ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

OBTAIN ANTERIOR/POSTERIOR POSITION OF THE TUMOR THROUGH MACHINE LEARNING

M.Sc. THESIS

Golshan GHOLAMPOUR

Department of Electronics & Communications Engineering

Biomedical Engineering Program

MAY 2023

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

OBTAIN ANTERIOR/POSTERIOR POSITION OF THE TUMOR THROUGH MACHINE LEARNING

M.Sc. THESIS

Golshan GHOLAMPOUR
(504201407)

Department of Electronics & Communications Engineering

Biomedical Engineering Program

Thesis Advisor: Prof. Dr. Ibrahim AKDUMAN

MAY 2023

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MAKİNE ÖĞRENME YOLUYLA TÜMÖRÜN ANTERİOR/POSTERİOR POZİSYONUNU ELDE EDİN

YÜKSEK LİSANS TEZİ

Golshan GHOLAMPOUR
(504201407)

Elektronik & Haberlesme Mühendisliği

Biomedikal Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Ibrahim AKDUMAN

MAYIS 2023

Golshan-Gholampour, a M.Sc. student of İTU Graduate School student ID 504201407, successfully defended the thesis/dissertation entitled “OBTAIN ANTERIOR/POSTERIOR POSITION OF THE TUMOR THROUGH MACHINE LEARNING”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Ibrahim AKDUMAN** İstanbul Technical University	
Co-advisor :	**Prof. Dr.**	
Jury Members :	**Prof. Dr. Ibrahim AKDUMAN** Istanbul Technical University	
	Prof. Dr. Hülya ŞAHINTURK Yildiz Technical University	
	Prof. Dr. Mehmet ÇAYOREN Istanbul Technical University	
(If exists)	**Prof. Dr. Name SURNAME** ..... University	.....
(If exists)	**Prof. Dr. Name SURNAME** ..... University	.....

Date of Submission : 25 May 2023
Date of Defense : 25 July 2023

To my parents
Roya Abazari
And, Kheirollah Gholampour
Who always believed in me
Even when my weak soul was withered
They always saw light in me,
Even when darkness was my only breath
If today I am capable to stand up
It is because the always healed my broken bones
With their endless love, support, and encouragement,
I owe you for the person I am now

TABLE OF CONTENTS

Page

ABBREVIATIONS

WHO : World Health Organization
MRI : Magnetic Resonance Imaging
PET : Positron Emission Tomography
MWI : Microwave Imaging
MG : Mammography
SAFE : Scan and Find Early
ML : Machine Learning
SVM : Support Vector Machine
KNN : K Nearest Neighbors
IDC : Invasive Ductal Carcinoma
ILC : Invasive Lobular Carcinoma
DCIS : Ductal Carcinoma in Situ
LCIS : Ductal Carcinoma in Situ
AEC : Automatic Exposure Control
MLO : Mediolateral Oblique
CC : Cranial Caudal
ACS : American Cancer Society
NCCN : National Comprehensive Cancer Coalition Network
18F : Fluorodeoxyglucose
CT : Computed Tomography
MARIA : Multistatic Array Processing for Radiowave Imaging Acquisition
NN : Neural Networks
IG : Information Gain
CSV : Comma Separated Values
AUC : Area Under Curve
TPR : True Positive Rate
TP : True Positive
TN : True Negative
FP : False Positive
FN : False Negative
CA : Classification Accuracy

SYMBOLS

kVp **:** kilo Voltage peak

$\boldsymbol{\mu m}$ **:** Mictometer

LIST OF TABLES

Page

LIST OF FIGURES

Page

OBTAIN ANTERIOR/POSTERIOR POSITION OF THE TUMOR THROUGH MACHINE LEARNING

Summary

Breast cancer is a disease in which an uncontrolled growth of cells in the breast happens. It can initiate in different sections of the breast and based on which cells are involved and turned to cancer, how it spreads, and how it behaves the type of breast cancer will be identified. Breast cancer can grow in fatty tissue and lobules or spread over the blood or lymph system to other organs and cause metastasis. After surpassing lung cancer in 2020 breast cancer became the most common cancer worldwide. It is a leading cause of cancer death in less developed countries and the second leading cause of cancer death in American women. According to World Health Organization (WHO), of almost 2.3 million women diagnosed with breast cancer globally, 685,000 died. The number of diagnosed patients and mortality have increased by more than 20 and 14 percent respectively since 2008. On the other hand, Survival following a diagnosis of breast cancer has improved dramatically in the past 20 years with the development of early diagnosis tools and new treatment regimens that is early detection and treatment through clinical evaluation playing a vital role.

Among different breast cancer imaging techniques like Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Conventional X-ray, Ultrasonography, Mammography is the most common technique for breast cancer abnormality detection. It is available widely all around the world and has reduced the breast cancer mortality by 40%. Mammography itself is limited by several known risks such as over diagnosis, overtreatment, false-positive examinations, and radiation exposure. Additional limitations such as relatively low sensitivity, patient discomfort, and pain during the procedure due to breast compression and false negativity in dense breasts also impact mammography screening reliability. To address the limitations mentioned earlier, implementation of microwave imaging (MWI) for breast cancer imaging has come to the scene. It is based on the use of harmless electromagnetic waves and can provide relevant initial diagnostic information without resorting to X-rays and has the potential of early diagnosis of breast cancer due to using non-ionizing radiation, safe and low-cost technology.

SAFE is a breast cancer imaging device designed and produced at MITOS lab at ITU that works based on the use of harmless electromagnetic waves. There has been studies and papers about the function of the system itself. Moreover, data driven by SAFE has been classified from different perspectives. As at the current stage SAFE can provide only 2-D images in the coronal plane, we feel we need the information about the effect of depth of the tumor on the situation of the patient and to investigate if we can get useful information from the raw data about the anterior/posterior position of the tumor, which would benefit the doctors to make more accurate clinical predictions.

In this thesis, three different machine learning approaches is going to be used to classify the raw clinical data to figure out the possible contribution of vertical location of the tumor on the study and early detection of the cancer. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Three are the approaches that will be utilized for classification of the existed data set.

MAKİNE ÖĞRENME YOLUYLA TÜMÖRÜN ANTERİOR/POSTERİOR POZİSYONUNU ELDE EDİN

Özet

Meme kanseri, meme dokusundaki hücrelerin kontrolsüz bir şekilde büyümesiyle ortaya çıkan bir hastalıktır. Farklı meme bölgelerinde başlayabilir ve hangi hücrelerin kanserleştiğine, nasıl yayıldığına ve nasıl davrandığına bağlı olarak farklı meme kanseri tipleri belirlenebilir. Meme kanseri yağ dokusunda ve lobüllerde büyüyebilir veya kan ve lenf sistemi aracılığıyla diğer organlara metastaz yapabilir. 2020 yılında akciğer kanserini geride bırakarak meme kanseri, dünya genelinde en yaygın görülen kanser türü haline gelmiştir. Bu hastalık, az gelişmiş ülkelerde kanserden kaynaklanan ölümlerin önde gelen sebebi olup Amerikan kadınlarında kansere bağlı ölümlerin ikinci sıradaki nedenidir. Dünya Sağlık Örgütü'ne göre, dünya genelinde yaklaşık 2,3 milyon kadına meme kanseri teşhisi konmuş ve bunlardan 685.000'i hayatını kaybetmiştir. Bu rakamlar, 2008 yılından bu yana teşhis edilen hasta sayısının %20'den fazla, ölüm oranının ise %14'ten fazla arttığını göstermektedir. Ancak, erken teşhis yöntemlerinin geliştirilmesi ve yeni tedavi rejimlerinin kullanılmasıyla birlikte meme kanseri tanısı alan hastaların sağkalım oranı son 20 yılda büyük ölçüde artmıştır. Erken teşhis ve tedavi, klinik değerlendirme yoluyla önemli bir rol oynamaktadır.

Magnetik Rezonans Görüntüleme (MRG), Pozitron Emisyon Tomografisi (PET), Geleneksel X-ışını, Ultrasonografi gibi farklı meme kanseri görüntüleme teknikleri arasında, Mammografi meme kanseri anormalliklerinin tespiti için en yaygın kullanılan tekniktir. Mammografi, dünya genelinde yaygın olarak kullanılmakta ve meme kanseri mortalitesini %40 oranında azaltmıştır. Ancak mammografi, aşırı teşhis, gereksiz tedaviler, yanlış pozitif sonuçlar ve radyasyon maruziyeti gibi birçok bilinen riskle sınırlıdır. Ayrıca, görece düşük duyarlılık, meme sıkıştırılması nedeniyle prosedür sırasında hastada rahatsızlık ve ağrı, yoğun meme dokusunda yanlış negatif sonuçlar gibi ek sınırlamalar da mammografi taramasının güvenilirliğini etkilemektedir. Bu nedenlerle, meme kanseri görüntüleme alanında mikrodalga görüntüleme (MWI) yöntemi devreye girmiştir. Bu yöntem, zararsız elektromanyetik dalgaların kullanımına dayanır ve X ışınlarına başvurmadan ilgili ilk tanısal bilgileri sağlayabilir. Ayrıca, iyonlaştırıcı olmayan radyasyon kullanımı, güvenli ve düşük maliyetli bir teknoloji olması nedeniyle meme kanserinin erken teşhisi için potansiyel sunmaktadır.

SAFE, İTÜ'deki MITOS laboratuvarında tasarlanan ve üretilen bir meme kanseri görüntüleme cihazıdır. Cihazın işleyişi hakkında yapılan çalışmalar ve yayınlar bulunmaktadır. Ayrıca, SAFE tarafından sağlanan veriler farklı perspektiflerden sınıflandırılmıştır. Şu anki aşamada, SAFE sadece koronal düzlemde 2 boyutlu görüntüler sağlayabilmektedir. Bu nedenle, tümörün derinliğinin hastanın durumu üzerindeki etkisi hakkında bilgi edinmek ve ham verilerden tümörün anterior/posterior

konumu hakkında faydalı bilgiler elde edebilmek mümkün olabilir. Bu durum, doktorların daha doğru klinik tahminler yapmalarına yardımcı olabilir.

Bu tezde, tümörün dikey konumunun çalışmaya olan olası katkısını ve kanserin erken teşhisi için ham klinik verilerin sınıflandırılmasını belirlemek amacıyla üç farklı makine öğrenimi yaklaşımı kullanılacaktır. Destek Vektör Makineleri (SVM), K-En Yakın Komşular (KNN) ve Karar Ağacı, mevcut veri setinin sınıflandırılmasında kullanılacak yöntemlerdir.

1. Introduction

An uncontrollable growth in some of the body's cells and its spread to the other parts of the body is called cancer which is a genetic disease. When the normal process of cell growth and its multiplication to form new cells breaks down, abnormal or damaged cells grow and multiply when they shouldn't. This process may generate tumors that can be cancerous or not (benign).

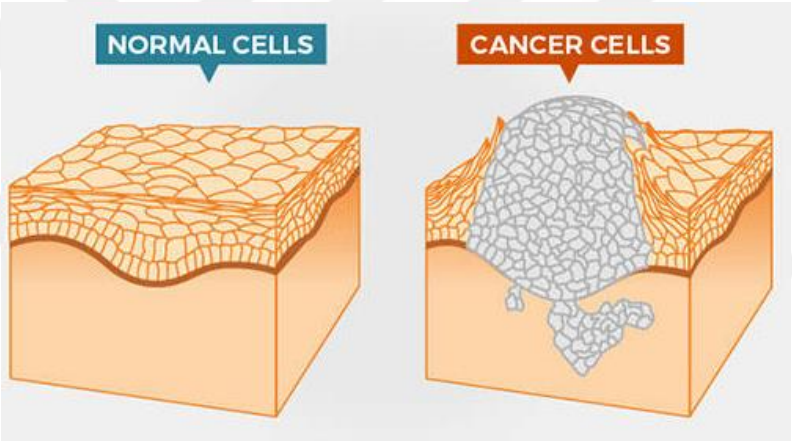


Figure 1.1: Normal and Cancerous cell.

Cancer has been seen among both genders, with a high mortality rate. According to the World Health Organization report [1], cancer is one of the leading causes of death worldwide, which accounts for about 10 million deaths in 2020. It has been shown that of almost 20 million new cases of cancer observed globally, nearly 10 million cases led to death. Figure 1.2 and 1.3 [2] shows that the most common cancer in terms of new incidence in 2020 was Breast cancer, with 2.26 million cases, and 9.6% of death in both sexes and all ages, globally.

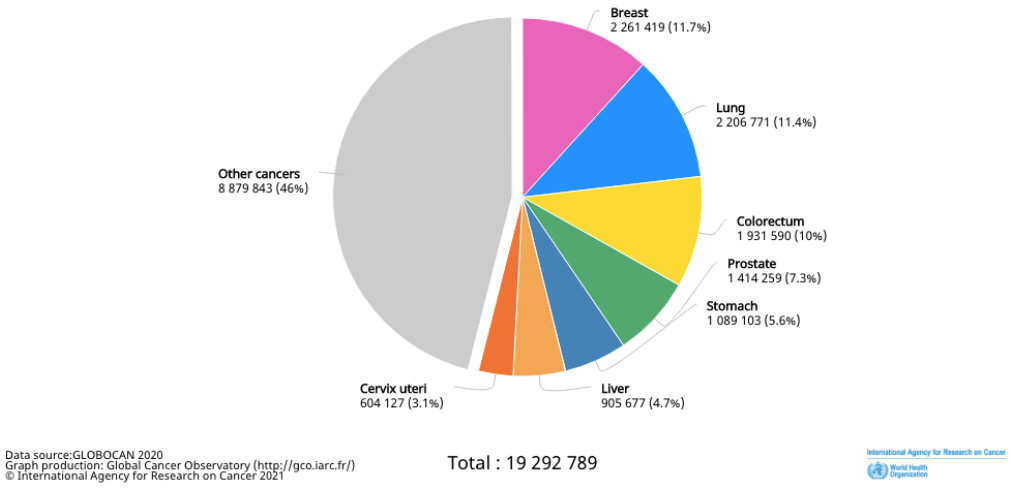


Figure 1.2 Number of new cancer cases in 2020, worldwide, both sexes, all ages.

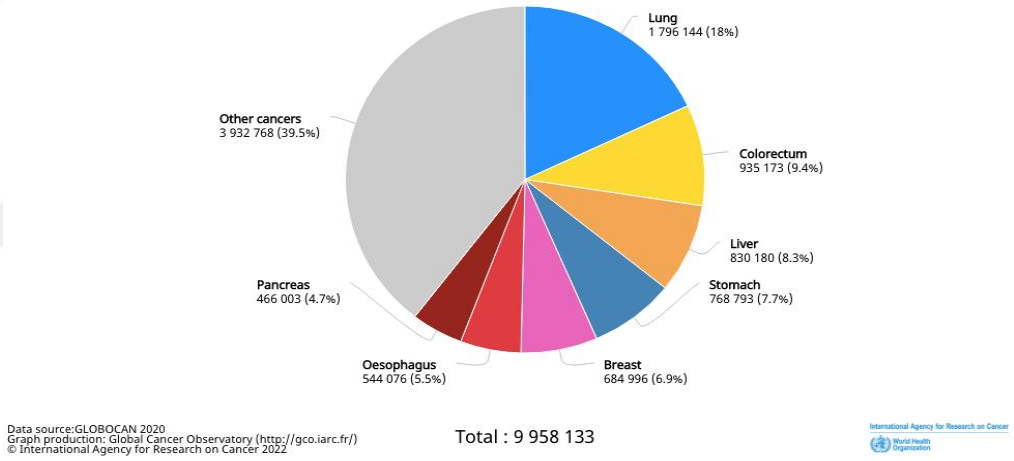


Figure 1.3: Number of cancer deaths in 2020, worldwide, both sexes, all ages.

The number and rate of new cases in women of all ages worldwide by 2020 are depicted in Figure 1.4. The highest incidence of cancer in women, with a 24.5% and 15.5% of mortality rate, is breast cancer which can be seen in Figure 1.5.

Based on this research, out of 2,261,419 detected new cases, 684,996 led to death [1]. As shown in the statistics, the incidence and mortality rate of breast cancer is distinguishably high. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past five years, making it the world's most prevalent cancer [1]. By a quick look at the future trend of cancer, we can figure out that there is an estimated 20 million new cases of cancer, and half of this number cause

death. This burden will increase by approximately 60% over the next two decades and an increase of 30 million new cases by 2040 with the greatest increase in low and middle-income countries if no further action is taken to prevent and control cancer [2].

Considering the future trend of cancer it is obvious that early detection of cancer is crucial in the success of therapy and reducing mortality rate.

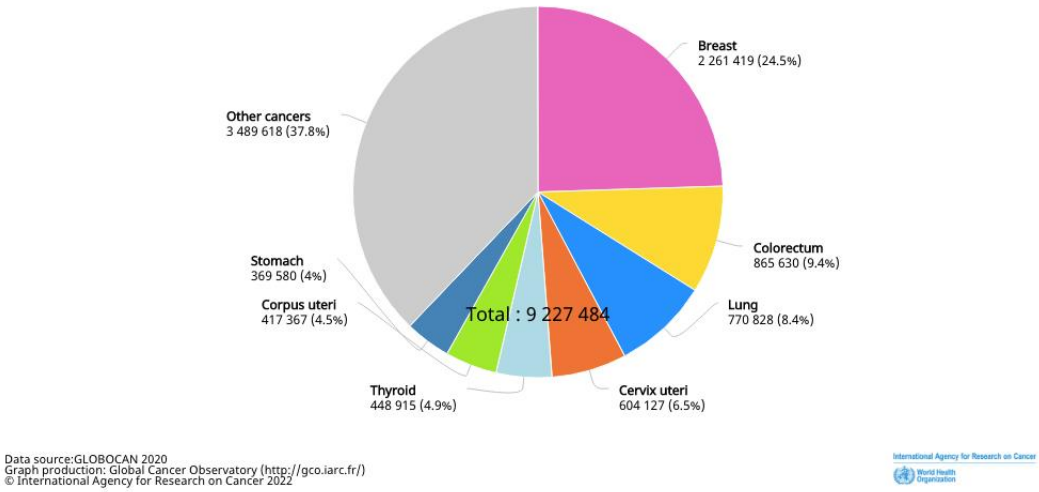


Figure 1.4: Number of new cancer cases in 2020, worldwide, female, all ages.

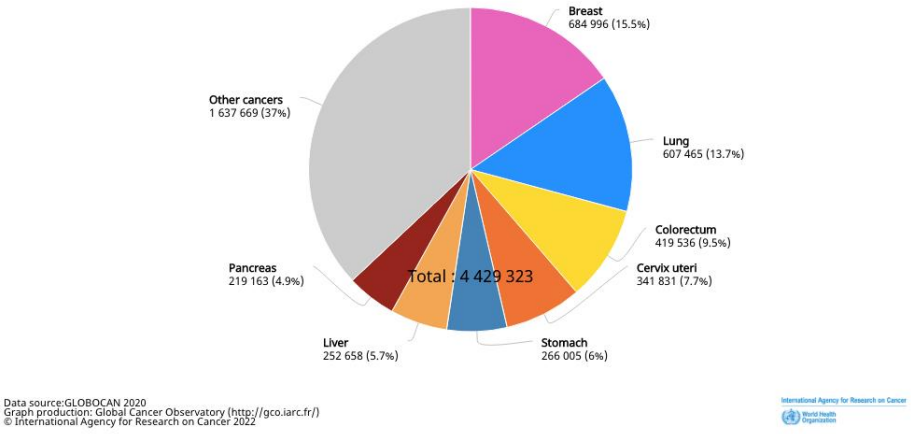


Figure 1.5: Number of cancer deaths in 2020, worldwide, female, all ages.

1.1Breast Anatomy

Female breast structure is complex. Its majority part is glandular tissue, which includes the breast lobes and breast ducts, and fatty tissues. The ratio of fat versus glandular varies depending on age, post-menopausal, post-partum, or pregnancy status. Besides

that, it has fibrous, supportive, connective tissue, blood vessels, ligaments, fat, lymph nodes, nipple, areola, alveoli, and nerves.

Glandular tissue is composed of 15-20 breast lobes in each breast around the nipple. Each breast lobe is further divided into multiple breast lobules about 20-40. Lobule branch out of the nipple and are connected by a network of tubes. Each lobule holds tiny glands named alveoli that produce and store milk in nursing women. Both the lobes and lobules are linked by milk ducts which carry the milk from the alveoli toward the areola and the nipple. Areola is the circular dark-colored area of skin surrounding the nipple and it has glands called Montgomery's glands that discharge a lubricating oil to protect the nipple and skin.

Breast has a very rich and intricate blood and lymphatic system. The lymphatic system of the breast which is composed of lymph, lymphatic tissue, lymph nodes, and lymphatic vessels is a part of the lymphatic system. It plays an essential role in systemic immunity by carrying bacteria and wastes out of the human body. Lymph nodes are small, bean-shaped organs that are located in the armpit, above the collarbone, in the neck, and in the chest produce and filter a colorless fluid called lymph. Lymph vessels filter and carry lymph fluid from the breast to the lymph nodes. Lymph nodes are responsible for filtering damaged and cancerous cells, as well as producing and storing lymphocytes and other cells of the immune system to attack and kill bacteria and other harmful substances in the fluid.

These breast structures are generally where cancer begins to form [3] [4] [5].

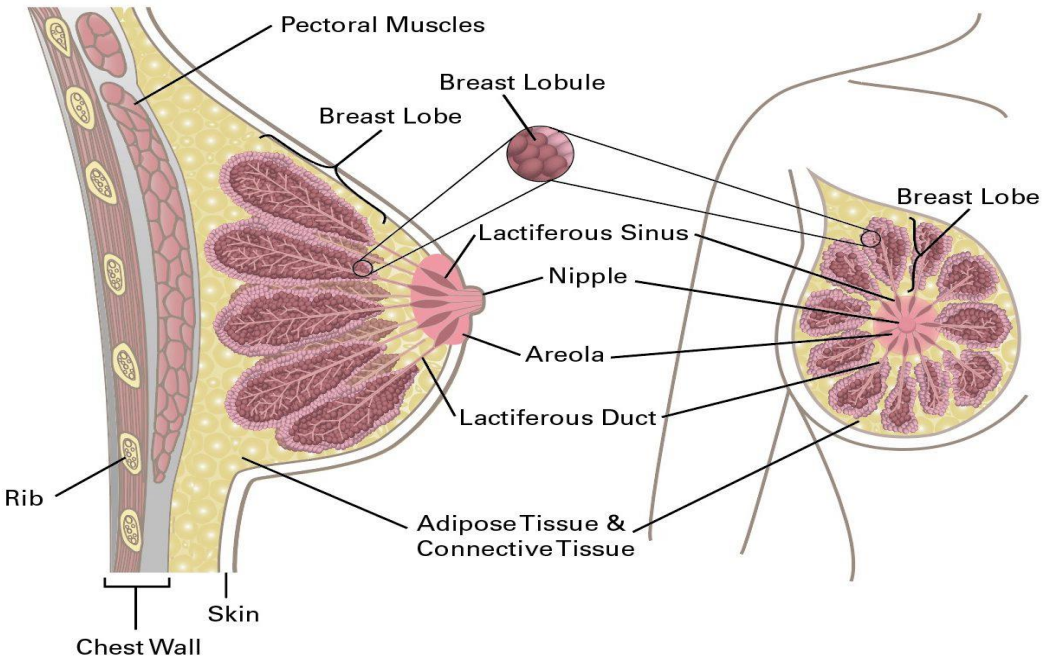


Figure 1.6: Female Breast Anatomy [5].

1.2 Breast Cancer

As mentioned earlier, breast cancer is a multifactorial disease that develops initially from breast tissues in which an uncontrollable growth of cell due to modification or mutation of DNA and or RNA happens. Breast cancer occurs almost entirely in women, but men can get breast cancer, too.

Technically any type of lump formed by cells is referred to as a tumors and there are various types of tumor, though not all tumors are malignant (cancerous). Fibroadenoma, Breast cysts, and fibrocystic changes are three common type of lumps that are benign (non-cancerous) that does not increase your risk for breast cancer. Majority of breast cancers originates in the passages that carries milk to the nipple (ductal cancers) or in the cells of the milk-producing glands (called the lobular cancers), a few initiate in other tissues and less commonly in the fatty and fibrous connective tissues of the breast [6] [7].

Breast cancer has dozen of types and subtypes. Some aggressively grows while others grows slower. It has two main broad categories: invasive and noninvasive.

Invasive cancer happens when the cancer spreads from the initial to the surrounding through the blood and lymph systems. In this type of the cancer cells grown through the lining of the ducts into the surrounding like breast tissue, lymph nodes or elsewhere in the body. Invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) are the most common types of invasive breast cancer. IDC accounts for roughly 70 to 80 percent of all cases is the cancer that generates in the lining of milk ducts and expands into other parts of the breast. Invasive lobular carcinoma as the name implies is invasive breast cancer that begins in the glands responsible for production of the milk named lobules. ILC accounts for about 10% of all invasive breast cancers [8] [9] [10].

In non-invasive type of breast cancer or Carcinomas in situ (in situ) cancerous cells remain in a specific place in breast and does not spread to surrounding tissue, lobules or ducts. However, in situ type of breast cancer may eventually turn to invasive type. It includes Ductal carcinoma in situ (DCIS), which is the most common form of non-invasive breast cancer (90%) which is not life-threatening but having DCIS increases the chance of developing invasive breast cancer in the future, and Lobular carcinoma in situ (LCIS). LCIS is considered as a marker for increased breast cancer risk [5] [10].

The symptoms of the breast cancer varies based on the size of the tumor, how fast it growth, and where the tumor locates in the breast. The first and major symptom of breast cancer that most women notice is a lump, sometimes as little as a pea, found in the breast or armpit or an area of thickened tissue in their breast. Breast Appearance Changes, nipple changes like suddenly nipple discharge that might be bloody or if it happens in one breast, pain in the nipple, inverted nipple, scaly or pitted skin on nipple, persistent tenderness of the breast, and unusual breast pain or discomfort are common signs of breast cancer [11] [12].

Being female, personal or family history of breast cancer, age, Genetic predispositions, obesity, excessive consumption of alcohol and tobacco, High lifetime estrogen exposure, dense breast tissue, radiation therapy in the chest area, postmenopausal hormone therapy, race and ethnicity, environmental pollutants, gender-affirming hormone therapy, oral contraceptives, are factors associated with increased risk increased risk of breast cancer [13].

2. Modalities

Imaging play a vital role in early detection and control of the breast cancer. It has also an important role in Survival rate of breast cancer. According to WHO more than 90% in high-income countries, to 66% in India and 40% in South Africa is the high survival rate of breast cancer for at least 5 years after diagnosis that is a proof to the importance of the imaging.

There are different types of imaging mechanisms and tools but some of them are more common that is as follow:

1. Mammography
2. Magnetic Resonance Imaging (MRI)
3. Ultrasound
4. Positron Emission Tomography (PET).

In the following sections every modality will be briefly explained.

2.1. Mammography

Mammography is the most common imaging modality in breast cancer imaging, used for early breast cancer detection. It has been used for decades for women with no breast complaints and who have breast problems, such as pain, nipple discharge, or lump. Mammography is an X-ray examination and a two-dimensional imaging modality to single out the suspicious findings in breast cancer. Asymmetric calcifications, deformed breast areas, and masses can be a part of this identification [14]. The mammography unit includes an X-ray tube that encloses the cathode and anode in which the breast will be placed on the detector and compressed by a parallel-plate compression device. Images are produced by penetrating low-dose X-rays, 20-32 kVp, through the tissue. Up until the emergence of digital mammography with which images have more consistent quality and higher soft-tissue contrast, a conventional screen film used to be used in mammography where the film had to be chemically processed [15].

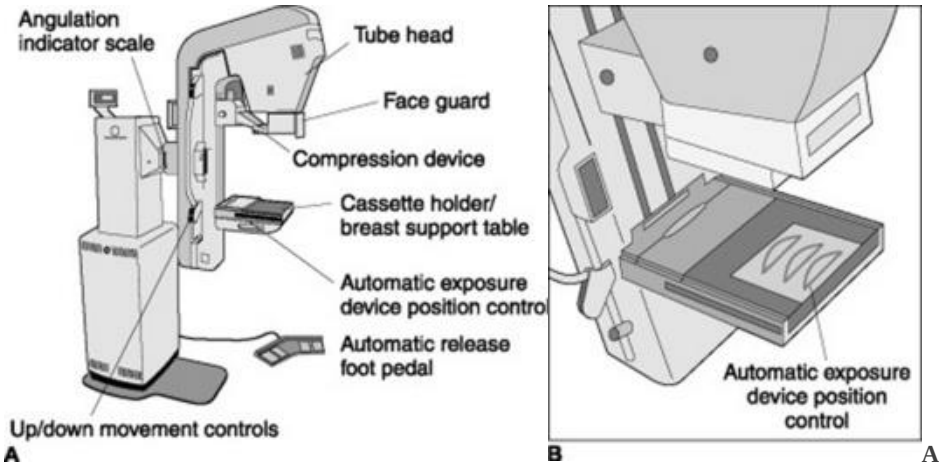


Figure 2.1: A mammography unit. B A mammography X-ray unit showing position of automatic exposure control (AEC).

Mammography demands two radiographic images of each breast i.e., the mediolateral oblique (MLO) view and the cranialcaudal (CC) view. The former one is an image taken from the side and reflects more of the breast in the upper-outer quadrant and gives the best view of the lateral side of the breast latter one is an image taken from above the breast and the view depicts the entire breast. [16].

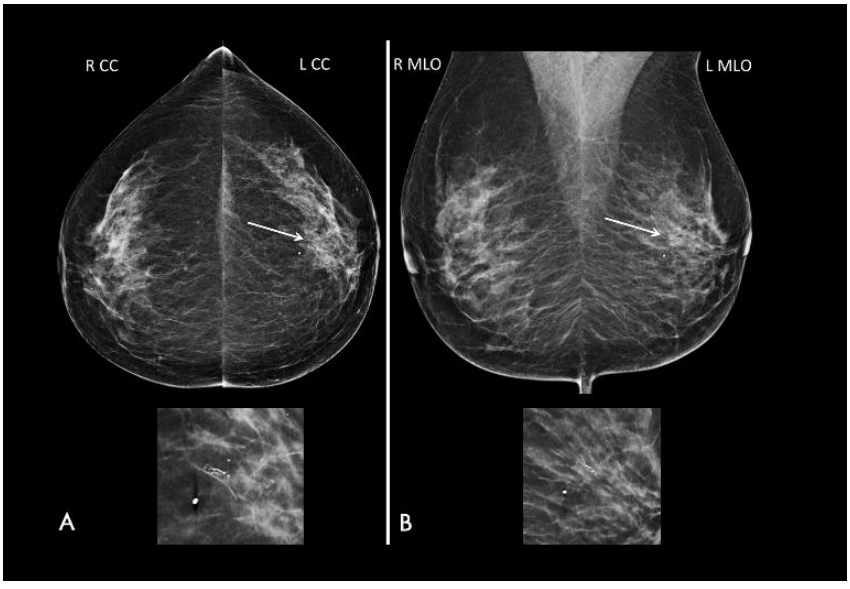


Figure 2.2: Two-view screening mammograms in, A, craniocaudal (CC) and, B, mediolateral oblique (MLO) [17]

There are two different types of mammography: screening mammography and diagnostic mammography. The first one refers to mammography in which women with

no symptoms will be targeted to detect breast changes with the purpose of reducing the breast cancer modality by early detection and the second one is for Patients with breast symptoms [18]. Mammography has a relatively high spatial resolution range (50-100 μm) that allows it to detect tiny deposits of calcium and nonpalpable soft tissue masses [19]. Although mammography has benefits as other procedures, it is limited by several known risks such as overdiagnosis, overtreatment, and false-positive examinations. Additional limitations include relatively low sensitivity, and pain during the procedure due to breast compression and false negativity in dense breasts [20]. Moreover, the exposure of the breast as a highly sensitive tissue to high radiation approximately 1-3 mGy during mammography can increase the risk of cancer occurrence [21].

2.2. Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive and painless imaging method that provides images utilizing the magnetic properties of the hydrogen atoms in the tissue that are exposed to a strong magnetic field Unlike mammograms, a breast MRI doesn't use X-rays (radiation). It has become widely used due to advances in surface coil technology, the introduction of new contrast agents, and fast imaging sequences. Compared to mammography and Ultrasound this method is more sensitive in the diagnosis of breast cancer and relatively cost-effective [14]. MRI is a non-ionizing radiation imaging system, which produces high-resolution, high-contrast cross-sectional (tomographic) images.

MRI unit for breast cancer wherein the patient lies face down on a special coil that has openings to hang breast tissue and contrast material is injected into the body that facilitates certain tissues and blood vessels to show up more clearly and in greater detail [22].

Due to the high sensitivity of MRI for breast cancer detection, several medical organizations including the American Cancer Society (ACS) and the National Comprehensive Cancer Coalition Network (NCCN) now recommend that women with particular high risk factors for breast cancer undergo annual screening breast MRI in addition to mammography [23]

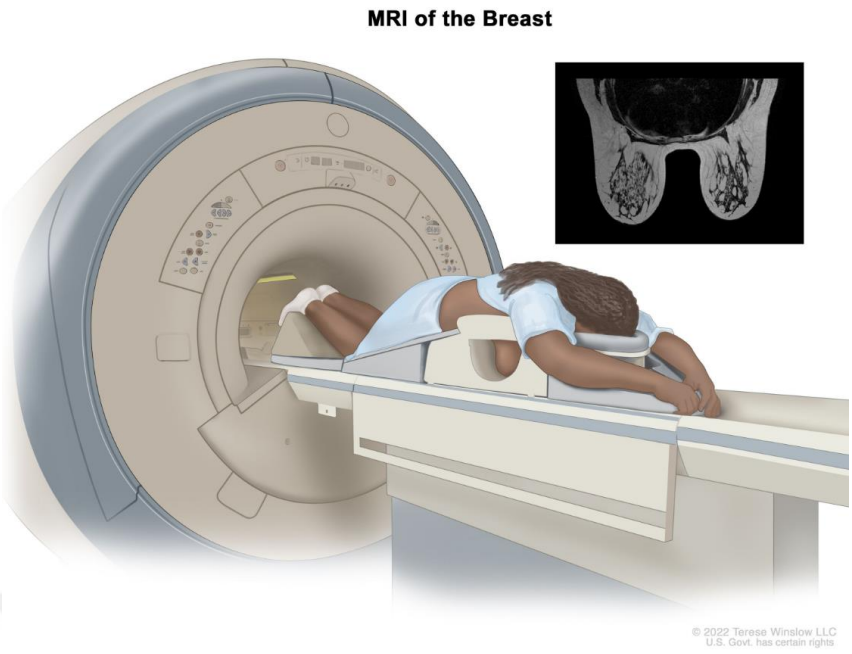


Figure 2.3: Breast MRI [24]

Three-dimensional format and strong soft tissue contrast of MRI for breast imaging are attractive features that permit the anatomical structure of the breast to be depicted in great detail, however, it is not sensitive to microcalcifications which is an early indication of breast [25]. MRI is a time-consuming and expensive imaging modality, yet its low risk and high image quality make it a suitable procedure for screening patients with a high risk of breast cancer (20-25%) [14]. MRI can play a pivotal role in diagnosing a patient with biopsy-proven cancer or a breast abnormality, yet it has its own shortcomings that happen when the low-to-moderate specificity of MRI is combined with its high sensitivity and leads to unnecessary biopsy and high costs [26]. Comparing the sensitivity and specificity of MRI imaging techniques with mammography, MRI and MG sensitivities were 92% and 75%, respectively, and their specificities were 71% and 70%, likewise [27].

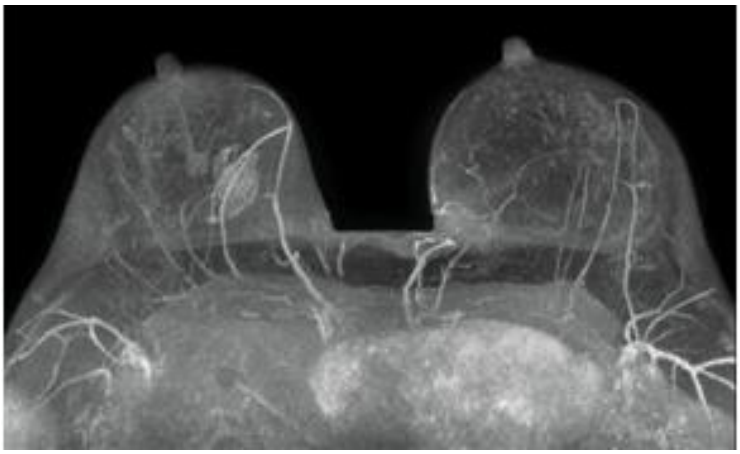


Figure 2.4: MRI scan of breast. [20]

2.3. Ultrasound

Ultrasound imaging is another commonly used modality for breast cancer detection and diagnosis that was shown to be successful in detecting tumors in dense breasts which may be missed using mammography. Ultrasound imaging of the breast uses sound waves and works based on the differences in acoustic impedances to produce pictures of the internal structures of the breast. The waves utilized in this approach have a frequency of 3-12 MHz and it is transmitted to the tissue with a probe [29]. Ultrasound is usually used with mammography to detect malignancy or benign nature of the tumor. The sensitivity of the ultrasound imaging method can be increased to 97.3% and specificity up to almost 76% by adding mammography and physical examination method to it [30]. The diagnostic accuracy in ultrasound intensely depends on the scanner's proficiency and correct selection of ultrasound parameters [31].

2.4. Positron Emission Tomography

Through the years, advances in imaging technologies and radiochemistry have been used to develop new radiolabeled molecules to target different tissues, organs, or even molecules rather than just anatomy. A positron emission tomography (PET) scan is an imaging test that utilizes pharmaceuticals labeled with positron-emitting radionuclides called tracers. Although many radiotracers have been developed for PET imaging to be attached to a drug and injected into an arm vein as a tracer, most breast cancer imaging studies have been performed with Fluorodeoxyglucose 18F or 18F-FDG [32]. The gamma rays emitted from the tracer are detected by special cameras and generate

a three-dimensional color image. Cancer cells tend to be more active than normal cells, and they absorb more of the radioactive as a result. That is why PET is an effective means of predicting response to neoadjuvant chemotherapy, as the tracer substance goes to places in the body where the cells are most active, especially highlighting cancerous tissue and ensuring that the most appropriate form of treatment is selected [33]. The incorporation of PET and computed tomography (CT) which is called PET-CT has been used to provide an anatomical and functional view of the suspect cells even for small tumors that cannot be detected by conventional imaging. This is particularly important when looking to see if cancer has spread from the breast to other parts of the body.

Many researches have shown high sensitivity of FDG-PET up to 68% when small (2 cm) tumors are present and a sensitivity of 92% in the case of bigger (2-5 cm) ones. Also high specificity of FDG-PET for the detection of primarily large and palpable breast tumors [34] [35]. It has been used recently for diagnosis, staging, restaging, and also treatment response evaluation, however, according to [36 [37] [38] [39] [40] there are some limitations and pitfalls that may lead to false-negative or false-positive. Mentioned studies have shown a variety of conclusions for the use of PET in breast cancer. PET scans are available in only very few centers, and they are expensive, sophisticated test that requires special expertise.

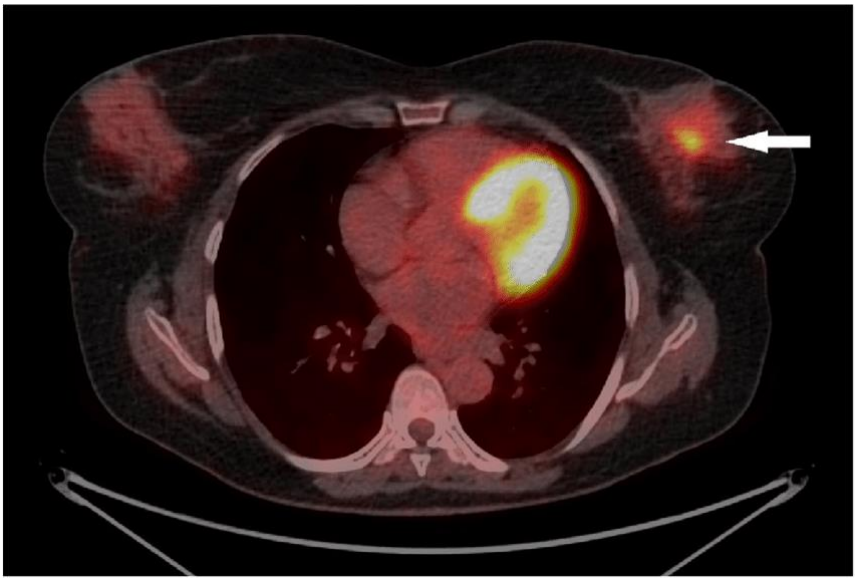


Figure 2.5: Transverse image of 18 F-FDG PET/CT in patient with breast cancer

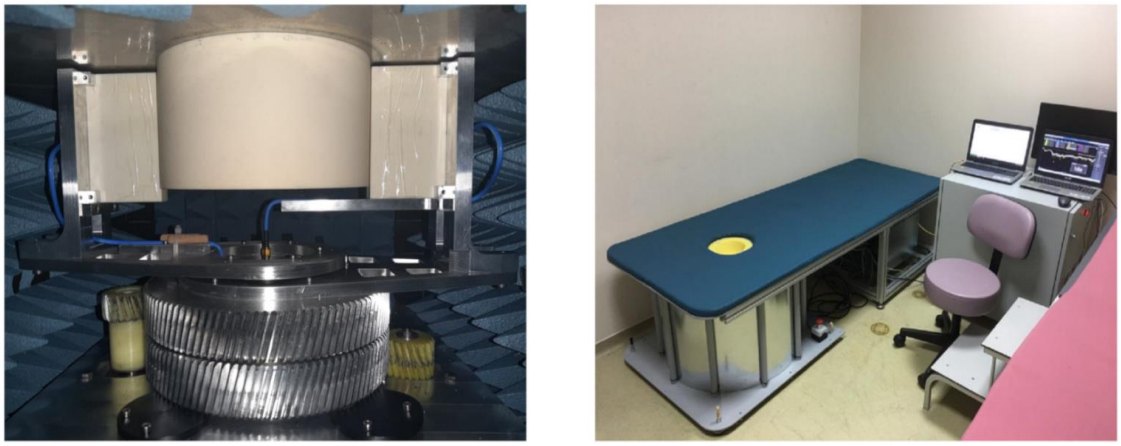
2.5. Microwave

As mentioned above, Mammography, MRI, and PET are relatively costly, time-consuming, and, in some cases detrimental methods due to high-dose radiation, especially in PET and Mammography. Furthermore, the low sensitivity and specificity of these methods lead to unnecessary biopsies, which means more waste of time and inconvenience for the patient. Therefore, the requirement for low-risk and fast alternatives for the initial examination and diagnosis arises.

The notable contrast of cells' dielectric properties at microwave frequencies attracted researchers' interest and gave birth to a promising alternative for breast cancer imaging known as Microwave Imaging (MWI) to overcome the prenominated disadvantages. MWI is a non-invasive, non-ionizing, and inexpensive technique that offers a safer, cheaper, and an effective way of detecting breast cancer [41] [42]. MWI does not require the breast to be pressed between two paddles as it works based on the microwave frequency range. As we mentioned earlier the other modalities are expensive. The low cost of MWI makes it easier to provide in developing countries, which can be a promising step to reduce the incidence of breast cancer and its mortality in deprived areas. This technology is in the clinical study stage and seven MWI prototypes have undergone clinical testing [41].

MWI functions are based on the differences in the dielectric properties of various tissues. Due to the presence of water, the electrical properties of the tissue exhibit a considerable frequency dependency in the microwave band. According to the fact that tumors contain more fluid or blood vessels than the surrounding healthy tissue, there is a significant difference between the permittivity and conductivity of breast cancer tissue compared to healthy tissues with respect to the frequency of an applied field [43]. When microwave signals irradiate and interact with the breast tissue some part of them gets transmitted and some gets reflected back. The signals that get reflected back are stored and processed through an imaging algorithm to generate a 2D image of the breast. This resultant image of the breast shows the presence of the tumor inside it [44].

There are seven clinically tested microwave imaging prototypes. Among all of them, just two of them were studies on a considerable number of patients which are Multistatic Array Processing for Radiowave Image Acquisition (MARIA®) and a system developed at the University of Bristol, UK [45][46]. SAFE (Scan and Find Early) was developed by Mitos Medical Lab inspired by the need for comprehensive MWI clinical assessments in terms of the number of patients. Currently MARIA, system Developed in Bristol, and SAFE have conducted studies with similar number of patients. SAFE system which is based on non-ionizing microwave radiation provides a pain-free procedure since it does not require breast compression. As it is shown in Figure 2.6 SAFE has a bed on which the patient can lie and a place for a size adjustable cup that the patient can insert the breast into the medium cup. When scanning starts, transmitter and receiver antennas turn around the cup to construct the images of the breast.



(a) (b)

Figure 2.6: Microwave breast imaging [20].

3. Literature

3.1. Machine Learning approaches for breast cancer detection

One of the artificial intelligence branches is Machine Learning which is known for its capability in combining a variety of statistical, probabilistic, and optimization techniques that allow computers to “learn” from past examples and to detect hard-to-

diagnosed patterns from massive, noisy, or complex data sets. These features are particularly well-suited to medical applications [47]. It has been widely used in the medical field to develop better diagnostic tools. Various research works have been conducted by researchers utilizing machine learning techniques to detect breast cancer and classify tumors. Ayush Sharma et. Al. in the work "Machine Learning Approaches for Breast Cancer Diagnosis and Prognosis" has applied three different ML approaches called Logistic Regression, Nearest Neighbor, and Support Vector Machines to predict breast cancer as benign or malignant using data set from Wisconsin Breast Cancer Data using sophisticated classifiers such as Logistic Regression [48]. A new machine learning approach, namely Gradient Boosting has been used by Janjic et. al in order to find the best-fitting ML model for automated breast lesion classification [49]. To the best of our knowledge, there has been no work that utilizes machine learning approaches to classify the raw clinical data to figure out the possible contribution of the vertical location of the tumor on the study and early detection of cancer.

4. Predictive Machine Learning Approaches

4.1. Support Vector Machine Classifier (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification and regression. It is basically a method to find a hyperplane to divide the examples of different outcomes. In this algorithm each data item is a plotted point in n-dimensional space (n the number of features we have) and the goal in SVM is to find the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. The optimal separation that is called optimal hyper plane is the one that finds a maximum distance to the closest point of the two classes. Vladimir Vapnik in 1963 proposed the optimal hyper plane first time as a linear classifier. Later on in 1992, Bernhard Boser, Isabelle Guyon and Vapnik applied kernel trick to hyperplane to suggest a nonlinear classifier. The resulting algorithm is formally similar, except that every dot product is replaced by a non-linear kernel function. It is

a powerful pattern, the solution to an SVM is unique, and it is easy to interpret geometrically.

In [50] a neural networks Support Vector Machine method for diagnosis of breast cancer was applied on a set of 683 samples of actual data, and 117 samples was generated by NN. The efficiency of the SVM 97% that is a high rate of accuracy compare to that of manual of with efficiency of 85% that can help the Doctor's decision to avoid Biopsy.

As we mentioned before the goal of support vector machine is to find that hyperplane that has the maximum margin from the closest points. In other words the goal is to maximize the minimum distance. In SVM we have some unknown points and we are interested in measuring which side of the hyperplane these points belong to. To figure it out we can utilize the concept of dot product.

If we define the equation of the hyperplane as follow:

$$w.x + b = 0 \quad (4.1)$$

Then we are able to use the hyperplane to make predictions:

$$h\,(x_i) = \begin{cases} +1, & if \;\; w.x + b \geq 0 \\ -1, & if \;\; w.x + b < 0 \end{cases} \quad (4.2)$$

And the optimization form of SVM will be as follow:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$\text{Subjected to } y_i(w.x + b) - 1 \geq 0\,, \;\; i = 1,2,\dots,m$$

4.2. Nearest Neighbors Classifier (KNN)

The k-Nearest Neighbors algorithm also known as KNN or k-NN is a non-parametric - which means it does not make any assumption on underlying data - and supervised learning algorithm that basically functions on samples that do not require a learning phase and utilizes the adjacency to classify or predict about an individual data point. The algorithm considers the similarity between the new data point and the classes of available nearest neighbors. And finally, the class that appears most among the neighbors is the one assigned to the new data point to be classified. The neighbors are weighted by the distance that separates them from the new elements to classify that is

the algorithm uses the distance function. For continuous variables most common distance functions are:

Euclidean distance

$$D(x,y) = \sqrt{\sum_{i=1}^{n}((x_i - y_i)^2)}$$

Manhattan distance

$$D(x,y) = \sum_{i=1}^{k}|x_i - y_i|$$

Minkowski distance

$$D(x,y) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{\frac{1}{p}}$$

In the case of categorical variables the Hamming distance must be used that is the number of values that are different between two vectors.

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \rightarrow D = 0$$

$$x \neq y \rightarrow D = 1$$

KNN is simple to implement, robust to the noisy training data, and can be more effective if the training data is large. A good choice of the value of k can be selected by different heuristic techniques such as cross-validation.

4.3. Decision Tree (DT)

The decision Tree algorithm is also a non-parametric and supervised learning algorithm that uses a tree-like model of decisions and their possible consequences. The aim of this model is to create a training model that predicts the class of a variable by

learning simple decision rules based on how a previous set of questions were answered. This method breaks down a dataset into smaller and smaller sets and makes a predictive model and has the root node that is the base of the decision tree from which a series of decision nodes flow that represent the decision to be made. Leaves are the representation of the consequences of each decision. A tree can be seen as a piecewise constant approximation in which leaves represent class labels and branches represent conjunctions of features leading to the class labels [50].

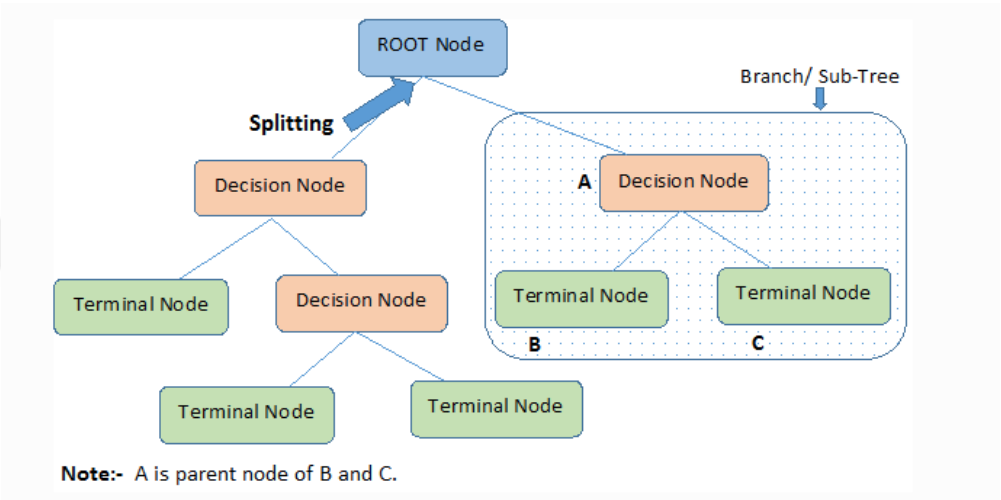


Figure 4.1: Decision Tree Algorithm

A decision tree is easy to understand and interpret and it can be visualized. It does not require long data preparation and can handle both numerical and categorical data. It is able to handle multi-output problems as well. However, this learner may cause overfitting which happens when the model creates over-complex trees that do not generalize the data well. Moreover, can be unstable because small variations in the data might result in a completely different tree being generated. The common algorithms for decision trees are ID3, C4.5, CART, J48, and Random Forest. Sumbaly et. Al utilized a J48 decision tree algorithm and 10-fold cross-validation to prepare training and test data for the early detection of breast cancer [50]. The decision tree has been applied in many research works [51] [52] successfully.

Implementation of the decision tree has its own challenges and its primary difficulty is to decide which attributes we need to consider as the root node and also each level. This procedure is called attribute selection. If we have a dataset with N attributes then the attribute selection procedure will be a complicated one and randomly selection

cannot lead us to a good result. To address this issue researchers worked and figured some criteria out such as Entropy, Information gain, and Gini index.

Entropy is nothing but the measure of randomness. It is hard to get any conclusion from the information when the entropy is high. The formula for entropy is shown below:

$$E(S) = -p_{(+)} \log p_{(+)} - p_{(-)} \log p_{(-)}$$

Where p+ is the probability of positive class, p– is the probability of negative class, and S is the subset of the training example. In decision tree problems the output is mostly "yes" or "no". Entropy measures the impurity of a node. It basically shows how random our data is and in the sub-split step we should be getting a "no" or it has to be a "yes". Using entropy gives us the impurity of a specific node but we do not have information about the parent entropy or the entropy of a particular node has decreased or not. Here is where "information gain" comes to the scene that is also called as IG. IG is a statistical feature that calculate how well an attribute classifies the training examples. The highest information gain and the smallest entropy is what a decision tree is supposed to find about an attribute. Mathematically, IG is represented as:

$$IG(T, X) = Entropy(T) - Entropy(T, X)$$

In a much simpler way, we can conclude that:

$$IG = Entropy(before) - \sum_{j=1}^{k} Entropy(j, after)$$

Where "before" is the dataset before the split, K is the number of subsets generated by the split, and (j, after) is subset j after the split.

Gini index is a kind of cost function that is used to evaluate splits in the dataset. It can be applied only on binary splits and it performs with the categorical target variables.

$$Gini = 1 - \sum_{i=1}^{c} (p_i)^2$$

5. Material and Methods

There are several researchers that proposed classification algorithms to investigate breast cancer data mostly to predict the classification of breast cancer type (malignant and benign). Classification algorithms have been selected to meet the most suitable one to predict cancer with high accuracy and efficiency. Different attributes such as age, sex, past history, medical diagnosis, Occupation, diet, weight, and more have been considered for classification. To the best of our knowledge, there has not been any research work with a focus on the vertical location of the tumor as an attribute. As at the current stage SAFE can provide only 2-D images in the coronal plane, we feel we need the information about the effect of the depth of the tumor on the situation of the patient and to investigate if we can obtain useful information from the raw data about the anterior/posterior position of the tumor, which would benefit the doctors to make more accurate clinical predictions.

5.1. Dataset

In this work, data is collected throughout the SAFE scanning procedure and taken from 100 patients at Marmara University Hospital Breast Center Hospital. The format used in this work is CSV (Comma Separated Value) format. The CSV format of breast cancer data is given as input via frequency and vertical location of the tumor.

5.2. Classification Algorithms

In this work, we have chosen three classifiers to apply to the dataset and figure the best one out. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Three are the approaches that are utilized for the classification of the existing data set. As the dataset is limited, 5-fold cross-validation was used to prepare training and test data.

5.3. Preprocessing

As we mentioned before the dataset is taken from 100 patients and it is in S36 parameter format. In the first stage, a 36x36x38 matrix has been made for each S36 file in which we have 36 different locations of the antenna and 38 different frequencies. For the sake of simplicity, we change the current matrix which is a 3D matrix into a

2D one of 1296 x 38. Since the device showed better performance in lower frequencies, we use only the first 15 frequencies that represents our features in the classification section.

We have different labels for vertical locations that display the place of tumor from chest to nipple. 1 is close to chest, 2 is in the middle of the breast, 3 is close to nipple. There are also other labels that represents those type of tumors that is spread form a place in the breast to another part. When the tumor is spread from chest to the middle of the breast it is labeled as 1 to 2. For the cases in which tumor is spread from chest to nipple we use 1 to 3 label. For the patients that were diagnosed with tumors with labels of 1 to 2 and 1 to 3, we consider each label, 1, 2, and 3 as valid one, so we replicate the data.

After primary preprocessing of the dataset we get a matrix of 115200 x 15 that is a huge matrix and we need to shrink this matrix to a smaller one. Thus, we do it as another step of preprocessing, as we noticed the difference in the mean of depth locations, we get a mean over each frequencies and eventually the dataset has been saved as a CSV file.

Procedure:

1. Acquire dataset from SAFE breast imaging system.
2. Pre-process data and transform it into a CSV file
4. Pre-processed dataset uploaded in Orange for analysis.
5. Applying SVM, Decision tree, and KNN data mining technique.

6. Result

When the training data is modeled, we evaluate the test data and predict the outcome. The labels of test data are recorded and the incorrectly predicted labels are counted, giving us the simplest form of evaluation of the model. The performance of the classifiers is usually evaluated by metrics like classification accuracy, sensitivity, specificity, the area under curve (AUC), etc.

Sensitivity that is known also as true positive rate (TPR) tells us what proportion of a class got correctly classified by the classifiers. The classifier that can correctly classify each group will have a higher result in sensitivity. Sensitivity is defined as follows:

$$Sensitivity\ (\%) = TPR = \frac{TP}{FN + TP} \times 100$$

For multi-class classifications the formula of sensitivity remains the same. We just need to apply it on all three classes individually.

Accuracy is the most crucial metric for model evaluation. It is the ratio of the number of correct predictions to the total number of input samples and it invariably combines specificity and sensitivity for the whole of the data combined. It is formulated as follows:

$$Accuracy\ (\%) = \frac{TP+TN}{TN+FP+TP+FN} \times 100$$

Area Under Curve (AUC)-ROC curve helps us visualize how well our classifier performs. It is a disposition of sensitivity and specificity over all possible thresholds. The higher the AUC, the better the model's performance at distinguishing between the classes.

$$AUC\ (\%) = \frac{1}{2}\left(\frac{TN}{TN + FP} + \frac{TP}{TP + FN}\right) \times 100$$

Specificity of a model evaluates the ability of the algorithm to predict true negative. It basically shows what fraction of negative samples are predicted as negative. It can be applied to categorical models. The equations for calculating these metrics are below:

$$Specifity = \frac{TN}{TN + FP}$$

Throughout this work, three models were closely scrutinized and evaluated for the dataset using the metrics such as sensitivity, accuracy, area under the curve, and specificity as shown in the Table 6.1.

Table 6.1: Summary table of sensitivity, classification accuracy, area under curve, and specificity values by each applied method. SVM: Support Vector Machine, KNN: k-nearest neighbors, DT: Decision tree.

Matric.	Support Vector Machine	k-nearest neighbor	Decision Tree
Sensitivity	52.1	35.7	22.9
Accuracy	52.1	35.7	22.9
AUC	58.2	36.8	32.1
Specificity	75.7	70.75	68.4

We see considering accuracy as the sole criterion for evaluation, that the support vector machine is the best classifier model for the data set. Accuracy itself cannot be used to decide the model to be chosen. At the same time, keeping Sensitivity, Specificity, and AUC as the criteria we have the SVM as the best classifier. The confusion matrix for each model is represented as follows to evaluate the models easily.

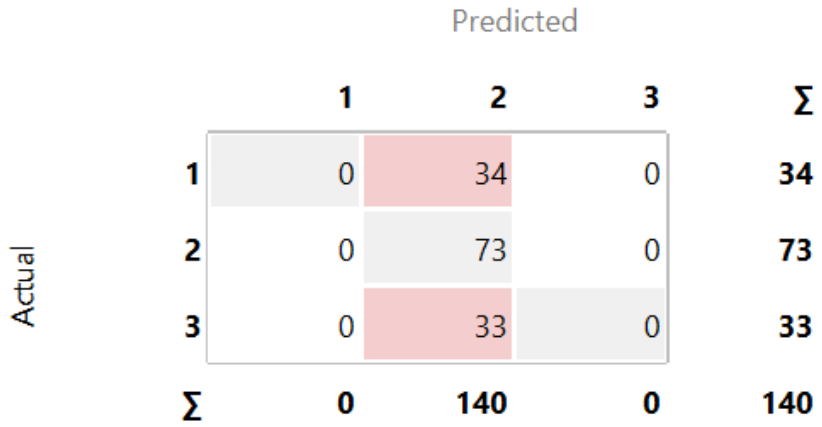


Figure 6.1: Confusion Matrix for SVM classifier

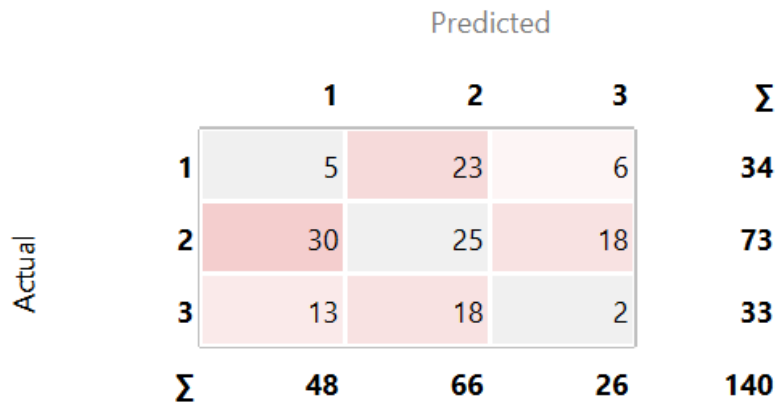


Figure 6.2: Confusion Matrix for Decision Tree classifier

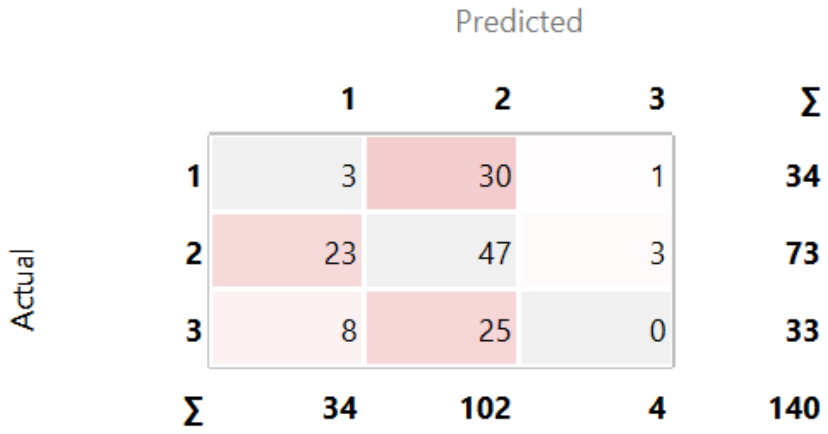


Figure 6.3: Confusion Matrix for KNN classifier

According to the confusion matrix for the SVM classifier that is shown in the Figure 6.1, a total number of 73 samples were correctly predicted out of the total 140 samples. For class 1 the True Positive (TP) is 0 which means none of the samples labeled as one was predicted as 1. False Positive (FP) is 34 which means 34 samples that were actually from class 1 were predicted as class 2 and 3. The True Negative (TN) in this classifier is 106 which shows the number of samples that belong to the classes except 1 and were classified truly. And finally, the False Negative (FN) that represents the number of samples that belongs to the negative group and are classified mistakenly is 0. As we can see in the Table 6.1 the overall sensitivity, specificity, and accuracy achieved for this classifier is 52.1 %, 75.7 %, and 52.1 %, respectively that is higher than the other two classifiers.

Table 6.2: Summary table of true positive, true negative, false positive, and false negative values by each applied method. SVM: Support Vector Machine, KNN: k-nearest neighbors, DT: Decision tree.

Matric.	Support Vector Machine	k-nearest neighbor	Decision Tree
TP	0	3	5
TN	106	75	63
FP	34	31	29
FN	0	31	43

As we mentioned before, for multi-class classifications the formula of sensitivity, specificity, accuracy, and area under the curve remains the same. We just need to apply it on all three classes individually and calculate true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Table 6.2 shows their values by each applied method. SVM: Support Vector Machine, KNN: k-nearest neighbors, DT: Decision tree.

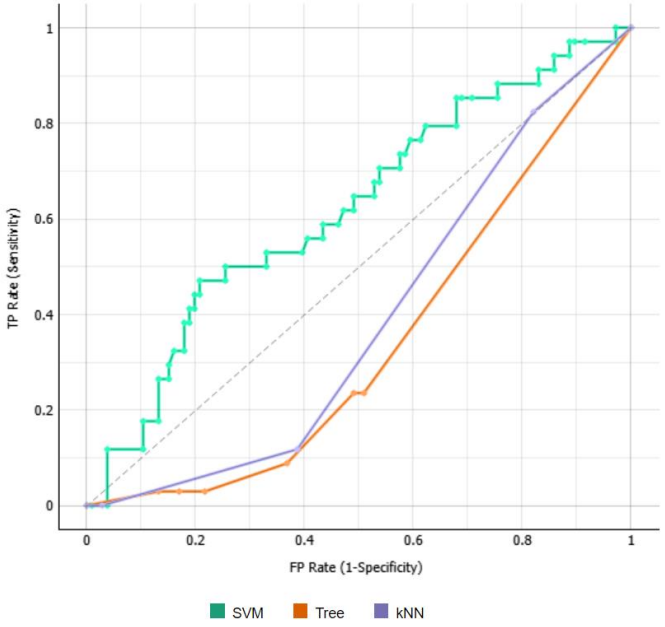


Figure 6.4: ROC analysis for target class 1

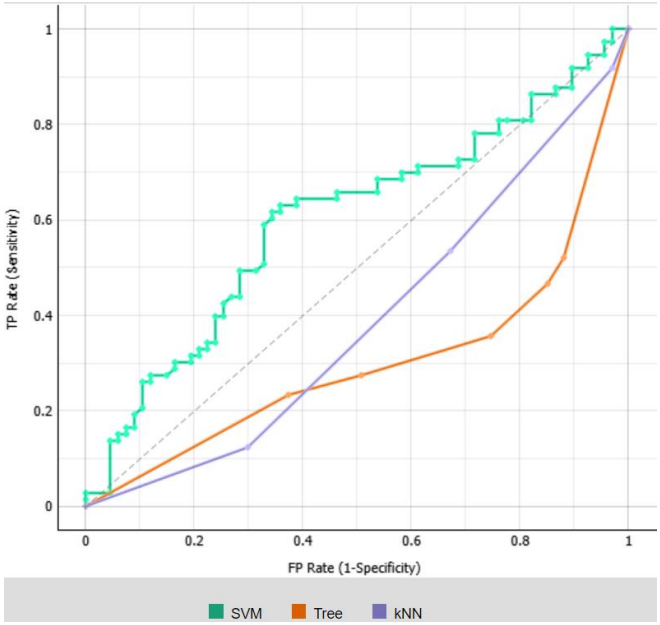


Figure 6.5: ROC analysis for target class 2

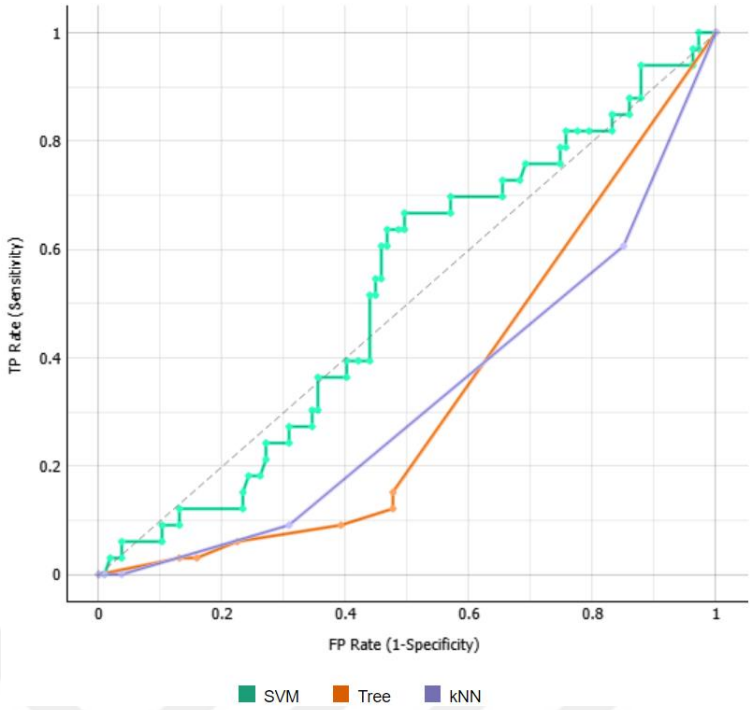


Figure 6.6: ROC analysis for target class 3.

The analysis regarding the effect of the vertical location of the tumor on the classification outcome in comparison to one another for each class is shown in Figures 6.4, 6.5, and 6.6. As it is also discussed based on the confusion matrix, SVM model functions better than the other applied models on our dataset and specifically for each class, however, our models did not provide a good result and high percentage of sensitivity, specificity, and accuracy (higher than 80-85 %) based on the features and attributes we chose. We cannot say with certainty that the vertical position of the tumor provides the doctors with accurate information for the early diagnosis of breast cancer. We may achieve better results by adding other attributes such as breast density, age, or other attributes to our dataset.

7. Conclusion and Discussion

Inspired by the lack of information about the effect of the depth of the tumor on the situation of the patient, in this thesis, three different machine learning approaches were applied to a set of data of 100 patients driven by SAFE microwave breast imaging

system at Marmara University Hospital Breast Center Hospital to figure out the possible contribution of the vertical location of the tumor on the study and early detection of cancer and investigate if we can get useful information from the raw data about the anterior/posterior position of the tumor, which would benefit the doctors to make more accurate clinical predictions.

Our data set has 15 different frequencies as features and three labels that are the depth of the tumor in the breast from closest to the chest labeled as 1, the middle of the breast, labeled as 2, and the one close to the nipple that is labeled as 3. After preprocessing of the data set, three machine learning approaches mentioned earlier were applied and a confusion matrix was driven for each one of the methods. Sensitivity, Specificity, Accuracy, and Area under the curve have been calculated for each algorithm.

Based on the results we reached throughout this work, we noticed that Support Vector Machine provides us with a better result in the sense of Sensitivity, Specificity, Accuracy, and Area under the curve, yet the results do not offer us certainty about the contribution of the vertical location of the tumor on the study and early detection with the features and attributes we chose. However, SVM performed well on the small data set and limited attributes, its performance based on a bigger dataset, considering other attributes such as age and breast density should be further investigated. Other ML approaches can be also considered a part of future works, where additional data collected will be used to enhance the proposed model and classification outcome.

REFERENCES

[1] **World Health Organization**. 2021, https://www.who.int/news-room/fact-sheets/detail/breast-cancer (htt)

[2] **Ferlay J, Laversanne M, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F** (2020). Global Cancer Observatory: Cancer Tomorrow. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/tomorrow, 2021].

[3] **National library of medicine** https://www.ncbi.nlm.nih.gov/books/NBK519575/

[4] **Gratia Medical Center** https://hk-gmc.com/

[5] **National Breast Cancer Foundation**, https://www.nationalbreastcancer.org/breast-anatomy

[6] **Stony Brook Cancer Center**, https://cancer.stonybrookmedicine.edu/breast-cancer-team/patients/bse/breastlumps

[7] **Centers for Disease Control and Prevention (CDC)** https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm

[8] https://www.breastcancer.org/types

[9] **Cancer Research UK,** https://www.cancerresearchuk.org/about-cancer/breast-cancer/stages-types-grades/types

[10] **City of Hope,** https://www.cancercenter.com/cancer-types/breast-cancer/types

[11] **Cancer Australia,** https://www.canceraustralia.gov.au/cancer-types/breast-cancer/symptoms-and-diagnosis/what-are-symptoms-breast-cancer

[12] **American Society of Clinical Oncology (ASCO)**, https://www.cancer.net/cancer-types/breast-cancer/symptoms-and-signs

[13] **University of California San Francisco health,** https://www.ucsfhealth.org/education/breast-cancer-risk-factors

[14] **Iranmakani, S., Mortezazadeh, T., Sajadian, F., Ghaziani, M. F., Ghafari, A., Khezerloo, D., & Musa, A. E.** (2020). A review of various modalities in breast imaging: technical aspects and clinical outcomes. *Egyptian Journal of Radiology and Nuclear Medicine*, *51*(1), 1-22.

[15] **Ebrahimi, Mehran**. "Breast imaging: Mammography, digital tomosynthesis, dynamic contrast enhancement." (2019): 501-504.

[16] **Mohamed, A. A., Luo, Y., Peng, H., Jankowitz, R. C., & Wu, S.** (2018). Understanding clinical mammographic breast density assessment: a deep learning perspective. *Journal of digital imaging*, *31*, 387-392.

[17] **Luiten, J. D., Voogd, A. C., Luiten, E. J., Broeders, M. J., Roes, K. C., Tjan-Heijnen, V. C., & Duijm, L. E.** (2020). Recall and outcome of screen-detected microcalcifications during 2 decades of mammography

screening in the Netherlands National Breast Screening Program. *Radiology*, *294*(3), 528-537.

[18] **Johns Hopkins Medicine,** https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/mammogram-procedure

[19] **Canadian Cancer Society**, 2019, www.cancer.ca

[20] **Janjic, A., Cayoren, M., Akduman, I., Yilmaz, T., Onemli, E., Bugdayci, O., & Aribal, M. E.** (2021). SAFE: A novel microwave imaging system design for breast cancer screening and early detection—Clinical evaluation. *Diagnostics*, *11*(3), 533.

[21] **National Research Council.** "Health risks from exposure to low levels of ionizing radiation: BEIR VII phase 2." (2006).

[22] **Nishimura, D. G.** (1996). Principles of magnetic resonance imaging: Dwight D. *Nishimura.(D. Nishimura, Sl, 2010), ed*, *1*.

[23] **DeMartini, W. B., Ichikawa, L., Yankaskas, B. C., Buist, D., Kerlikowske, K., Geller, B., ... & Lehman, C. D.** (2010). Breast MRI in community practice: equipment and imaging techniques at facilities in the Breast Cancer Surveillance Consortium. *Journal of the American College of Radiology*, *7*(11), 878-884.

[24] **National Cancer Institute** https://www.cancer.gov/

[25] **Morris, E. A., Liberman, L., & Hylton, N. M.** (2005). Breast magnetic resonance imaging techniques. *Breast MRI: Diagnosis and Intervention*, 7-14.

[26]

[27] **Zhang, Y., & Ren, H.** (2017). Meta-analysis of diagnostic accuracy of magnetic resonance imaging and mammography for breast cancer. *Journal of cancer research and therapeutics*, *13*(5), 862-868.

[28] **Hedrick, W. R., Hykes, D. L., & Starchman, D. E.** (2005). Ultrasound physics and instrumentation: practice examinations.

[29] **Brem, R. F., Lenihan, M. J., Lieberman, J., & Torrente, J.** (2015). Screening breast ultrasound: past, present, and future. *American Journal of Roentgenology*, *204*(2), 234-240.

[30] **Guo R, Lu G, Qin B, Fei B** (2018) Ultrasound imaging technologies for breast cancer detection and management: A review. Ultrasound Med Biol 44(1): 37–70

[31] **Vercher-Conejero, J. L., Pelegrí-Martinez, L., Lopez-Aznar, D., & Cózar-Santiago, M. D. P.** (2015). Positron emission tomography in breast cancer. *Diagnostics*, *5*(1), 61-83.

[32] **Koga, S., Nakano, S., Honma, Y., & Ogasawara, N.** (2007). FDG-PET (positron emission tomography) in the detection of primary breast cancer and lymph node involvement. *Nihon rinsho. Japanese journal of clinical medicine*, *65*, 379-384.

[33] **Scheidhauer, K., Walter, C., & Seemann, M. D.** (2004). FDG PET and other imaging modalities in the primary diagnosis of suspicious breast

lesions. *European journal of nuclear medicine and molecular imaging*, *31*, S70-S79.

[34] **Hodgson, N. C., & Gulenchyn, K. Y.** (2008). Is there a role for positron emission tomography in breast cancer staging?. *Journal of Clinical Oncology*, *26*(5), 712-720.

[35] **Krammer, J., Schnitzer, A., Kaiser, C. G., Buesing, K. A., Sperk, E., Brade, J., ... & Wasser, K.** (2015). 18 F-FDG PET/CT for initial staging in breast cancer patients–Is there a relevant impact on treatment planning compared to conventional staging modalities?. *European radiology*, *25*, 2460-2469.

[36] **Krammer, J., Schnitzer, A., Kaiser, C. G., Buesing, K. A., Sperk, E., Brade, J., ... & Wasser, K.** (2015). 18 F-FDG PET/CT for initial staging in breast cancer patients–Is there a relevant impact on treatment planning compared to conventional staging modalities?. *European radiology*, *25*, 2460-2469.

[37] **Liu, Y.** (2014). Role of FDG PET-CT in evaluation of locoregional nodal disease for initial staging of breast cancer. *World journal of clinical oncology*, *5*(5), 982.

[38] **Zucchini, G., Quercia, S., Zamagni, C., Santini, D., Taffurelli, M., Fanti, S., & Martoni, A. A.** (2013). Potential utility of early metabolic response by 18F-2-fluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography in a selected group of breast cancer patients receiving preoperative chemotherapy. *European Journal of Cancer*, *49*(7), 1539-1545.

[39] **Kalles, V., Zografos, G. C., Provatopoulou, X., Koulocheri, D., & Gounaris, A.** (2013). The current status of positron emission mammography in breast cancer diagnosis. *Breast Cancer*, *20*, 123-130.

[40] **Ohring, M., & Kasprzak, L.** (2014). *Reliability and failure of electronic materials and devices*. Academic Press.

[41] **Solimene, R., Cuccaro, A., Ruvio, G., Tapia, D. F., & O'Halloran, M.** (2016). Beamforming and holography image formation methods: An analytic study. *Optics Express*, *24*(8), 9077-9093.

[42] **Nikolova, N. K.** (2011). Microwave imaging for breast cancer. *IEEE microwave magazine*, *12*(7), 78-94.

[43] **Saleh, W., & Qaddoumi, N.** (2009). Potential of near-field microwave imaging in breast cancer detection utilizing tapered rectangular waveguide probes. *Computers & Electrical Engineering*, *35*(4), 587-593.

[44] **O'Loughlin, D., O'Halloran, M., Moloney, B. M., Glavin, M., Jones, E., & Elahi, M. A.** (2018). Microwave breast imaging: Clinical advances and remaining challenges. *IEEE Transactions on Biomedical Engineering*, *65*(11), 2580-2590.

[45] **Modiri, A., Goudreau, S., Rahimi, A., & Kiasaleh, K.** (2017). Review of breast screening: Toward clinical realization of microwave imaging. *Medical physics*, *44*(12), e446-e458.

[46] **Selvanambi, R., & Jaisankar, N.** (2019). Healthcare: Prediction of breast cancer stage using social spider-inspired optimization algorithm. *International Journal of E-Health and Medical Communications (IJEHMC)*, *10*(2), 63-85.

[47] **Sharma, A., Kulshrestha, S., & Daniel, S.** (2017, December). Machine learning approaches for breast cancer diagnosis and prognosis. In *2017 International conference on soft computing and its engineering applications (icSoftComp)* (pp. 1-5). IEEE.

[48] **Janjic, A., Akduman, I., Cayoren, M., Bugdayci, O., & Aribal, M. E.** (2022). Gradient-Boosting Algorithm for Microwave Breast Lesion Classification—SAFE Clinical Investigation. *Diagnostics*, *12*(12), 3151.

[49] **Sawarkar, S. D., Ghatol, A. A., & Pande, A. P.** (2006, June). Neural network aided breast cancer detection and diagnosis using support vector machine. In *Proceedings of the International conference on Neural Networks, Cavtat, Croatia* (pp. 158-163).

[50] **Keerthika, J., Sruthi, D., Swathi, D., Swetha, S., & Vinupriya, R.** (2021, March). Diagnosis of breast cancer using decision tree data mining technique. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 1530-1535). IEEE.

[51] **Al-Salihy, N. K., & Ibrikci, T.** (2017, February). Classifying breast cancer by using decision tree algorithms. In *Proceedings of the 6th International Conference on Software and Computer Applications* (pp. 144-148).

[52] **Elsayad, A. M., & Elsalamony, H. A.** (2013). Diagnosis of breast cancer using decision tree models and SVM. *International Journal of Computer Applications*, *83*(5).

CURRICULUM VITAE

Name Surname : Golshan GHOLAMPOUR

EDUCATION :

- **B.Sc. :** 2014, Azad University, Electrical Engineering, Electrical and Electronics Engineering

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Gholampour, G., Shayea, I., Saad, S. A., & Roslee, M.** (2021, December). Outage Probability of a CRAN in the presence of Channel Estimation Error. In *2021 IEEE 15th Malaysia International Conference on Communication (MICC)* (pp. 138-142). IEEE.
- **Hamza, B. J., Saad, W. K., Shayea, I., Ahmad, N., Mohamed, N., Nandi, D., & Gholampour, G.** (2021). Performance enhancement of SCM/WDM-RoF-XGPON system for bidirectional transmission with square root module. *IEEE Access*, *9*, 49487-49503.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS: