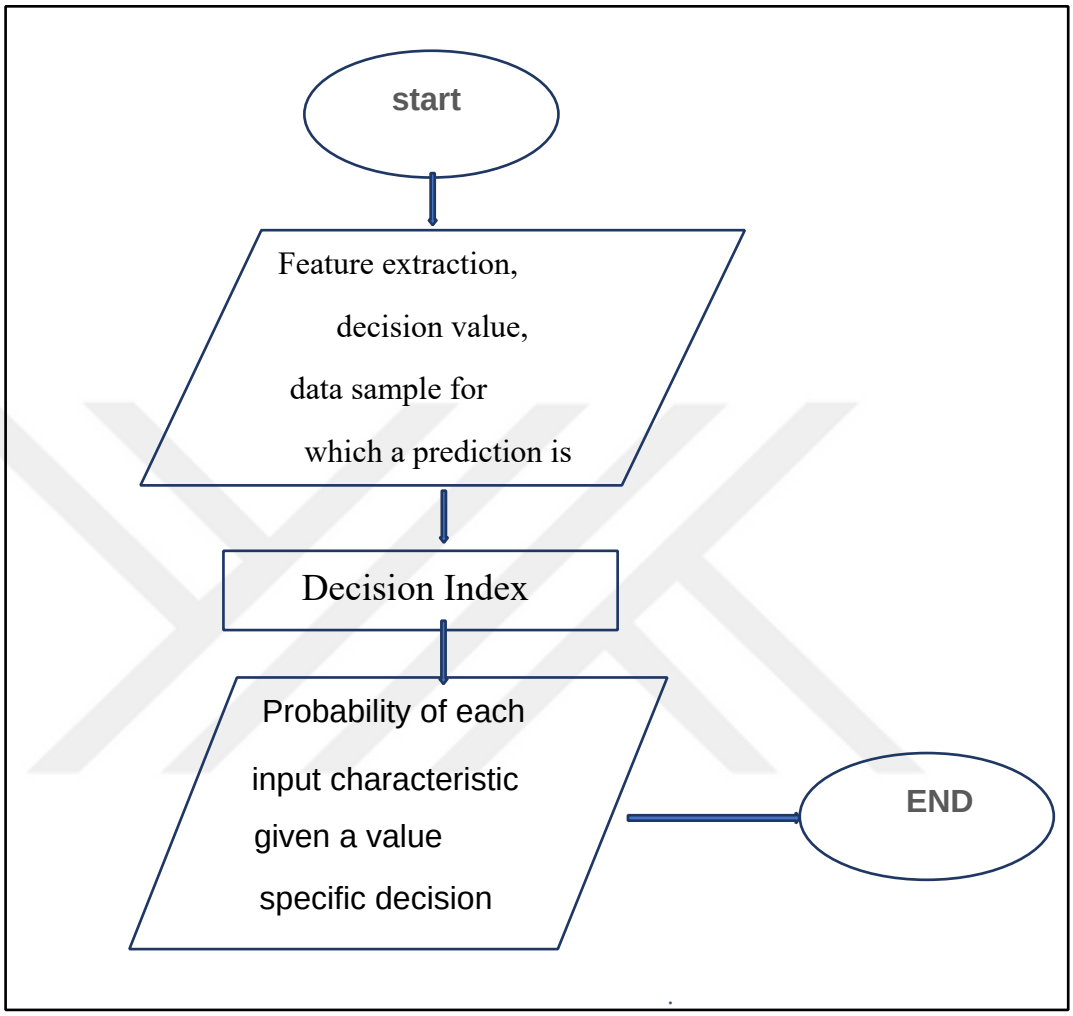

Figure 3.6: NB Flowchart.

f. Support Vector Machine:

SVM is a supervised ML classification technique commonly used in the field of cancer diagnosis and prognosis. SVM separates categories by selecting key samples from all categories called support vectors and generating a linear function that uses these support vectors to divide them as widely as possible

Support Vector Machines (SVM) was originally developed by Vapnik and colleagues in the 1960s based on the statistical learning theory of Vapnik & Chervonenkis in 1992. SVM has been successfully applied to many applications including handwriting

recognition, time series prediction, speech recognition, protein sequence problems, breast cancer diagnosis, etc.

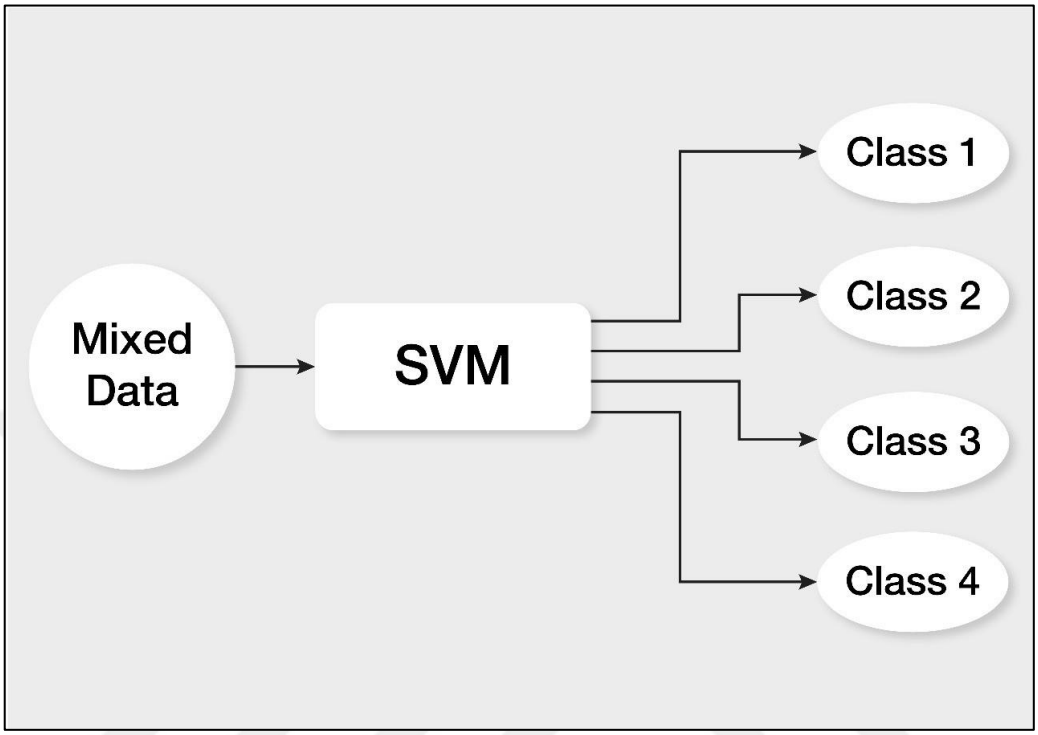

Figure 3.7: Principle of Classification in SVM (Vector Machine Support).

g. K-Nearest Neighbors:

K Nearest Neighbors is an intuitive supervised classification method often used in machine learning. This is a generalization of the nearest neighbor (NN) method. NN is a special case of KNN The K Nearest Neighbor (KNN) method has been widely used in data mining and machine learning applications due to its simple implementation and outstanding performance. However, it has been proven that setting all test data to the same value in the previous KNN method makes these methods impractical in practical applications.

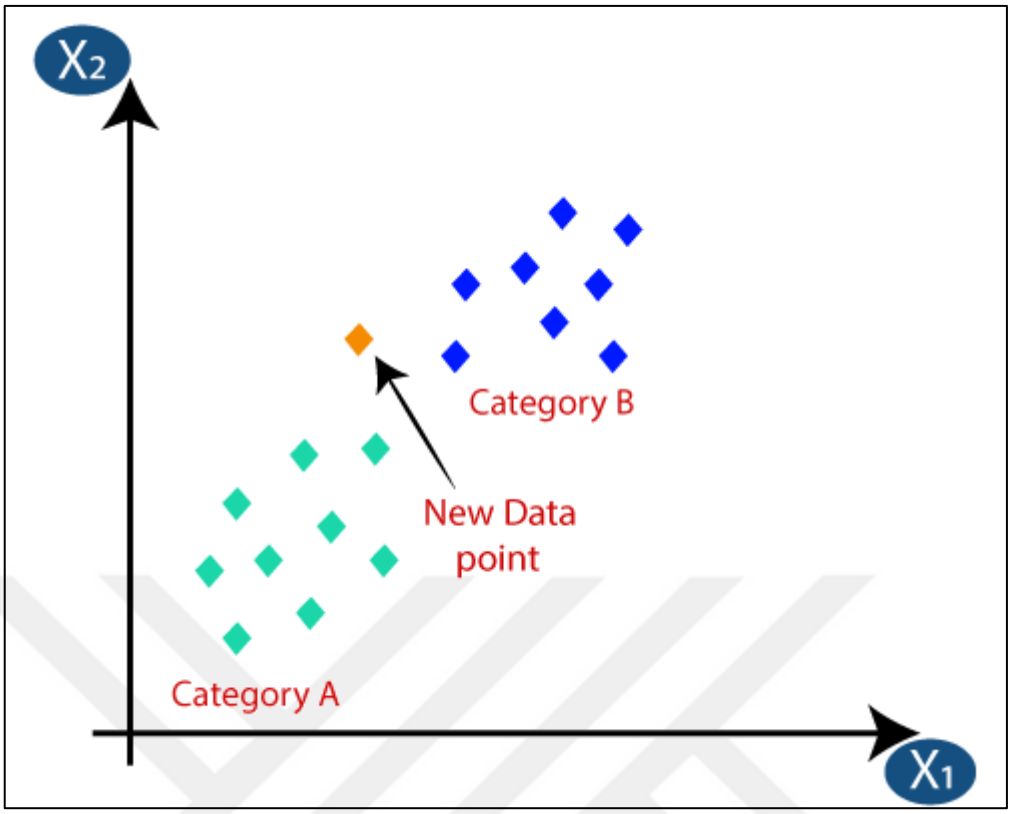

Figure 3.8: How K-NN Works For Machine Learning.

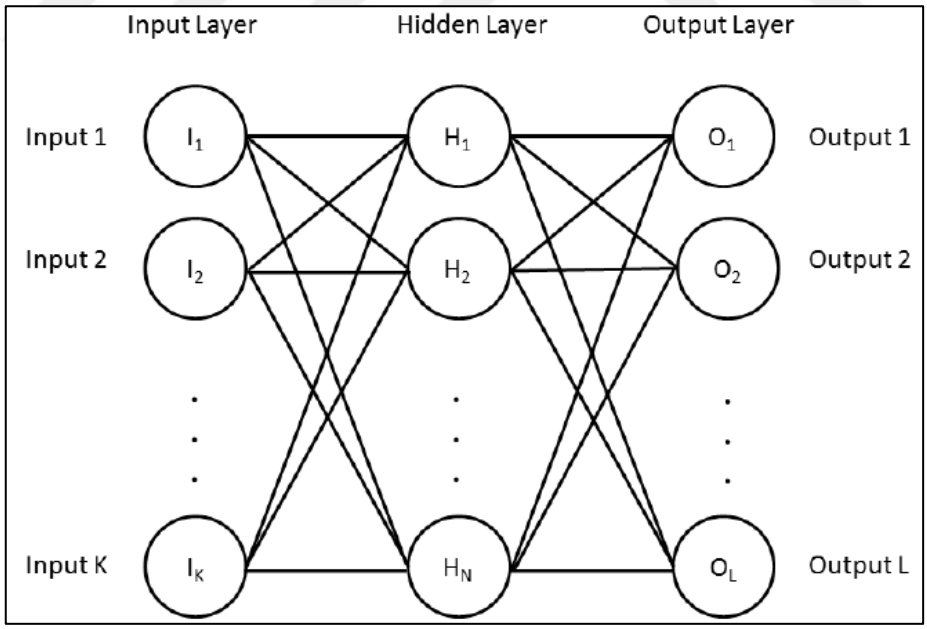

Figure 3.9: Schematic of a Three-Layer Feedforward Neural Network, With An Input Layer, A Hidden Layer, and an Output Layer.

3.3 REMINDER OF SOME SCIENTIFIC WORK ON BREAST CANCER:

This involves research work carried out in a biomedical field related to the classification of cancers, in particular breast cancer, using automatic learning algorithms (supervised learning).

Machine Learning Breast Cancer Diagnosis Data Logistic Regression Algorithm Research used a logistic regression algorithm to rank the data set of breast cancer patients, the data set has was obtained from the UCI Wisconsin standard. The author first used the 33 features of the dataset to train the model and eventually achieved 90% accuracy. Then the author, using a feature selection technique, extracted two main features from the 33, namely the maximum texture and the maximum perimeter to achieve an accuracy of 96.5%, which is a improvement compared to the result obtained from the 33 characteristics.

Mammogram data was used by logistic regression model to predict patient history risk factor, logistic regression prediction is used to check physician’s prognosis and is also used to correct incorrect predictions. The authors' work can help radiologists correctly diagnose breast cancer using mammography and referring to the patient's history.

Using Naïve Bayesian as a classifier on the Wisconsin dataset of 10 features tried to estimate the success and error of the classification and prediction algorithm when the data is chosen at random.

The Naïve Bayes model showed an approximate success rate of 85% to 95% and an error rate of 10 to 15% for classification and prediction.

4. PROPOSED METHOD

This chapter contains the various work phases in detail. Specifically, it describes the dataset used, the ways in which the images are characterized, the methods used for classifying the lesions and the tests performed.

4.1 DATASETS

The dataset consists of C-View synthetic mammography images from 153 patients who underwent the diagnostic exam. The images were acquired using the Dimensions Mammography system by Hologic and provided by the university of Wisconsin the dataset includes C-views of both breasts along the craniocaudal and medial lateral oblique views. All images have been made anonymous and refer to exams carried out between September 2014 and October 2022. The patients who underwent the exam were between 30 and 87 years old, with an average of 60 years.

The dataset used in the project is made up of a limited number of images compared to the dataset as some cases were excluded because they did not meet the following criteria:

a. Presence of both CC and MLO projections.
b. Presence within the image of the breast of at least one tumor lesion attributable to a single type (mass and opacity) clearly visible in both projections.
c. Presence within the breast image of lesions with a diameter of no more than 600 pixels.

Since each breast may contain more than one tumor, the images have been labeled and named according to the lesion present within them. In this way a dataset was obtained consisting of 414 images each including a single tumor lesion, half of which in CC projection and the other half in MLO projection. For each lesion, the class (benign or malignant) proven by histological examination or radiology diagnosis was provided. Available data include 293 benign and 121 malignant lesions. In addition to the class, the breast doctors who diagnosed the tumor also introduced the contours of the lesions into the dataset. The outlines have been colored red to be easily recognizable as can be seen in the Figure 4.1.

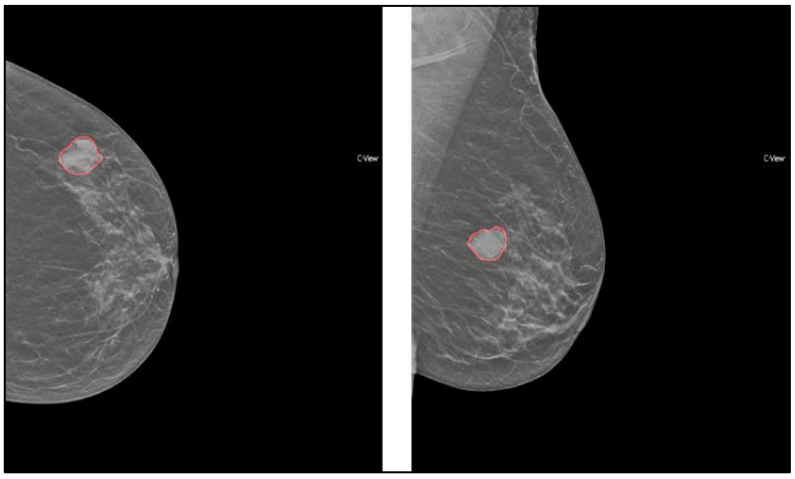

Figure 4.1: Example of two Projections (CC on The Left and MLO on the Right) in Which The Lesion is Highlighted Through The Red Outline Positioned by the Breast Specialists.

4.2 PIPELINES

The aim of the project is the creation of an automatic software that allows to predict the class to which a lesion belongs (benign or malignant) starting from the synthetic mammography image in which the lesion itself is present. All machine learning algorithms that have the purpose of classifying an object are based on a process called feature extraction This process makes it possible to transform the raw data (in this project represented by the C-View images) into numerical characteristics which make it possible to distinguish the two objects to be classified (benign and malignant lesions).

There are several methods of feature extraction, two different approaches are used in this work: an automatic one and a manual one.

a. In a first phase, an automatic extraction of the functionalities is carried out through the use of deep learning algorithms that do not require human intervention. In particular, four SVM architectures are implemented This extraction is performed only in the area of the image where the tumor is present, thus obtaining the peculiar elements that allow the characterization of the two types of lesions for the purpose of classification.
b. In the second phase of the project, manual functionalities are extracted which, on the basis of scientific research, belong to the first and second order descriptors and LBP (explained in more detail in the following paragraphs). These features are extracted on the entire breast to describe the nature of the tissue.

Once all the features have been extracted, a single vector is created that combines the two groups of features, useful for the characterization of both the lesion and the breast tissue The

feature vector thus created is given as input to a classifier based on a machine learning algorithm. Three types of classifiers are used, such as RF, SVM, and KNN, whose final classification is compared in order to allow the selection of the best algorithm in terms of performance [49].

During the automatic extraction phase, in addition to extracting the features, the input images were also classified, whose performances were compared with those of the ML algorithms trained with the entire set of features. This process allowed us to determine the impact of the nature of the breast tissue (described by the manual features) on the classification of the lesions.

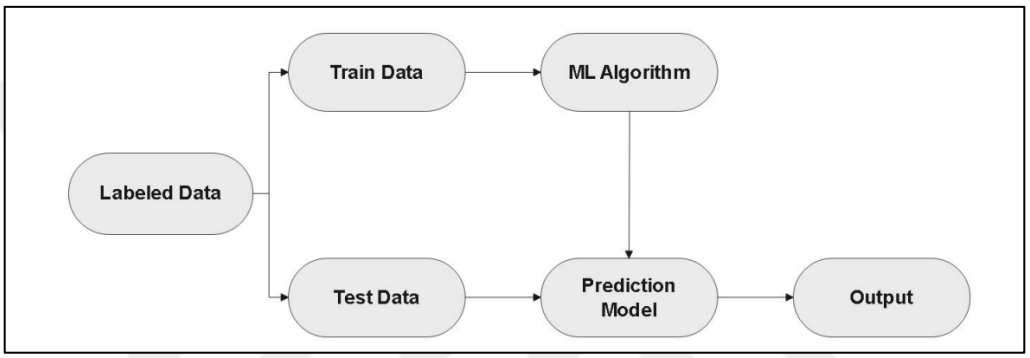

Figure 4.2: Flow Chart of the Project.

4.3 PREPROCESSING

Before the feature extraction phase, all the images are resampled, i.e. the pixel dimensions are varied in order to have comparable image quality. Resampling takes place via the Pixescaling value (a value that indicates the physical distance between the centers of two pixels of the image), a property present in every DICOM image. During the resampling process, higher resolution images lose quality. To limit this condition, it was decided to carry out a resampling with a Pixelscaling value equal to the median of the Pixelscaling of all the images. As can be seen from Fig 4.3, this value is equal to 0.1 mm/pixel.

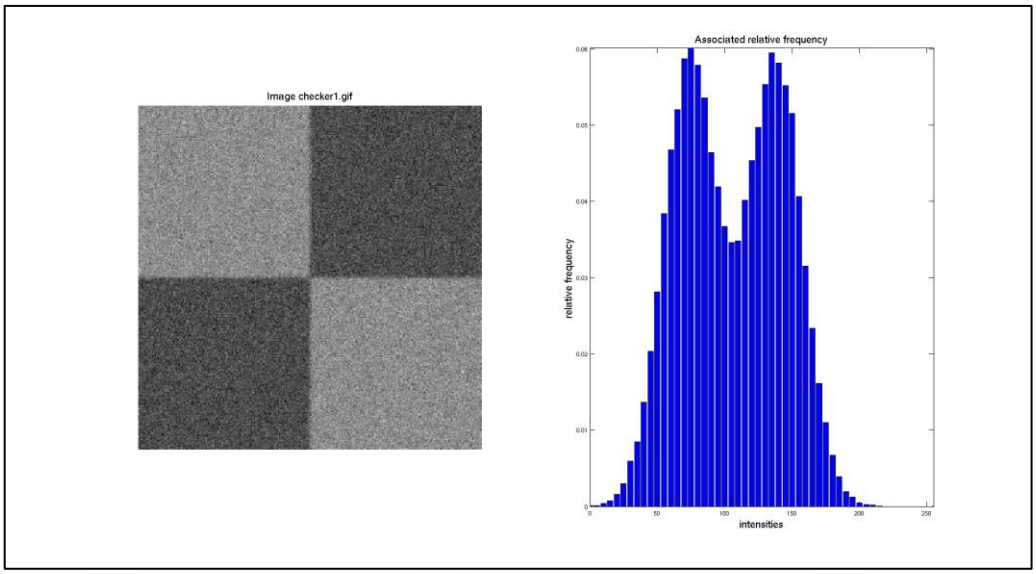

Figure 4.3: Histogram of Pixelscaling Values and Their Median Value.

Subsequently, depending on the features to be extracted (automatic on the lesions or manual on the breast), two different pre-processing steps are carried out.

a. In the first case, the image is cropped so as to identify only the area where the tumor is present. To do this, a rectangular bounding box is created around the manual outline (colored red) proposed by the breast specialists.

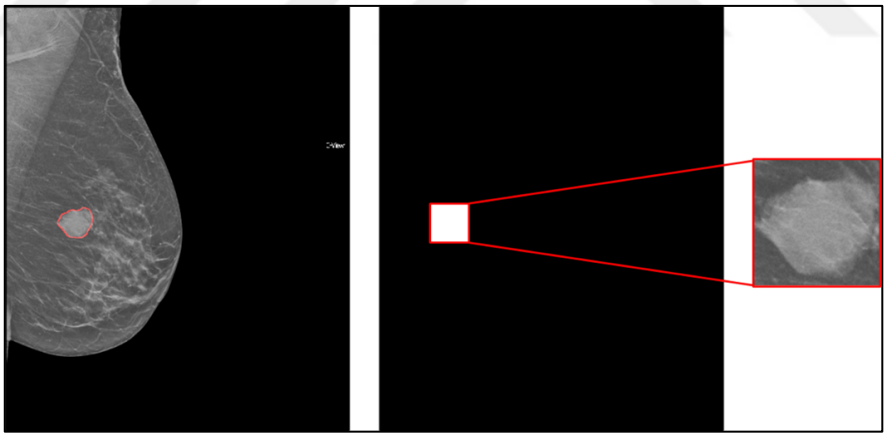

Figure 4.4: Preprocessing Applied to the Image to Locate The Lesion.

b. In the second case, all the parts of the image that do not contain the breast are eliminated, ie the background and in the case of MLO projections the pectoral and the upper part of the axillary cavity. To do this, a mask is created in which the white pixels represent the area of the breast on which the features are extracted. This area is identified through specific points that have been positioned and saved automatically by the acquisition system.

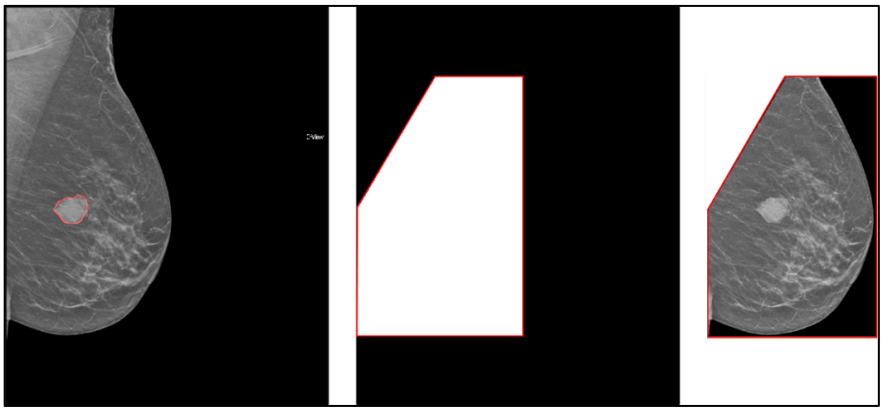

Figure 4.5: Preprocessing Applied to the Image to Locate The Breast.

4.4 LOCAL FEATURE EXTRACTION

The following paragraphs show the methods for extracting features on lesions using the automatic approach.

4.4.1 Image Resizing

Automatic feature extraction using convolutional networks requires homogeneity of input data sizes. For this reason, since the tumors have different sizes, it is necessary to resize the bounding boxes that enclose the lesions to a common size. In the figure 4.6the distribution of the values of the side of the bounding boxes, the value of the average dimension and the value chosen for the resizing, i.e. the power of two closest to the average value, are reported. The cropped images containing the tumors are then scaled to square ROIs of size 128x128 pixels.

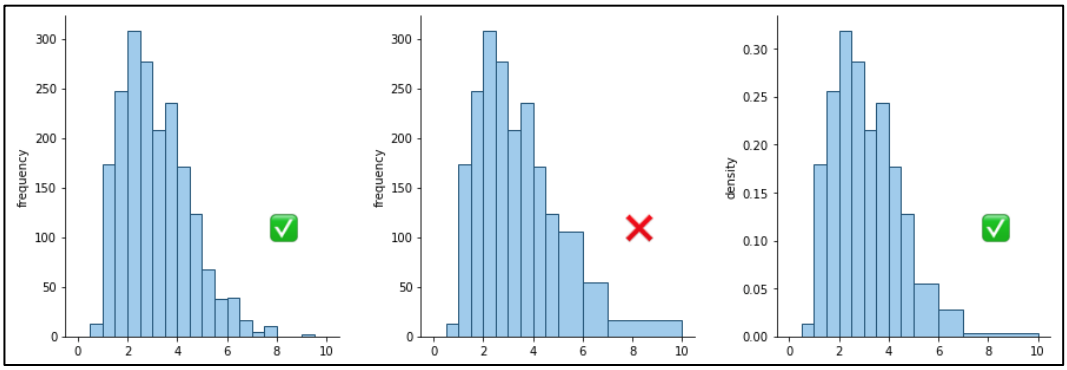

Figure 4.6: Histogram of Lesion Bounding-Box Dimensions. it Also Depicts the Average Size and the Selected One.

4.4.2 Construction and Division of the Dataset

Machine learning algorithms are based on dividing input data into three data sets: Training, Validation and Test sets. The split is necessary because you cannot evaluate the model on

the same data used during training. This is because we want to create an automatic system that can correctly classify not only the images present in the starting dataset, but also data other than those on which it has been trained. In doing so, the algorithm does not specialize on the input data, but manages to extract features and information that can be generalized to various new cases. The dataset is then divided into three groups with the following functions:

a. The model is initially adapted on a training data set called Training set. This set is formed by a series of examples that are used to adapt the parameters of the model (in the case of artificial neural networks it allows to adapt the synaptic connections between neurons). During the training phase, the system builds relationships between the input images and the class of the lesion present in the image. It then performs a comparison between the predicted and expected result on the basis of which the parameters will be adjusted.
b. Subsequently, the trained model is used to predict the outputs of a second dataset called a Validation dataset. This data set makes it possible to evaluate the quality of the trained system and its ability to generalize, or to correctly predict images other than those used in the training phase. If the performances on the Validation set are not high, the hyperparameters of the model are modified and the algorithm is trained again on the Training set, until the result on the validation set is not satisfactory.
c. Finally, the last data set is called a test set and is used to provide an unbiased assessment of a data set that the system has never seen. This allows to obtain performances similar to those that would be obtained with any new set of data.

The starting dataset composed of 414 lesions (293 benign and 121 malignant) was then divided into the three sets just described. The percentage of images in each set is as follows: 70% of the dataset is part of the Training set, 20% of the Validation set and the remaining 10% of the Test set. As can be seen, both from the percentages just listed and from the graph below, the training set is the most numerous since a large number of images is needed in the training phase to obtain an algorithm that has the ability to generalize. Furthermore, to avoid that the ratio between benign and malignant lesions is different in the various sets, the percentages are applied to the data of both classes separately (that is, the benign lesions are divided from the malignant ones and subsequently the subdivision is carried out for the two groups in sets)

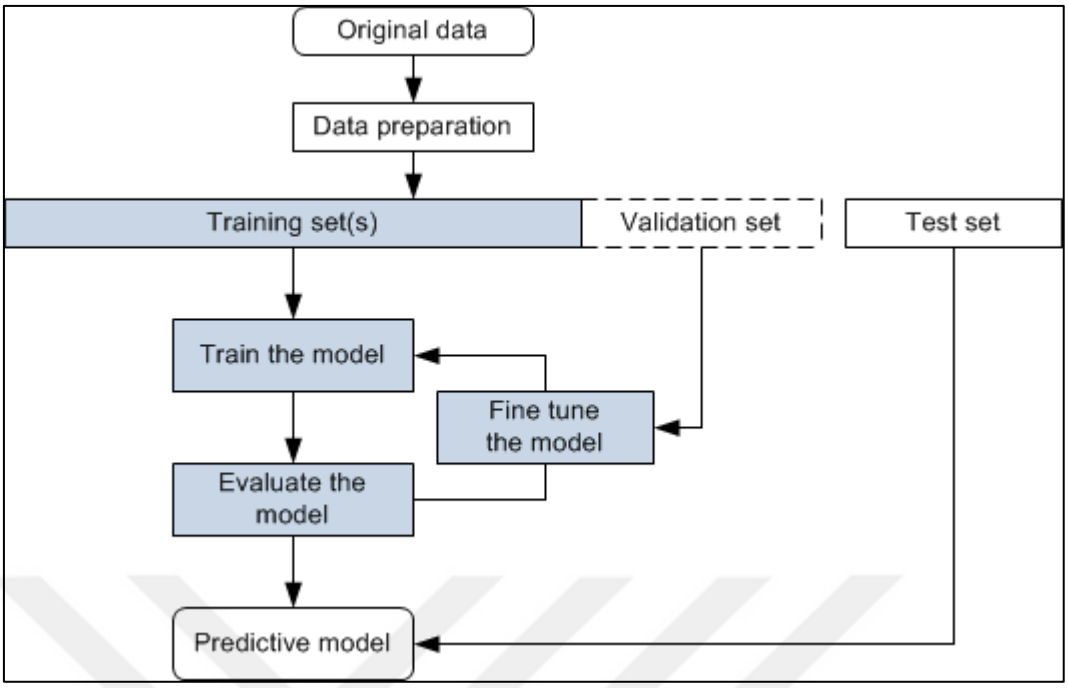

Figure 4.7: Flow-Chart of The Use of Datasets During Model Training and Validation.

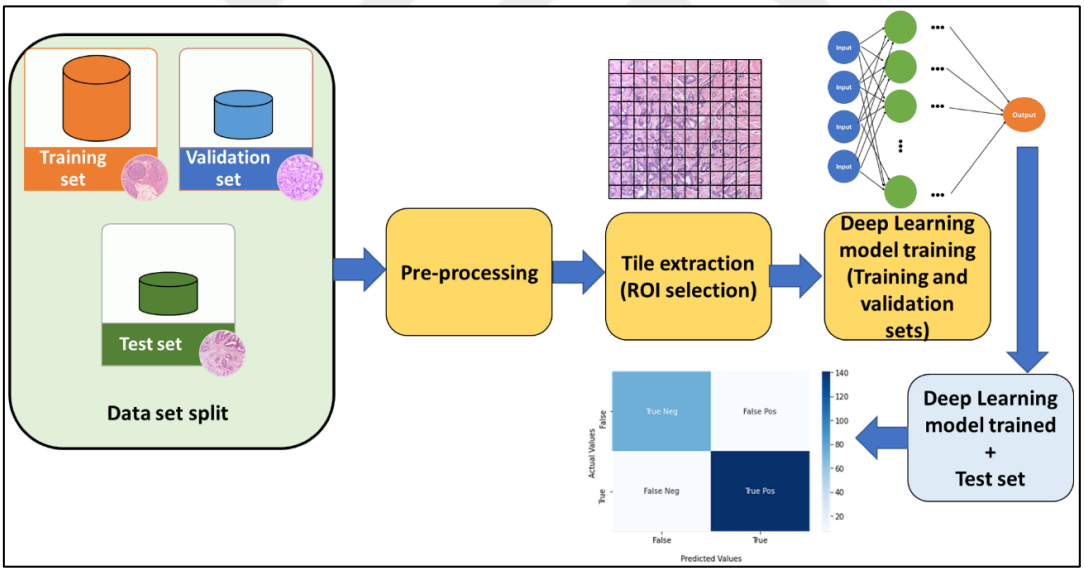

Figure 4.8: Percentages by Which The Dataset Split on the Left and Number of Lesions For Each Set on The Right.

4.4.3 Tuning of the Parameters

Starting from a model it is then possible to make various modifications both at the structural level and at the level of the hyperparameters (that is, all the variables of the network that can be initialized and optimized before the training phase). In this work it was decided not to make changes to the architectures (except for the classification layers that have been adapted

to our purpose), while numerous experiments were carried out to find the best combination of hyperparameters.

4.4.3.1 Cost function

As already mentioned, neural networks learn by comparing the output they predict with the real one. This comparison is expressed in mathematical terms through a function which is called the cost function. In order for the answer to be as exact as possible, this function must be minimized. There are different types of functions that can be used based on the predefined task, in this work cross-entropy was used, defined in mathematical terms by the following expression:

$$\frac{1}{2m}\sum_{1}^{m}(h(x^{(i)})-y^{(i)})^2 \tag{4.1}$$

where y is the vector of target outputs, a is the vector of outputs predicted by the network, n represents the number of inputs and c in number of classes (in our case equal to 2).

4.4.3.2 Learning rate

Another fundamental parameter for network training is the learning rate, also called learning rate. It determines the movement towards the minimum of the cost function which is carried out at each iteration. While the direction of descent is usually determined by the gradient of the loss function, the learning rate determines how large the step is in that direction. Because it affects the extent to which newly acquired information replaces old information, it metaphorically represents how quickly a machine learning model "learns". If the learning rate is very small, the model will converge too slowly, and if it is too large, the model will diverge. In order to achieve faster convergence,

In this work numerous workouts are carried out with different learning rate values, in particular we started from a learning rate of 0.001 and decreased it up to 0.0001.

4.4.3.3 Batch sizes

In the case of numerous datasets, often the network is not trained with all the images, but it is preferable to divide the set into smaller groups, called batches. The batch size then defines the number of samples that are propagated through the network. This allows you to train the network using less memory and faster. The limitation of this approach is that the smaller the lot, the less accurate the gradient estimate will be.

For the four architectures proposed above, it was decided to carry out three tests by dividing the dataset into batches containing respectively 8, 16 and 32 images.

4.4.4 Training and Overfitting

Following the initialization of the hyperparameters, the weights and the GPU setting (when available), the real training of the network on the Training set begins. In this phase it is of fundamental importance to understand what the possible prediction errors could be. These errors cause a poor generalization capacity of the network called the overfitting problem. In this situation, the model overlearns the Training set examples, extracting not only the fundamental information for classification, but also random details and noises that cause unsatisfactory performance on a new data set (Validation set). The figure 4.9 describes this situation in terms of loss, which decreases for both sets up to a certain point in which the Training function continues to decrease while that of the Validation set begins to grow.

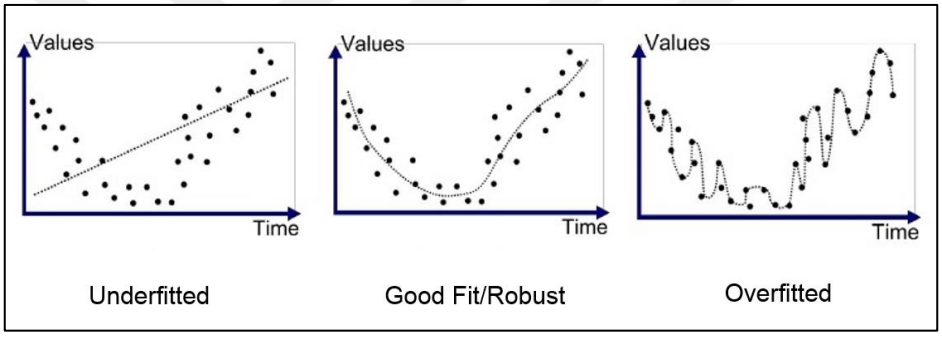

Figure 4.9: Example of Network Overfitting.

There are several approaches to limit overfitting, the main ones will be covered in the next paragraphs.

4.4.4.1 Date augmentation

Data augmentation, also called data augmentation, is a first method to limit overfitting and increase the generalization capacity of the algorithm [51] [52]. It consists in "artificially" increasing the number of datasets, applying transformations to the images (including rotation, flip, contrast or brightness variation, distortions, zoom, etc.) which allow for new examples during the training phase. This technique is widely used in the literature, especially to obtain better performance when having limited datasets.

In this work, two types of data augmentation are implemented: one online and one offline.

a. In the online technique, the images of each batch are transformed randomly, before each iteration during model training (therefore, copies of images are not created but the

original ones are modified). In this case, the transformed images are not saved to disk and it is impossible to know which image has been modified. The transformations that are applied with this technique are as follows: horizontal flip, vertical flip, contrast shift, and brightness shift all applied with a 50% chance to occur.

b. In the offline technique, data augmentation occurs across the entire dataset before training begins. The dataset thus created is saved on disk and will contain as many "copy" images as increases are made. The model is then trained through a dataset composed of many more images than the original one which allows to increase its robustness and limit overfitting. The transformations that are applied with this technique are the following: rotation of 90°, 180°, 270°, horizontal flip and vertical flip. The figure shows the two types of data augmentation applied to an image of the Training set.

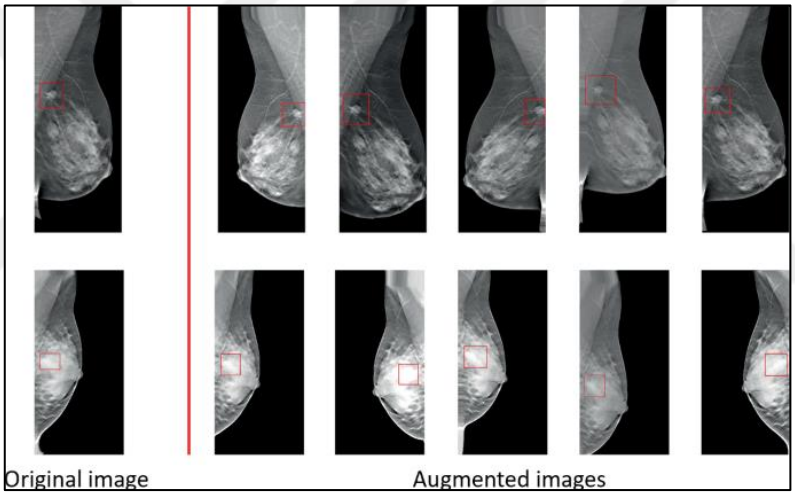

Figure 4.10: Example of Offline and Online Data Augmentation.

4.4.4.2 Number of epochs and early stop

Machine learning algorithms increase their performance as training time increases, i.e. as the number of training epochs increases. However, in the case of overfitting it is not efficient to train the algorithm on a very large number of epochs since at a certain point the performances will increase only on the Training set and no longer on the validation set. In this project, networks are trained for 50 eras, applying a particular technique called early stopping. This technique allows training to stop when performance on a validation set stops increasing over a number of epochs, thus preventing overfitting.

4.4.4.3 Regularization

A further technique which limits overfitting involves the introduction of a regularization term in the description of the cost function. This term is defined as the sum of the squares of the weights multiplied by a constant term lambda. The value of lambda determines the importance of the regularization term: the higher it is, the more the cost function is influenced by this term.

$$\text{Costfunction} = \text{Lossfunction} + \text{RegularizationTerm} \tag{4.2}$$

Thanks to this term, the loss function (and therefore the error committed by the network) increases if the weight values are very high. By reducing the value of the weights, you limit the importance of the Training set input data which can cause low generalization to other data sets. By doing so, overfitting is reduced and since the weights tend to zero, the complexity of the model is moderated.

4.4.4.4 Dropouts

Dropout is a technique that "ignores" certain neurons from the network during training. During each iteration of the Training, some neural units are randomly silenced, which causes a decrease of neurons working in the learning process. This process leads to a simpler network structure and, at the same time, a more robust model, since the learning of the weights cannot rely only on some specific neurons. The network will therefore be forced to give importance to all the neurons present, without becoming aware of some of them that could be silenced during training.

4.5 EXTRACTION OF GLOBAL FEATURES

The following paragraphs show the types of features extracted on breast images through the manual approach.

4.5.1 Features Extracted

The manual extraction of features in the area of the images where the breast is present takes place through texture analysis. This technique allows to describe the nature of the tissue through the information relating to the absolute value of the pixels and their distribution within the image. To extract the features, precise mathematical definitions, called "descriptors", are used, which are independent of the imaging modality. The descriptors that extract features starting from the native values of the images, called first order, are based on the study of the luminosity histogram, while the descriptors that analyze the relationships between two or more image pixels, called second order or higher , they are based on particular matrices on which the descriptors are subsequently calculated. The following paragraphs describe in detail the descriptors used in this project.

4.5.2 First Order Descriptors

The first order descriptors are obtained from the brightness histogram, i.e. a graph which shows the gray values on the x axis and the number of times that this value appears in the pixels of the image on the y axis and therefore describes the distribution of gray values within the breast. These descriptors are statistical operators which do not provide information on a morphological level, but make it possible to obtain an estimate of the brightness and homogeneity of the image.

4.5.2.1 Local binary patterns

Local Binary Patterns (LBP) are descriptors that investigate local contrast within the image. To do this, the 8-connected neighborhood of each pixel is analyzed and the value 1 is assigned if a pixel of the neighborhood has a value higher than that of the central pixel while the value 0 is assigned if it has a value lower than the central one. Subsequently, the values associated with the 8-connected set (zero and one) are read, clockwise or counterclockwise, first transforming them into a binary number and then into decimal coding. This is done for all the pixels of the image and the values obtained are averaged to obtain a single parameter.

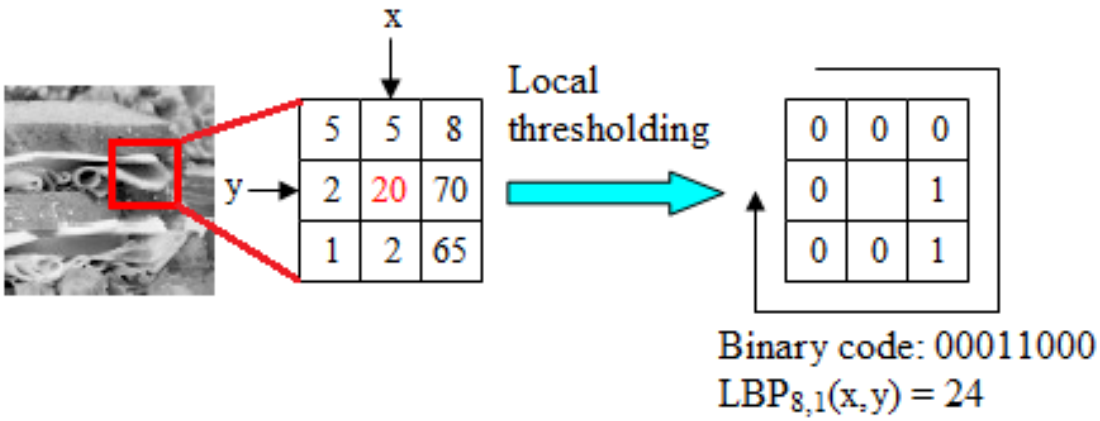

Figure 4.11: LBP Descriptors.

4.6 FEATURE SELECTION

Depending on the chosen architecture, there are a different number of features extracted, starting from 512 features extracted from SVM up to 4096 Given the large number of features extracted by deep learning compared to those extracted manually, it was decided to try to rebalance the two groups in order to give more weight to global features. For this reason it was decided to perform feature selection through the Random Forest technique applied exclusively to local features.

The Random Forest is a machine learning algorithm based on the set of many decision trees (or Decision Tree) whose outputs are combined to reach a single prediction. In turn, decision trees are supervised learning algorithms that are, however, subject to the problems of bias and overfitting. The power of the Random Forest consists in the combination of many trees which reduces the problems just mentioned. By merging different predictions, more accurate and less biased results are obtained, as each tree performs a different misclassification which, once averaged, will have less weight.

The problem of overfitting is, however, overcome by training the Decision Trees with different inputs. Trees don't use the entire dataset for training, but each algorithm is trained on a random sample of data from the starting set. The greater the number of trees, the less overfitting will be, however at the expense of performance on the Testing set. In this project, a number of decision trees equal to 5 was selected which made it possible to limit the overfitting, but at the same time did not worsen the performance on the Test set.

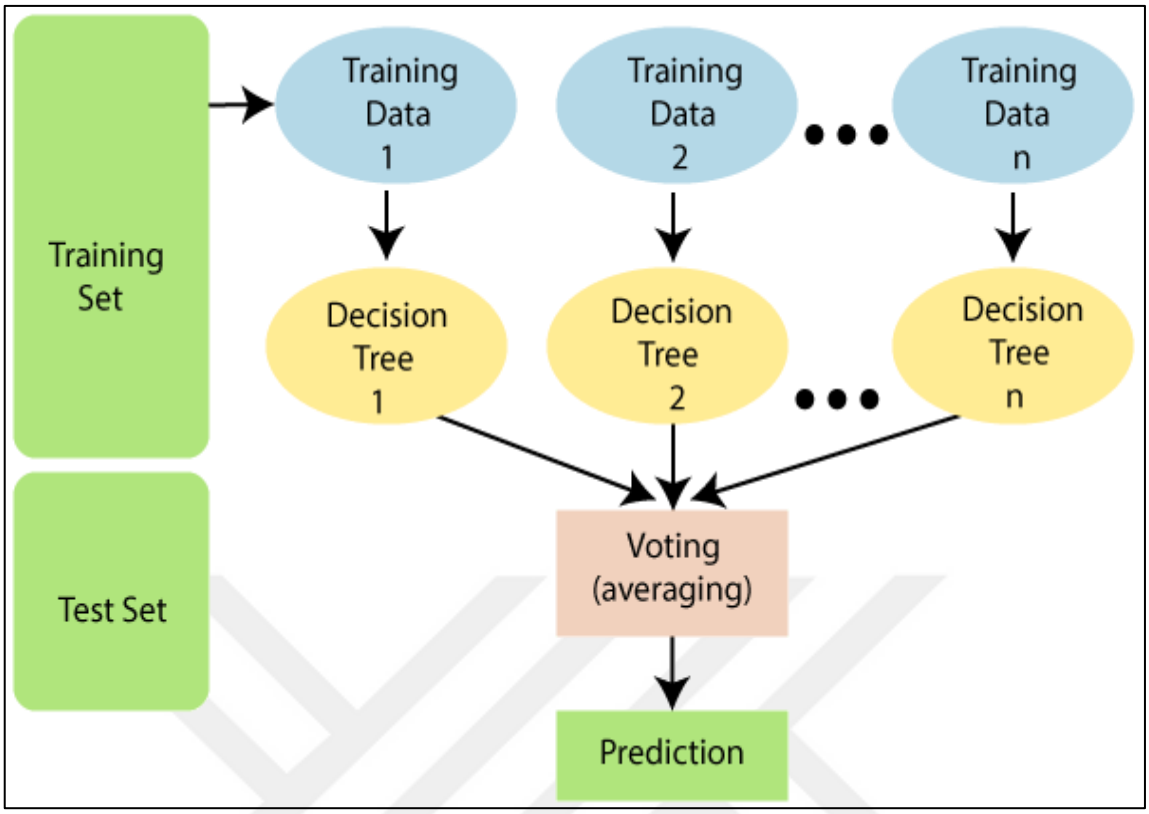

Figure 4.12: Random Forest Algorithm Structure.

The Random Forest, in addition to predicting the class, also determines an importance score associated with each variable. These scores make it possible to identify the features that according to the algorithm are of greater importance, eliminating the features that have little impact in the classification. In this work, an importance threshold has been set which varies between 0.1 and 0.2 in order to extract an average of 50 features for each set of features.

After selecting the most important features, these features were concatenated with the manual ones. Finally, all the characteristics have been normalized through the min-max scaling technique which allows to improve the direct comparison between values with very different quantities.

4.7 CLASSIFICATION METHODS

The last part of the project is to classify the features that were previously extracted. Classification is performed both for only deep features extracted from SVM architectures, and for merging deep features and manually extracted ones. In the first case, the classification takes place through a simple KNN while in the second case the two groups of functionalities are concatenated and classified through three machine learning algorithms, often used in the literature: an SVM, a KNN and an RF The different classification methods

have been chosen among those most commonly used in the literature and will be described in more detail in the next paragraphs. All algorithms have the common goal of classifying images into benign and malignant classes.

4.7.1 SVM

support-vector machines, is a supervised learning method used for the classification of features extracted both manually and through deep learning. This technique is based on the identification of a function that allows the separation of elements of different classes. The boundary that separates the classes is called the decision boundary. The model represents the input data as points in space, mapped such that elements belonging to two different classes are separable by as large a space as possible. While elements that are located away from the hyperplane are most likely classified in the correct class, elements close to it are at risk of being misclassified as new points are added.

The hyperplane equation is of the type wx - b = 0 and is found by solving an optimization problem with the aim of maximizing the distance between the support vectors.

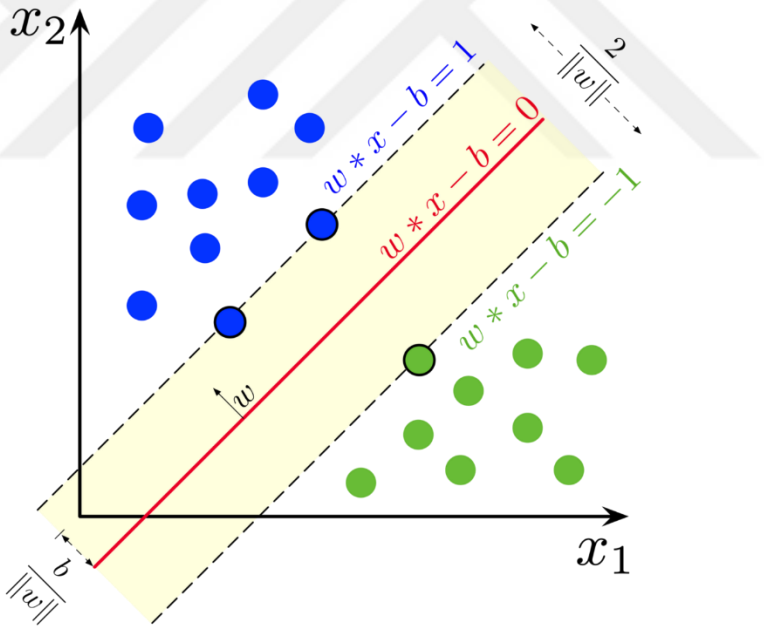

Figure 4.13: Example of Division Into Two Classes Using the SVM Classifier.

4.7.2 KNN Extension

The k-nearest neighbors (k-NN) is one of the simplest classification methods used in machine learning. It is an algorithm that classifies the data according to the characteristics of the objects close to the one under examination. To establish the proximity between the elements, the latter being represented as points in the multidimensional space. The

classification of a new element takes place on the basis of the most frequent class among the k closest points. Proximity is based on the calculation of the distance between points (usually the Euclidean distance is used, but other types of distance are equally usable).

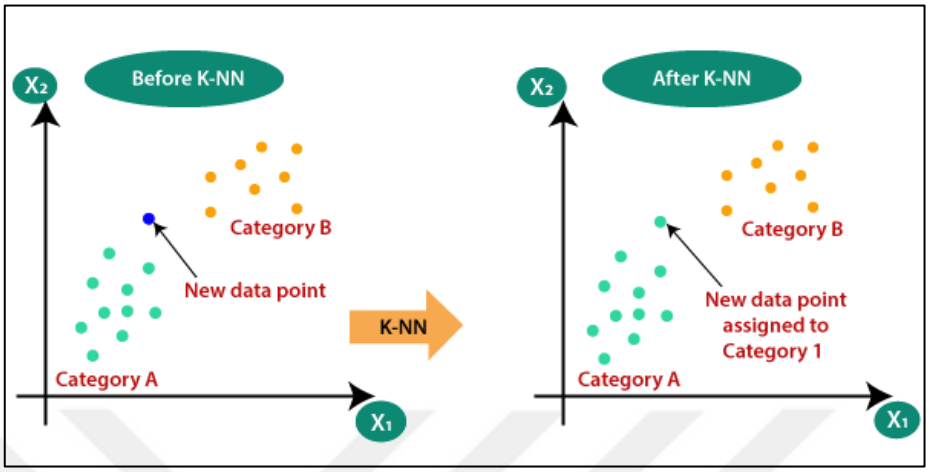

Figure 4.14: Example of Classification of A New Element Using The K-NN Classification Method.

Therefore, the parameter that is defined in the design phase is of fundamental importance. k is a typically not very large positive integer. If k were equal to one, then the object would be assigned to the class of the element closest to it. In a binary context in which there are only two classes, it is advisable to choose odd k to avoid ending up in situations of parity. Generally, as k increases, noise and overfitting decrease, but the selection criterion for the class becomes more unstable. Considering only the votes of the k neighboring objects there is the drawback due to the predominance of the classes with more objects. In this case it can be useful to weigh the contributions of the neighbors in order to give, in the calculation of the average, greater importance based on the distance from the considered object. The choice of k depends on the characteristics of the data. Finally, since this algorithm is based on distance, it is necessary to normalize the data because otherwise very different physical quantities can impact on the classification of the object. In this project the algorithm with k equal to 7 based on the Euclidean distance has been implemented.

4.7.3 RF

Random Forest (RF) is a method that aims to find a linear combination of characteristics that separates the classes under examination as much as possible. It is often used when the dataset is not large due to the algorithm's ability to generalize. To classify elements, this method projects input features from a higher dimensional space to a lower dimensional space. To do

this it is necessary to create a new axis on which the variables will be projected. To reduce dimensionality, the new axis must possess the following criteria: maximize the distance between the means of the two classes and minimize the variance within the single class. The mean value of each input for each of the classes can be calculated by dividing the sum of values by the total number of values, while the variance is calculated across all classes as the mean of the square of the difference of each value from the mean. RF models are based on the calculation of the probability that a new set of input data belongs to one of the output classes. The function used to calculate this probability is called the discriminant function, implemented as:

$$d(x)i = \log(f(x)i) + \log(pi) \tag{4.3}$$

In the classification phase, for each new datum the value of the function belonging to each class is calculated and the highest value represents the class to which it will be assigned.

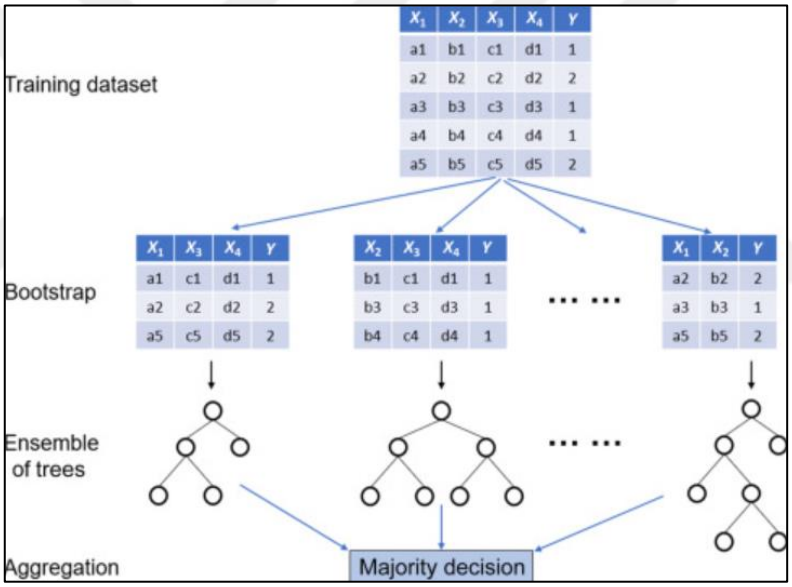

Figure 4.15: Example of Division Into Three Classes Using the RF Classifier.

4.8 EVALUATION METHODS

To numerically quantify the goodness of the classification results, the validation metrics discussed in the following paragraphs will be used. In general, in classification problems, the expected class is compared with the one predicted by the algorithm. The results will be compared with manual annotations, called ground truths, made by two expert breast specialists.

4.8.1 Confusion Matrix

In the realm of artificial intelligence, the confusion matrix or contingency table is the most commonly used validation statistic for measuring the performance of an algorithm. As shown in (Figıre 4.16), consists of a table with the actual class (actual condition) on the rows and the expected class (prediction condition) on the columns. The matrix has dimensions nxn, where n is the total number of classes in the dataset, which in this study is two: benign and malignant lesions. Positive (positive) and negative (negative) are the two classes in a binary classification. The positive class, which is related with the benign class in this project, is also indicated as 0, whereas the negative class, which is associated with the malignant class in this project, is indicated as 1.

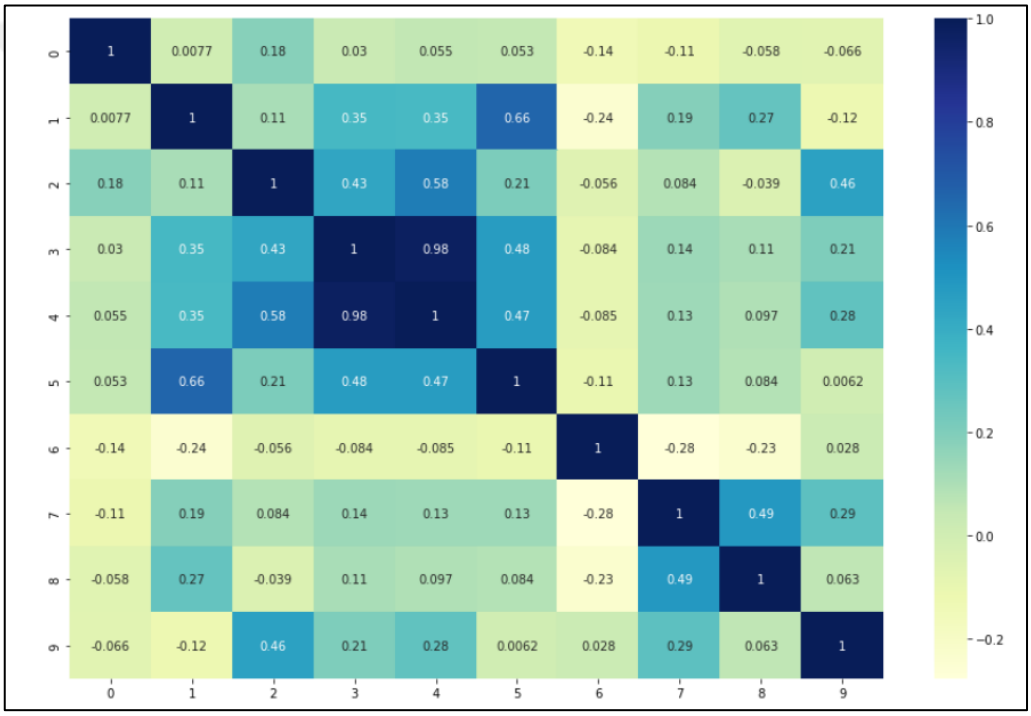

Figure 4.16: Confusion Matrix LBP.

The table shows:

a. The number of correct elements classified along the main diagonal. In particular, TN and TP which are, respectively, the correct classified for the positive class and the correct classified for the negative class;

b. The number of misclassified elements (that is, the classification errors committed by the algorithm) along the other diagonal. In particular, FN and FP which are, respectively, the cases that have been predicted negative even if in reality they are positive and vice versa.

Starting from the values present in the confusion matrix, various evaluation metrics can be calculated. The main ones are the following:

a. Accuracy: This is the most commonly used metric for a classification problem and represents the proportion of correct predictions out of the total.

$$\text{Accuracy} = \text{TP} + \text{TN TP} + \text{TN} + \text{FP} + \text{FN} \tag{4.4}$$

b. Recall/ Sensitivity/ True positive rate (TPR): represents the proportion of correctly classified positive predictions (TP) on the total of real positive instances.

$$\text{recall(TPR)} = \text{TP TP} + \text{FN} \tag{4.5}$$

c. Precision/Positive Predictive Value (PPV): Represents the proportion of positive predictions correctly classified (TP) out of the total positive predictions (both true and predicted).

$$\text{Precision(PPV)} = \text{TP TP} + \text{FP} \tag{4.6}$$

d. Specificity/True negative rate (TNR): represents the proportion of correctly classified negative predictions (TN) on the total negative real instances.

$$\text{Specificityà(TNR)} = \frac{\text{TN extension}}{\text{TN extension} + \text{FP}} \tag{4.7}$$

e. Negative predictive value (NPV): represents the proportion of correctly classified negative predictions (TN) on the total negative predictions (both real and predicted).

$$\text{NPV extension} = \text{TN TN} + \text{FN} \tag{4.8}$$

f. F1 score: represents the weighted harmonic mean of the Accuracy and Recall metrics.

$$\text{f1score} = \text{RecallPrecision Recall} + \text{Precision} \tag{4.9}$$

5. RESULTS

Table 5.1: Fitting Original Dataset (on the Default Algorithm Parameters, Training:80%, Testing:20%).

Module	Test Accuracy	F1_score	Specificity	sensitivity	AUC
SVC(RBF) :	70.8%	77.4%	66.7%	92.3%	68.9%
RF :	66.7%	71.4%	66.7%	76.9%	65.7%
LR :	58.3%	68.8%	57.9%	84.6%	55.9%
KNN :	75.0%	78.6%	73.3%	84.6%	74.1%

Table 5.2: Check Fitting Each Feature Alone.

Feature(s)	Module	Test Accuracy	F1_score	Specificity	sensitivity	AUC
Age	SVC(RBF) :	67%	71%	67%	77%	66%
BMI	SVC(RBF) :	63%	74%	59%	100%	59%
Insulin	SVC(RBF) :	63%	53%	83%	38%	65%
Resistin	SVC(RBF) :	63%	64%	67%	62%	63%
Glucose	SVC(RBF) :	58%	58%	64%	54%	59%
HOMA	SVC(RBF) :	58%	50%	71%	38%	60%
MCP.1	SVC(RBF) :	58%	72%	57%	100%	55%
Leptin	SVC(RBF) :	54%	70%	54%	100%	50%
Adiponectin	SVC(RBF) :	25%	25%	27%	23%	25%
HOMA	RF :	71%	70%	80%	62%	72%
Age	RF :	67%	71%	67%	77%	66%
Glucose	RF :	54%	56%	58%	54%	54%
Adiponectin	RF :	54%	62%	56%	69%	53%
Insulin	RF :	50%	57%	53%	62%	49%
BMI	RF :	46%	55%	50%	62%	44%
Leptin	RF :	46%	48%	50%	46%	46%
Resistin	RF :	42%	42%	45%	38%	42%
MCP.1	RF :	38%	48%	44%	54%	36%
BMI	LR :	67%	75%	63%	92%	64%
Glucose	LR :	63%	64%	67%	62%	63%
Resistin	LR :	63%	71%	61%	85%	60%
Age	LR :	54%	70%	54%	100%	50%
HOMA	LR :	54%	48%	63%	38%	56%
Leptin	LR :	54%	70%	54%	100%	50%
MCP.1	LR :	54%	70%	54%	100%	50%
Insulin	LR :	50%	45%	56%	38%	51%
Adiponectin	LR :	46%	61%	50%	77%	43%
Age	KNN :	75%	79%	73%	85%	74%
BMI	KNN :	71%	77%	67%	92%	69%
Insulin	KNN :	63%	64%	67%	62%	63%
Glucose	KNN :	58%	58%	64%	54%	59%
HOMA	KNN :	58%	58%	64%	54%	59%
Adiponectin	KNN :	54%	67%	55%	85%	51%
Resistin	KNN :	50%	57%	53%	62%	49%
MCP.1	KNN :	42%	53%	47%	62%	40%
Leptin	KNN :	25%	31%	31%	31%	24%

(MCP.1, Adiponectin, Leptin) in the lowest prediction priority in the most prediction algorithms)

Statistical Analysis:

a. Checking missing data.

b. Checking Duplicate Values.

c. Checking Negative Values.

d. Checking Normality:

e. Using (Shapiro Wilk tests): non-Parametric data.

f. Checking Correlation:

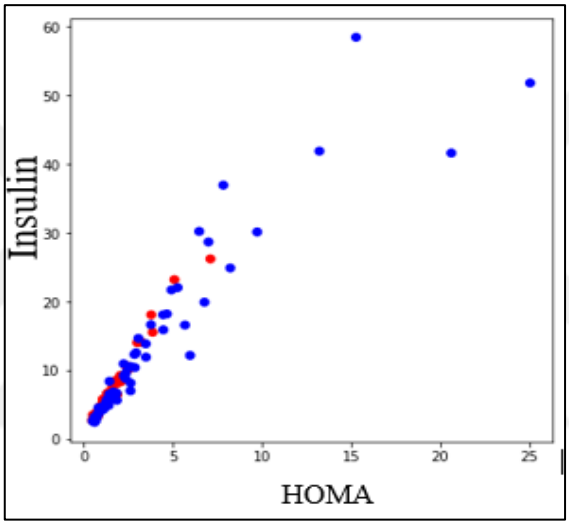

Figure 5.1: Correlation Between "Insulin" and "HOMA" 98% .

	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponectin	Resistin	MCP.1
Mann-Whitney U	1536.000	1433.500	783.500	1264.000	1127.000	1652.000	1610.000	1103.000	1543.000
Wilcoxon W	3616.000	3513.500	2161.500	2642.000	2505.000	3030.000	2988.000	2481.000	2921.000
Z	-0.711	-1.280	-4.891	-2.221	-2.981	-0.067	-0.300	-3.114	-0.672
Asymp. Sig. (2-tailed)	0.477	0.201	0.000	0.026	0.003	0.947	0.764	0.002	0.502

Figure 5.2: Correlation.

a. Fins Best parameters for algorithms (Support Vector Classifier(SVC), K-Nearest Neighbor(KNN), Random Forest(RF), Logistic Regression(LR)):

i. Training: 80%, Testing: 20%)

ii. SVC(kernel = 'rbf', C=2.333, gamma = 3, class_weight ='balanced', random_state = 0)

iii. KNN(n_neighbors = 8, weights = 'distance', algorithm = 'auto', leaf_size = 8, p = 1)

iv. RF(n_estimators = 3, criterion = 'entropy', max_features = 'auto', class_weight = 'balanced_subsample',random_state=140)

v. LR(C =1.5)

b. Excluded Insulin depending on high Correlation with HOMA %98 as appear in Spearman's Correlation:

i. Correlation between “Insulin” and “'HOMA”: 98%

c. Excluded (MCP.1, Adiponectin, Leptin) they are the lowest prediction priority in the most prediction algorithms as appear in table2 above, and additionally they have high p-value as appear in Mann-Whitney U Test:

i. Leptin: 0.947

ii. Adiponectin: 0.764

iii. MCP.1: 0.502

d. Add 2 new features depending on original features:

i. NewFeature1 = $(BMI)^4$ * Age

ii. NewFeature2 = $(Adiponectin)^2$ *Leptin

Table 5.3: Checking Mann-Whitney U Test.

Module	Test Accuracy	F1_score	Specificity	sensitivity	AUC
SVC(RBF) :	95.8%	96.0%	100.0%	92.3%	96.2%
KNN :	95.8%	96.0%	100.0%	92.3%	96.2%
RF :	95.8%	96.0%	100.0%	92.3%	96.2%
LR :	75.0%	80.0%	70.6%	92.3%	73.4%

6. CONCLUSIONS AND FUTURE DEVELOPMENTS

The best results, in terms of variability and mean value, are obtained with the SVM classifier applied to the concatenated feature vector. The combination of global features, descriptive of the nature of the breast tissue, with those extracted using deep learning algorithms, seem to bring a slight increase in performance compared to the classification obtained through local features alone. From the 95% accuracy achieved through the SVM architecture, it reaches 85.4% when the two types of features are combined. Even if the improvement is not high, it demonstrates that additional information beyond that which can be extracted from the lesion alone can contribute to the analysis of the tumour. This approach is typical of the clinical evaluation by doctors who combine the visual information obtained from the tumor area with that relating to the nature of the breast and the patient's medical record. For a possible future development, one can therefore think of integrating the analysis of the lesions carried out through deep learning with further data concerning both the nature of the breast tissue and the patient's clinical history.

This project is the continuation of a previous work that always had the objective of classifying the malignancy of lesions, but which focused on the development of machine learning algorithms trained on manual features. The techniques introduced in the new work have led to better performance than simple machine learning algorithms demonstrating that deep learning, as often reported in the literature, allows to achieve better results thanks to the higher level of abstraction of the features.

Following the analysis of false classifieds conducted various considerations were made. First, classifiers fail to recognize injuries that have not been well represented in the Training dataset. It is therefore advisable to create a dataset with good representability, i.e. the presence of a good number of examples for each type of tumor is necessary. In doing so, the algorithms do not have to recognize elements that are very different from those they have trained on. Of course it's not possibleintroduce any type of existing lesion into the dataset (due to the enormous anatomical variability), but the more complete the dataset, the better the classification performance will be.

Another consideration is the size of the lesions. It has been noted that algorithms often tend not to correctly classify very small or very large lesions. In the first case (ie when the lesions are small) the error was considered acceptable by the doctors as they themselves require further investigations when faced with lesions of this type. This occurs because it is not

always possible to identify the salient characteristics that allow discrimination due to the small size. In the second case (when very large lesions are considered) the error was considered important as very often this type of lesion appears more easily classifiable in their eyes. Since tumor size seems to play a crucial role in the reliability of the results, it will therefore be necessary to introduce this variable in a hypothetical future study. To avoid resizing all the lesions to a common size, different strategies can be implemented including the analysis of small areas of the image when the original size of the lesions is very large or the construction and training of different classifiers depending on the size of the tumors.

The software implemented in this study is based on clinicians finding and contouring the tumor within the image. Tests were carried out to evaluate whether small variations in the creation of the bounding box affected the performance of the classifier. It has been obtained that the algorithms are independent of these variations, a particularly important aspect which determines a high usability of the tool.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 71, no. 3, p. 212, 2021. 1

[2] A. C. Society, "Breast cancer facts & figures 2019–2020," Am. Cancer Soc, pp. 1–44, 2019. 1, 5

[3] C. P. A. Cancer, "Breast cancer screening in canada; monitoring & evaluation of quality indicators–results report, january 2011 to december 2012," 2017. 1, 4

[4] M. Løberg, M. L. Lousdal, M. Bretthauer, and M. Kalager, "Benefits and harms of mammography screening," Breast Cancer Research, vol. 17, no. 1, pp. 1–12, 2015. 1

[5] A. Chong, S. P. Weinstein, E. S. McDonald, and E. F. Conant, "Digital breast tomosynthesis: concepts and clinical practice," Radiology, vol. 292, no. 1, pp. 1–14, 2019. 1

[6] M. L. Marinovich, K. E. Hunter, P. Macaskill, and N. Houssami, "Breast cancer screening using tomosynthesis or mammography: a meta-analysis of cancer detection and recall," JNCI: Journal of the National Cancer Institute, vol. 110, no. 9, pp. 942–949, 2018. 1, 4

[7] D. B. Kopans, "Digital breast tomosynthesis from concept to clinical care," American Journal of Roentgenology, vol. 202, no. 2, pp. 299– 308, 2014. 1, 4, 6

[8] M. L. Giger, M. F. Inciardi, A. Edwards, J. Papaioannou, K. Drukker, Y. Jiang, R. Brem, and J. B. Brown, "Automated breast ultrasound in breast cancer screening of women with dense breasts: reader study of mammography-negative and mammography-positive cancers," American Journal of roentgenology, vol. 206, no. 6, pp. 1341–1350, 2016. 1

[9] K. Doi, “Computer-aided diagnosis in medical imaging: historical review, current status and future potential,” Computerized medical imaging and graphics, vol. 31, no. 4-5, pp. 198–211, 2007. 1

[10] M. L. Giger, N. Karssemeijer, and J. A. Schnabel, “Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer,” Annual review of biomedical engineering, vol. 15, pp. 327– 357, 2013. 1

[11] J.-Z. Cheng, Y.-H. Chou, C.-S. Huang, Y.-C. Chang, C.-M. Tiu, K.-W. Chen, and C.-M. Chen, “Computer-aided us diagnosis of breast lesions by using cell-based contour grouping,” Radiology, vol. 255, no. 3, pp. 746–754, 2010. 1

[12] B. Van Ginneken, C. M. Schaefer-Prokop, and M. Prokop, “Computeraided diagnosis: how to move from the laboratory to the clinic,” Radiology, vol. 261, no. 3, pp. 719–732, 2011. 1

[13] K. H. Allison, L. M. Reisch, P. A. Carney, D. L. Weaver, S. J. Schnitt, F. P. O’Malley, B. M. Geller, and J. G. Elmore, “Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel,” Histopathology, vol. 65, no. 2, pp. 240–251, 2014. 1

[14] L. J. Grimm, A. L. Anderson, J. A. Baker, K. S. Johnson, R. Walsh, S. C. Yoon, and S. V. Ghate, “Interobserver variability between breast imagers using the fifth edition of the bi-rads mri lexicon,” American Journal of Roentgenology, vol. 204, no. 5, pp. 1120–1124, 2015. 1

[15] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, “Deep learning for identifying metastatic breast cancer,” arXiv preprintar Xiv:1606.05718, 2016. 1

[16] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. S. Corrado, A. Darzi et al., “International evaluation of an ai system for breast cancer screening,” Nature, vol. 577, no. 7788, pp. 89–94, 2020. 1

[17] A. Ibrahim, P. Gamble, R. Jaroensri, M. M. Abdelsamea, C. H. Mermel, P.-H. C. Chen, and E. A. Rakha, “Artificial intelligence in digital breast pathology: techniques and applications,” The Breast, vol. 49, pp. 267– 273, 2020. 1, 7

[18] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, http://www.deeplearningbook.org. 1, 2, 3

[19] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3d deep learning on medical images: a review," Sensors, vol. 20, no. 18, p. 5097, 2020. 1

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, pp. 1097–1105, 2012. 1

[21] K. Suzuki, "Overview of deep learning in medical imaging," Radiological physics and technology, vol. 10, no. 3, pp. 257–273, 2017. 1, 7

[22] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights into imaging, vol. 9, no. 4, pp. 611–629, 2018. 1, 2, 3

[23] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai et al., "Recent advances in convolutional neural networks," Pattern Recognition, vol. 77, pp. 354–377, 2018. 2

[24] F. Amherd and E. Rodriguez, "Heatmap-based object detection and tracking with a fully convolutional neural network," arXiv preprint arXiv:2101.03541, 2021. 2

[25] S. Zheng, Y. Song, T. Leung, and I. Goodfellow, "Improving the robustness of deep neural networks via stability training," arXiv preprint arXiv:1604.04326, 2016. 2

[26] A. LeNail, "Nn-svg: Publication-ready neural network architecture schematics," Journal of Open Source Software, vol. 4, no. 33, p. 747, 2019. [Online]. Available: https://doi.org/10.21105/joss.00747 2, 3

[27] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," arXiv preprint arXiv:2106.11342, 2021. 2

[28] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015. 2

[29] J. Wu, “Introduction to convolutional neural networks,” National Key Lab for Novel Software Technology. Nanjing University. China, vol. 5, no. 23, p. 495, 2017. 2

[30] G. Kang, K. Liu, B. Hou, and N. Zhang, “3d multi-view convolutional neural networks for lung nodule classification,” PloS one, vol. 12, no. 11, p. e0188290, 2017. 2

[31] J. D. Hunter, “Matplotlib: A 2d graphics environment,” Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007. 2, 3

[32] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, pp. 436–444, 2015. 2, 3

[33] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” arXiv preprint arXiv:1710.05941, 2017. 2

[34] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation functions: Comparison of trends in practice and research for deep learning,” arXiv preprint arXiv:1811.03378, 2018. 2

[35] M. Sun, Z. Song, X. Jiang, J. Pan, and Y. Pang, “Learning pooling for convolutional neural network,” Neurocomputing, vol. 224, pp. 96–104, 2017. 3

[36] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, “Deep learning in radiology: an overview of the concepts and a survey of the state of the art,” arXiv preprint arXiv:1802.08717, 2018. 3

[37] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv preprint arXiv:1609.04747, 2016. 3

[38] X. Ying, “An overview of overfitting and its solutions,” in Journal of Physics: Conference Series, vol. 1168, no. 2. IOP Publishing, 2019, p. 022022. 4

[39] W. S. Sarle et al., “Stopped training and other remedies for overfitting,” Computing science and statistics, pp. 352–360, 1996. 4

[40] Z. Liu, M. Sun, T. Zhou, G. Huang, and T. Darrell, “Rethinking the value of network pruning,” arXiv preprint arXiv:1810.05270, 2018. 4

[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014. 4

[42] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” Journal of Big data, vol. 3, no. 1, pp. 1–40, 2016. 4

[43] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” in International conference on artificial neural networks. Springer, 2018, pp. 270–279. 4

[44] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, “A survey on deep learning techniques for image and video semantic segmentation,” Applied Soft Computing, vol. 70, pp. 41–65, 2018. 4, 5, 6

[45] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 233–240. 4

[46] L. A. Jeni, J. F. Cohn, and F. De La Torre, “Facing imbalanced data–recommendations for the use of performance metrics,” in 2013 Humaine association conference on affective computing and intelligent interaction. IEEE, 2013, pp. 245–251. 4, 7

[47] D. Chicco and G. Jurman, “The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation,” BMC genomics, vol. 21, no. 1, pp. 1–13, 2020. 4, 7

[48] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” Journal of Big Data, vol. 6, no. 1, pp. 1–54, 2019. 4, 7

[49] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, “Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports,” Academic radiology, vol. 11, no. 2, pp. 178– 189, 2004. 4

[50] M. El Adoui, S. A. Mahmoudi, M. A. Larhmam, and M. Benjelloun, “Mri breast tumor segmentation using different encoder and decoder cnn architectures,” Computers, vol. 8, no. 3, p. 52, 2019. 4

[51] N. Dhungel, G. Carneiro, and A. P. Bradley, “A deep learning approach for the analysis of masses in mammograms with minimal user intervention,” Medical image analysis, vol. 37, pp. 114–128, 2017. 5

[52] M. A. Al-Antari, M. A. Al-Masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, “A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification,” International journal of medical informatics, vol. 117, pp. 44–54, 2018. 5

[53] H. Chougrad, H. Zouaki, and O. Alheyane, “Deep convolutional neural networks for breast cancer screening,” Computer methods and programs in biomedicine, vol. 157, pp. 19–30, 2018. 5

[54] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, “Detecting and classifying lesions in mammograms with deep learning,” Scientific reports, vol. 8, no. 1, pp. 1–7, 2018. 5

[55] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig et al., “Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network,” Expert Systems with Applications, vol. 139, p. 112855, 2020. 5

[56] W. T. Tran, A. Sadeghi-Naini, F.-I. Lu, S. Gandhi, N. Meti, M. Brackstone, E. Rakovitch, and B. Curpen, “Computational radiology in breast cancer screening and diagnosis using artificial intelligence,” Canadian Association of Radiologists Journal, vol. 72, no. 1, pp. 98–108, 2021. 4, 7

[57] E. F. Conant, W. E. Barlow, S. D. Herschorn, D. L. Weaver, E. F. Beaber, A. N. Tosteson, J. S. Haas, K. P. Lowry, N. K. Stout, A. Trentham-Dietz et al., “Association of digital breast tomosynthesis vs digital mammography with cancer detection and recall rates by age and breast density,” JAMA oncology, vol. 5, no. 5, pp. 635–642, 2019. 4

[58] E. A. Rafferty, M. A. Durand, E. F. Conant, D. S. Copit, S. M. Friedewald, D. M. Plecha, and D. P. Miller, “Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts,” Jama, vol. 315, no. 16, pp. 1784–1786, 2016.

[59] K. P. Lowry, R. Y. Coley, D. L. Miglioretti, K. Kerlikowske, L. M. Henderson, T. Onega, B. L. Sprague, J. M. Lee, S. Herschorn, A. N. Tosteson et al., “Screening performance of digital breast tomosynthesis vs digital mammography in community practice by patient age, screening round, and breast density,” JAMA network open, vol. 3, no. 7, pp. e2011792–e2011792, 2020. 4

[60] E. F. Conant, E. F. Beaber, B. L. Sprague, S. D. Herschorn, D. L. Weaver, T. Onega, A. N. Tosteson, A. M. McCarthy, S. P. Poplack, J. S. Haas et al., “Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the prospr consortium,” Breast cancer research and treatment, vol. 156, no. 1, pp. 109–116, 2016. 4

[61] X.-A. Phi, A. Tagliafico, N. Houssami, M. J. Greuter, and G. H. de Bock, “Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts–a systematic review and metaanalysis,” BMC cancer, vol. 18, no. 1, pp. 1–9, 2018. 4

[62] P. Pattacini, A. Nitrosi, P. Giorgi Rossi, V. Iotti, V. Ginocchi, S. Ravaioli, R. Vacondio, L. Braglia, S. Cavuto, C. Campari et al., “Digital mammography versus digital mammography plus tomosynthesis for breast cancer screening: the reggio emilia tomosynthesis randomized trial,” Radiology, vol. 288, no. 2, pp. 375–385, 2018. 4

[63] J. A. Baker and J. Y. Lo, “Breast tomosynthesis: state-of-the-art and review of the literature,” Academic radiology, vol. 18, no. 10, pp. 1298– 1310, 2011. 4

[64] J. V. Horvat, D. M. Keating, H. Rodrigues-Duarte, E. A. Morris, and V. L. Mango, “Calcifications at digital breast tomosynthesis: imaging features and biopsy techniques,” Radiographics, vol. 39, no. 2, pp. 307–318, 2019. 4

[65] S. H. Kim, H. H. Kim, and W. K. Moon, “Automated breast ultrasound screening for dense breasts,” Korean journal of radiology, vol. 21, no. 1, pp. 15–24, 2020. 5

[66] H. J. Shin, H. H. Kim, and J. H. Cha, “Current status of automated breast ultrasonography,” Ultrasonography, vol. 34, no. 3, p. 165, 2015. 5

[67] C. D. Lehman, J. M. Lee, W. B. DeMartini, D. S. Hippe, M. H. Rendi, G. Kalish, P. Porter, J. Gralow, and S. C. Partridge, “Screening mri in women with a personal history of breast cancer,” Journal of the National Cancer Institute, vol. 108, no. 3, p. djv349, 2016. 5

[68] R. M. Mann, N. Cho, and L. Moy, “Breast mri: state of the art,” Radiology, vol. 292, no. 3, pp. 520–536, 2019. 5

[69] H. Greenspan, G. Oz, N. Kiryati, and S. Peled, “Mri inter-slice reconstruction using super-resolution,” Magnetic resonance imaging, vol. 20, no. 5, pp. 437–446, 2002. 5

[70] R. Z. Shilling, T. Q. Robbie, T. Bailloeul, K. Mewes, R. M. Mersereau, and M. E. Brummer, “A super-resolution framework for 3-d highresolution and high-contrast imaging using 2-d multislice mri,” IEEE transactions on medical imaging, vol. 28, no. 5, pp. 633–644, 2008. 5

[71] M. Yousefi, A. Krzyzak, and C. Y. Suen, “Mass detection in dig-˙ ital breast tomosynthesis data using convolutional neural networks and multiple instance learning,” Computers in biology and medicine, vol. 96, pp. 283–293, 2018. 6, 7

[72] D. H. Kim, S. T. Kim, J. M. Chang, and Y. M. Ro, “Latent feature representation with depth directional long-term recurrent learning for breast masses in digital breast tomosynthesis,” Physics in Medicine & Biology, vol. 62, no. 3, p. 1009, 2017. 5, 6

[73] S. Liu, D. Xu, S. K. Zhou, T. Mertelmeier, J. Wicklein, A. Jerebko, S. Grbic, O. Pauly, W. Cai, and D. Comaniciu, “3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes,” arXiv preprint arXiv:1711.08580, 2017. 5, 6

[74] Y. Zhang, X. Wang, H. Blanton, G. Liang, X. Xing, and N. Jacobs, “2d convolutional neural networks for 3d digital breast tomosynthesis classification,” in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2019. 5, 6, 7

[75] G. Liang, X. Wang, Y. Zhang, X. Xing, H. Blanton, T. Salem, and N. Jacobs, “Joint 2d-3d breast cancer classification,” in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2019. 5, 6

[76] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv preprint arXiv:1512.03385, 2015. 5, 6

[77] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in International Conference on Learning Representations, 2015. 5

[78] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Computer Vision and Pattern Recognition (CVPR), 2016. 5

[79] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” CVPR, 2017. 5

[80] X. Zhang, Y. Zhang, E. Y. Han, N. Jacobs, Q. Han, X. Wang, and J. Liu, “Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks,” IEEE transactions on nanobioscience, vol. 17, no. 3, pp. 237–242, 2018. 7

[81] M. Fan, H. Zheng, S. Zheng, C. You, Y. Gu, X. Gao, W. Peng, and L. Li, “Mass detection and segmentation in digital breast tomosynthesis using 3d-mask region-based convolutional neural network: A comparative analysis,” Frontiers in molecular biosciences, vol. 7, 2020. 6, 7