UNIVERSITA' POLITECNICA DELLE MARCHE

ANCONA

DEPARTMENT OF ENGINEERING

MASTER'S DEGREE IN BIOMEDICAL ENGINEERING

DEEP LEARNING EVALUATION OF THERAPEUTIC EFFICACY IN BREAST CANCER FROM RADIOLOGICAL IMAGES

Candidate: Aurora Lanciotti

Supervisor: Prof. Laura Burattini

Co-Supervisor: Dott. Agnese Sbrollini

Academic year 2022/2023

ABSTRACT

Breast cancer is the leading cause of cancer-related death in women world-wide. Among all the malignant diseases, it accounts for 23% of all cancer deaths. The highest incidence occurs in women over age 50. The heterogeneity of breast carcinoma can arise from neoplastic transformations in myoepithelial or epithelial cells, or even from a stem cell that has the ability to change its nature. Breast cancer can be classified by different criteria regarding histopathological type, grade, stage and receptor status.

It is generally diagnosed through screening or in response to the presence of symptoms that prompt further diagnostic examinations. Imaging techniques are the most used for tumor detection and the principal ones are: mammography, ultrasound, positron emission tomography and magnetic resonance imaging. Specifically, tumor identification in breast MRI images relies on the different contrast enhancement between normal tissues and breast lesions.

Among all different types of tumors treatments, surgical resection and chemoradiotherapy are often the principal ones. Recently, new therapies that reactivate immune responses against cancer, have discovered the presence of specific killing lymphocytes in the tumor microenvironment, the so called tumor-infiltrating lymphocytes (TIL). The presence of TIL in breast cancer prior to treatment can predict the response to therapy and it is associated with a better prognosis.

In recent years, there has been significant progress in Deep Learning (DL) techniques, particularly in the automatic analysis of radiological images for tumor detection and the prediction of therapeutic efficacy. The aim of this study is to assess TIL from breast cancer MRIs, acquired from the Cancer Genome Atlas Breast Invasive Carcinoma data collection. The study relies on three experiments performed using three different classifiers: the VGG-16 and two simpler CNNs, one with a single convolutional layer and the other with four convolutional layers.

All these models were fed with MRIs pre-processed with the following customized pipeline: normalization, bias field correction and inter-slice distance modification. Afterward, the hyper-parameter tuning was utilized to find the best parameters to train, validate and test the models.

The performances of the three model architectures were compared in terms of accuracy, sensitivity, specificity, area under (AUC) the curve and F_1-score. It turned out that the model of intermediate complexity reached better performances with AUC of 0.943, 0.962, 0.990, 0.985 and 0.988 in different splits.

The achieved results demonstrate that the classifier used in the third experiment is able to accurately discriminate among low TIL and high TIL tumors from radiological images. This is a very

promising starting point for the development of a clinically useful computer aided system to assess TIL levels from radiological images.

CONTENTS

INTRODUCTION

Breast cancer is one of the principal leading cause of cancer-related deaths and the most diagnosed cancer in women across the world, with an estimated 55,720 cases of non-invasive carcinoma and 297,790 cases of invasive disease in 2023. Its incidence and death rates have increased over the last three decades due to the change in risk factor profiles, better cancer registration and detection. Currently, about 80% of breast cancer patients are individuals aged over 50. Several procedures such as screening programs are potentially reducing breast cancer incidence rate. There are specific factors that increase the risk of breast cancer development including: advancing age, obesity, use of alcohol, breast cancer family history, radiation exposure, reproductive history (such as age at first menstrual period and age at first pregnancy), smoking and postmenopausal hormone therapy. Approximately half of breast cancers develop in women who have no identifiable breast cancer risk factor other than gender (female) and age (over 40 years). Furthermore, specific inherited gene mutations significantly elevate the risk of breast cancer, with BRCA1, BRCA2, and PALB-2 gene mutations being the most prevalent. Among the various diagnostic techniques, imaging plays a central role in the identification and assessment of breast cancer. In particular, techniques as mammography, magnetic resonance imaging, positron emission tomography, single-photon emission computed tomography, breast specific gamma imaging and ultrasound have been emerged as powerful tools for detection and monitoring of response to therapy. All of these techniques are currently digitized, enabling the possibility to implement deep learning in breast imaging. Deep learning is nowadays applied to different tasks, such as lesion classification and segmentation, image reconstruction, cancer risk evaluation and prediction and assessment of therapy response. Breast cancer is commonly treated by various combinations of surgery, radiation therapy, chemotherapy, hormone therapy and immunotherapy. Recently, new therapies that reactivate anticancer immune responses, have entered clinical practice and got an improved outcome. Several clinical studies have evaluated the prognostic and predictive importance of tumor-infiltrating lymphocytes (TILs) in breast cancer. TILs are the core of an experimental cell therapy for the treatment of solid tumors. Lymphocytes, made up of T cells and B cells, are part of the immune system and are constantly controlling the body, identifying modified cells, including cancerous ones. As tumors progress, lymphocytes discern these abnormal cells and infiltrate the tumor to build a defence against their presence.

In scientific literature, there are few sources regarding the prediction of therapy outcomes combined with deep learning applications, especially concerning TILs.

This is largely due to the fact that TILs are not still officially recognized as a biomarker. Studies exploring the use of TIL as a predictive parameter for the effectiveness of breast anti-tumor therapies are still in experimental phase.

Thus, the aim of this study is to use different DL models for the assessment of TIL from breast MRI cancer images.

1. BREAST CANCER

Breast cancer is the leading cause of cancer-related death in women world-wide [1]. Among all the malignant diseases, it accounts for 23% of all cancer deaths. The incidence of breast cancer rises after age 40. The highest incidence (approximately 80% of invasive cases) occurs in women over age 50 [2].

1.1 ANATOMY OF THE BREAST

The breast of an adult woman is a milk-producing, tear-shaped gland. It is supported by and attached to the front of the chest wall on either side of the breast bone or sternum by ligaments. The skin of the breast is thin and contains hair follicles, sebaceous glands, and sweat glands. Two fascial layers are present within the breast. The posterior surface of the breast rests on segment of the fasciae of the pectoralis major, rectus abdominis muscles, external abdominal oblique and serratus anterior [1]. The superficial fascia lies deep to the dermis and the deep fascia lies anterior to the pectoralis major muscle. The breast has no muscle tissue and a layer of fat surrounds the glands and extends throughout the breast. The breast is responsive to a complex interplay of different hormones that cause the tissue to develop, enlarge and produce milk. The three major hormones affecting the breast are estrogen, progesterone and prolactin, which are responsible for changes of glandular tissue in the breast and of the uterus during the menstrual cycle [3].

The breast is composed of approximately 15 to 20 lobes and these lobes are further divided into lobules. The lobules are made up of branched alveolar glands. Each lobe drains into a major lactiferous duct (Figure 1.1).

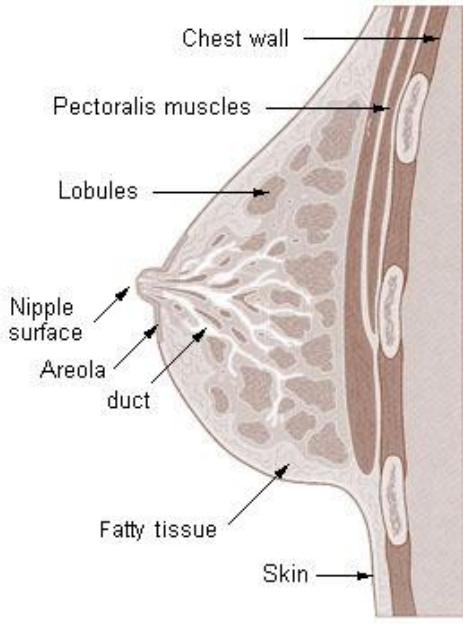

Figure 1.1: Anatomy of the breast

The lactiferous ducts dilate into a lactiferous sinus beneath the areola, a spherical darker-pigmented area, and then open through a constricted orifice onto the nipple. The space between the lobes is filled by connective tissue and subcutaneous adipose tissue. The fat that covers the lobes gives the breast its size and shape. During the development of the breasts, at the time of puberty, the ducts grow and divide and form terminal end buds. The terminal end buds then form new branches and small ductules termed alveolar buds. The alveolar buds differentiate into the terminal structure of the resting breast called acines or ductules [4]. Typically, there are hundreds of acinar cells within each breast. The Terminal Ductal Lobular Unit (TDLU) refers to the basic functional unit of the breast with 30 to 50 acinar cells grouped into a lobule and the associated ducts as reported in Figure 1.2.
In the immature breast, ducts and alveoli are lined by a two-layer epithelium that consists of a basal cuboidal layer and flattened surface layer. In the adult breast, the epithelium proliferates and becomes multilayered. Myoepithelial cells are located surrounding the alveoli and are contractile units that are stimulated by hormones such as prolactin and oxytocin [1]. If pregnancy occurs, the breast lobules which are composed of epithelial lobular cells and underlying contractile myoepithelial cells differentiate for lactation. With the presence of prolactin hormone, the lobules are stimulated to produce and secrete milk.

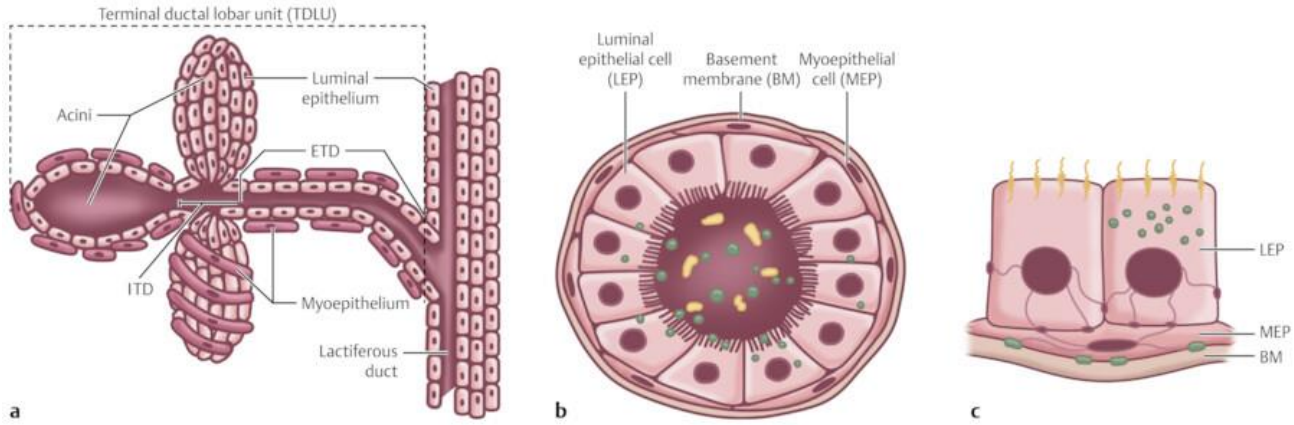

Figure 1.2: (a-c) Terminal Ductal Lobular Unit (TDLU) composed of Intralobar Terminal Duct (ITD) and Extralobar Terminal Duct (ETD). The inner secretory luminal epithelial cells (LEP) is shown at cellular level lining the acini (ductules). The outer layer is made of myoepithelial cells (MEP) which are the contractile units. The basement membrane (BM) is deep to the myoepithelial cell layer.

1.2 CARCINOGENESIS AND CAUSES

Breast cancer is a malignant disease that initiates in breast cells. In the breast, the stratified epithelium is composed of two dissimilar cell populations, myoepithelial and epithelial. The heterogenicity of the breast carcinoma might happen from the neoplastic change of either myoepithelial or epithelial cells, or yet from a stem cell that has the ability to develop into myoepithelial or epithelial cell [5]. According to the oncology of breast cancer, neoplastic cells differ from the normal body cells. Normal tissues of the body have limited growth promotion and regulation which helps to keep the structure and functions of tissues in normal conditions. Instead, cancerous cells have prolonged and chronic proliferation without any external stimuli [6] by overcoming the growth suppressor genes. There are numerous causes that can increase the possibility of developing breast cancer. For example, injuries to the deoxyribonucleic acid (DNA) and hereditary alteration can guide to breast cancer and they have been associated with the exposure to estrogen. Some patients inherit fault in the deoxyribonucleic acid (DNA) and genes like the P53, BRCA1 and BRCA2 and the ones with a family history of breast or ovarian cancer have more possibility of developing breast cancer [7]. The immune system usually tries to find out cancer cells and cells with injured deoxyribonucleic acid (DNA) in order to demolish them. But, if there is a malfunctioning of the immune defence and surveillance system, breast cancer might develop. Moreover, breast cancer commonly occurs due to an association between genetic and environmental factors: usually, RAS/MEK/ERK pathway and PI3K/AKT pathway defend normal cells from cell suicide. When mutation occurs in genes that are involved in encoding these protective pathways, the cells become unable of committing suicide when they are no longer required leading to cancer. These mutations were confirmed to be experimentally associated with estrogen exposure [8]. Modifications or mutations can occur spontaneously or they may be induced by other factors such as nuclear radiation, electromagnetic radiation (microwaves, X-rays, Gamma-rays, Ultraviolet-rays etc.), viruses, bacteria and fungi, parasites, heat, chemicals in the air, water and food, mechanical cell-level injury, free radicals and ageing of DNA and RNA. In addition, deformity in the growth factors signalling or over expression of leptinin breast adipose tissue, enhances proliferation of cell [9]. In fact, in cancer cells, enzyme telomerase turns away the chromosomal shortening and allows the extensive replication of cells [10]. Moreover, tumour cells get their nutrients and oxygen supply by angiogenesis, the physiological process through which new blood vessels are born from pre-existing vessels: they break their boundaries and can enter into the blood, lymphatic tissues and other tissues of the body to produce a secondary tumour [1].

1.3 BREAST CANCER CLASSIFICATION

Breast cancer can be classified by different criteria regarding histopathological type, grade, stage and receptor status [7].

The grading of breast cancer depends on the microscopic similarity of cancer cells to normal breast tissue and, classifies the cancer as well differentiated (grade 1), moderately differentiated (grade 2) and poorly differentiated (grade 3 or grade 4).

According to site, breast cancer is divided into invasive and non-invasive: non-invasive breast cancer is a cancer that has not extended away from the lobule or ducts where it situated [12], as Lobular carcinoma in situ (LCIS) and Ductal carcinoma in situ (DCIS). Instead, invasive breast cancer exists when abnormal cells from within the lobules or milk ducts split out into close proximity of breast tissue, as Infiltrating lobular carcinoma (ILC) and Infiltrating ductal carcinoma (IDC).

Less common types of breast cancer include [3]:

- Inflammatory breast cancer: it diffuses brawny infiltration; the breast appears red or inflamed and tends to spread quickly
- Medullary carcinoma: it originates in central breast tissue
- Mucinous carcinoma: it is invasive and usually occurs in postmenopausal women
- Paget disease of the nipple: it originates in the milk ducts and spreads to the skin of the nipples or areola
- Phyllodes tumor: it is a tumor with a leaf-like appearance that extends into the ducts and that rarely metastasizes
- Tubular carcinoma: small tumor that is often undetectable by palpation

1.4 BREAST CANCER STAGES

Breast cancer staging is based on tumour size and it describes if and where the cancer has spread.

The international standard for classifying the malignancy of a tumor is the TNM Classification of Malignant Tumors (TNM) system: it describes the anatomic extent of cancer and determines its stage based on a list of factors involving tumor (T), node (N) and metastases (M) [13]. The T category refers to the cancer size and describes the original primary tumor:

- TX: the tumor can't be assessed.
- T0: there isn't any evidence of the primary tumor.

- Tis: the cancer is in situ (the tumor has not started growing into healthy breast tissue).
- T1, T2, T3, T4: These numbers are based on the size of the tumor and how it has grown into the surrounding breast tissue. The higher the T number, the larger the tumor and its spread thrown the breast tissue.

The N category refers to lymph node involvement and describes whether or not the cancer has reached nearby lymph nodes:

- NX: nearby lymph nodes can't be assessed.
- N0: nearby lymph nodes do not contain cancer.
- N1, N2, N3: These numbers are based on the number of lymph nodes involved and how the cancer has been spread between them. The higher the N number, the greater the extent of the lymph node involvement.

The M category express whether or not there is evidence of diffusion of the cancer in other parts of the body:

- MX: metastasis can't be assessed.
- M0: there is no distant metastasis.
- M1: distant metastasis is present.

Once the T, N, and M categories have been determined, it is possible to assess the cancer overall stage.

1.4.1 STAGE 0

This is the non-invasive stage of tumour which indicates that both cancerous and non-cancerous cells are within the boundaries of that part of the breast in which the tumor begins to grow [14]. Stage 0 is used to describe non-invasive breast cancers, such as Ductal carcinoma in situ (DCIS).

1.4.2 STAGE 1

This stage describes the invasive breast carcinoma as when the microscopic invasion is possible [15] and it is divided into two subcategories known as IA and IB. In general, stage IA describes invasive breast cancer in which the tumor measures up to 2 cm and the cancer has not spread outside the breast (no lymph nodes are involved). Instead, stage IB describes invasive breast cancer in which small groups of cancer cells (larger than 0.2 mm but not larger than 2 mm) are found in the lymph nodes and there may be or not tumor in the breast.

1.4.3 STAGE 2

Stage 2 describes that the tumour can be found in axillary lymph nodes or in sentinel lymph nodes and is referred to as early breast cancer. This stage is divided into two subcategories: IIA and IIB.

In general, stage IIA describes invasive breast cancer in which: no tumor can be found in the breast, but cancer (larger than 2 mm) is found in one to three axillary lymph nodes (the lymph nodes under the arm) or in the lymph nodes near the breastbone. In other cases, the tumor measures 2 cm or smaller and has spread to the axillary lymph nodes or the tumor is larger than 2 cm but smaller than 5 cm and it has not spread to the axillary lymph nodes.

Stage IIB refers to invasive breast cancer in which the tumor is larger than 2 cm but smaller than 5 cm and small groups of breast cancer cells (larger than 0.2 mm but smaller than 2 mm) may be found or not in the lymph nodes or in one to three axillary lymph nodes or into lymph nodes near the breastbone.

1.4.4 STAGE 3

Stage 3 is referred to as locally advanced breast cancer. This is an invasive breast cancer that may be large (typically bigger than 5 cm) or has spread to several lymph nodes in the underarm area or other areas near the breast. Stage III is divided into subcategories known as IIIA, IIIB, and IIIC.

In general, stage IIIA describes invasive breast cancer in which no tumor is found in the breast but is found in four to nine axillary lymph nodes or in the lymph nodes near the breastbone. In other cases, the tumor in the breast is larger than 5 cm and small groups of breast cancer cells are found in the lymph nodes or, cancer has spread to one to three axillary lymph nodes or to the lymph nodes near the breastbone.

Stage IIIB describes invasive breast cancer in which the tumor may be any size and it has spread to the chest wall and/or skin of the breast and caused swelling or an ulcer and may have spread to up to nine axillary lymph nodes or lymph nodes near the breastbone.

Stage IIIC refers to invasive breast cancer in which there may be no sign of cancer in the breast or, if there is a tumor, it may be any size and may have spread to the chest wall and/or the skin of the breast and the cancer has spread to 10 or more axillary lymph nodes or lymph nodes above or below the collarbone or breastbone.

1.4.5 STAGE 4

This is the advanced and metastatic stage of cancer and this stage describes the spread to other organs of the body that could be lungs, bones, liver, brain etc [16].

1.5 TUMOR INFILTRATING LYMPHOCYTE

Maintenance of tissue homeostasis by continuous immunosurveillance and initiation of inflammatory reactions is a fundamental role of the immune system that involve the coordinated activation of innate and adaptive immune cells [17]. Neoplastic transformation alters the structure of tissues and induces immune responses that can eliminate incipient tumors. In situations where elimination is incomplete, neoplastic transformation of cells is able to escape immune control and as a consequence, a cancer can develop.

Among all different types of tumors treatments, surgical resection and chemoradiotherapy are often the principal ones. These traditional treatments focus on the tumor cells, without taking into account the complex tumor microenvironment (TME). TME denotes the complex internal environment generated by tumor cells, including cell components, such as tumor cells, fibroblasts, immune cells, and extracellular components such as cytokines, growth factors and extracellular matrix. There is an interactive relationship between tumors and TME. A tumor can influence its microenvironment by producing cell signalling molecules to enhance tumor angiogenesis and trigger immune tolerance. Immune cells in the microenvironment can influence the growth as well as the development of cancerous cells. Recently, new therapies that reactivate immune responses aganist cancer, have discovered the presence of specific killing lymphocytes in the tumor microenvironment, the so called **tumor-infiltrating lymphocytes (TIL)** [18].

TIL can affect TME through the interaction between different cells, thus influencing the occurrence and the development of the tumor exerting an anti-tumor effect directly or indirectly. In addition, TIL establishes a complex network of intercellular interactions that help to maintain the immunosuppressive microenvironment, promoting immune escape and triggering tumors.

So, TIL are associated with cancer immune regulation and TME, and play a fundamental role in tumor genesis, development, metastasis and prognosis.

On this basis, scientists use in vitro culture methods to enrich some of the tissue lymphocytes and then transfuse them to patients in order to improve the anti-tumor immunological response.

Compared to other types of cellular immunotherapies, TIL is derived from patients themselves, without genetic modification and so, it has a specific improved killing effect on tumor cells [19].
TIL mainly consists of T-cells, B-cells and Natural killer cells (NK) and it acts differently in different subtypes of breast cancer.
In general, Intratumoral lymphocytes (iTu-Ly) are defined as intraepithelial mononuclear cells within tumor cell nests or in direct contact with tumor cells and are reported as the percentage of the tumor epithelial nests that contain infiltrating lymphocytes (Figure 1.3). Instead, Stromal lymphocytes (str-Ly) are defined as the percentage of tumor stroma area that contains a lymphocytic infiltrate without direct contact to tumor cells [20].

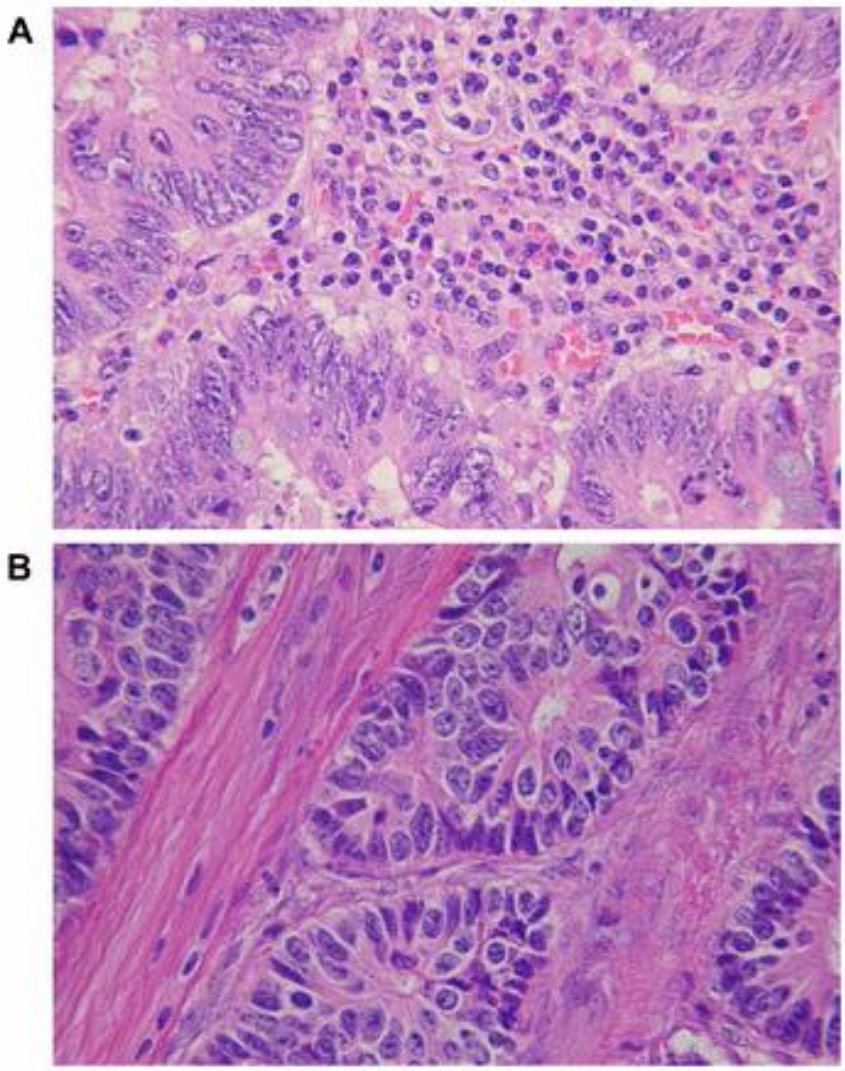

Figure 1.3: Representative pictures of tumor-infiltrating lymphocytes: (A) High density of inflammatory cell infiltration. (B) Low density of inflammatory cell infiltration.[20]

1.5.1 TUMOR INFILTRATING LYMPHOCYTE AS A MEASURE OF IMMUNOTHERAPY EFFICACY PREDICTION

Several recent clinical studies have evaluated the prognostic and predictive importance of tumor-infiltrating lymphocytes.

Infiltration of immune cells, particularly infiltration of anti-tumor lymphocytes, is associated to an improved prognosis in many different tumor types including colon, ovarian, lung and breast cancer. Historically breast cancer was not thought to be immunologically active, particularly when compared to tumors such as melanoma. However recent studies have demonstrated that the presence of TILs in breast cancer prior to treatment can predict the response to therapy and moreover, it is associated with a better prognosis [21]. Not only the amount of lymphocytic infiltration but also the phenotype of that infiltrate is able to determine clinical outcome. Histological as well as molecular data indicate that immunological parameters, including stromal TILs are associated with higher rates of pathological complete remission (pCR), independent of other clinico-pathological prognostic factors or of the chemotherapy regimen [18]. It has been shown that B-cells and T-cells markers besides chemokines and their receptors are expressed in the stromal infiltrate and are linked to therapy response.

It may be speculated that the destruction of tumor cells by therapeutic agents may release tumor-associated antigens. This could trigger an immune response against the tumor cells, which will be particularly strong in those patients who have a sensitization of the immune system against some tumor antigens prior on the onset of therapy. Therefore, the combination of the therapeutic destruction of tumor cells and an increased immune response could result in a pCR.

Some studies explored the effect of TIL in primary tumors on the efficacy of radiotherapy after breast-conserving surgery. The results showed that high density of TIL was associated with a lower risk of recurrence in patients with ipsilateral breast cancer. Other studies pointed out that the increased TIL density in HER2-positive and triple-negative breast cancer (TNBC) patients often means a longer survival time. However, for patients with HER2-negative ductal breast cancer, the density of TIL was negatively correlated with prognosis. This suggests that the tumor subtype has different immunological osmosis [19].

Generally, the higher expression of an immune signature reflects a greater presence of immune cells, that are associated with a better prognosis. Despite the immune system's failure to contain the growing tumor, the presence of follicular helper T (Tfh) cells is associated with organized immune structures adjacent to the tumor bed that potentially participate in propagating sustainable

and effective long-term antitumor immunity. Leukocytes within the tertiary lymphoid structures (TLS) could thus dynamically respond to evolving local tumor antigens, maintain activation, and generate long-lived memory cells while sequestered in a protected microenvironment. The mixture of activation and suppression of regulators of immune responses as CD4+ TIL, may reflect the presence of functionally activated cells from the TLS combined with suppressed cells migrating through the intratumoral and peritumoral regions. Despite their imminent suppression, once they move from the TLS, some tumor-specific memory cells survive and maintain effective immunosurveillance that functions over the long-term to detect and eliminate residual tumor cells. Thus, patients whose immune systems have an organized response to their tumors, specifically detectable by the Tfh signature, would be predicted to have a better response to preoperative chemotherapy or post-surgery [22].

2. BREAST CANCER DIAGNOSIS

Breast cancer is the most frequent cancer and different types of diagnostic examinations are available, such as mammography, Magnetic Resonance Imaging (MRI), biopsy, ultrasound (US) and molecular imaging. Radionuclide-based imaging methods including Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) are useful in diagnosis and treatment of the cancer. The radiolabelling of chemo drugs with nanoparticles should be recommended from the standpoint of an early diagnosis and effective treatment of breast cancer [23]. Therefore, many novel technologies are being developed for early detection of primary tumors, as well as distant metastases and recurrent disease. Theranostics has emerged as a new paradigm for the simultaneous diagnosis, imaging and treatment of cancers. It has the potential to provide timely and improved patient care via personalized therapy. In nanotheranostics, cell-specific targeting moieties, imaging agents and therapeutic agents can be embedded within a single formulation for effective treatment [24].

2.1 SCREENING

Breast cancer is generally diagnosed through either screening or the presence of symptoms (pain or a palpable mass) that prompt a diagnostic exam. Screening of healthy women is associated with the detection of tumors that are smaller, have lower odds of metastasis, are more amenable to breast-conserving and limited axillary surgery, and are less likely to require chemotherapy [25]. This scenario translates to reduced treatment-related morbidity and improved survival. The only screening modality proven to reduce breast cancer-specific mortality is mammography [26]. The age at which various organizations recommend beginning screening mammography and the frequency at which mammography is recommended vary based on the weight given to the perceived risks (false-positive examinations and the possibility of over-diagnosis) and benefits of screening (mortality reduction and less invasive treatment options). Some groups suggest mammography screening for all women starting at age 50 as an annual or biennial examination. Based on a review of subsequent meta-analyses, the American College of Radiology (ACR) recommends annual screening beginning at 40 years of age [27]. According to the Italian Ministry of Health, breast cancer screening programs involve women between 50 and 69 years old and consist of a free mammography examination every two years. In some regions, different screening programs involving a wider age range are being experimented, including women between 45 and 49 years old, invited to perform mammography every year. Instead, for women between 50 and 74

years old it is recommended a mammography every two years. The addition of digital breast tomosynthesis to a conventional full-field digital mammography examination reduces false-positive results and increases cancer detection. Further, the use of MRI as an adjunct to mammography had a higher sensitivity for malignancy (92.7%) than the use of ultrasound as an adjunct to mammography (52%). As a result, for women with a lifetime risk of breast cancer of greater than 20%, breast MRI as an adjunct to mammography is recommended [28].

2.2 DIAGNOSTIC IMAGING TECHNIQUES FOR BREAST CANCER DETECTION

Early detection and monitoring of patients in response to various types of therapies are important aspects of breast cancer therapy. The utilization of imaging techniques is one of main approaches for tumor detection and assessment response to therapy in patients with breast cancer. Multiple studies indicated that these techniques could provide effective information on breast tumor morphology and functional and metabolic processes within the tumor at various levels [27]. The principal types of diagnostic imaging techniques actually used for breast cancer detection are: Mammography, Digital Breast Tomosynthesis, Contrast Enhanced Spectral Mammography, Magnetic Resonance Imaging, Dynamic Contrast Agent Enhanced breast MRI, Diffusion Weighted Imaging, Magnetic Resonance Elastography, Magnetic Resonance Spectroscopy, Positron Emission Tomography, Single-photon Emission Computed Tomography, Breast Specific Gamma Imaging and Ultrasound. However, further improvements of these medical imaging techniques are needed in order to overcome current limitations and to increase their effectiveness in detecting breast cancer.

2.2.1 MAMMOGRAPHY

Mammography is a dedicated radiographic technique for imaging the breast and the resultant images are known as mammograms. The images are obtained by applying a small dose of radiation (x-ray) through the breast post compression between two plates [29]. It is considered to be the gold standard for early detection of breast cancer and it has the ability to detect very small lesions (less than 1 cm size) such as microcalcifications or architectural distortions that can be a signal of a non-symptomatic breast tumor. The predictive diagnostic value of mammography largely depends on breast density: glandular tissue appears white and opaque on the mammograms, such as calcifications or tumor lesions (Figure 2.1). Thus, the efficiency of tumor detection is higher in fatty

breasts (up to 90%) that look darker on mammogram while it decreases with very dense breasts (80-85%).

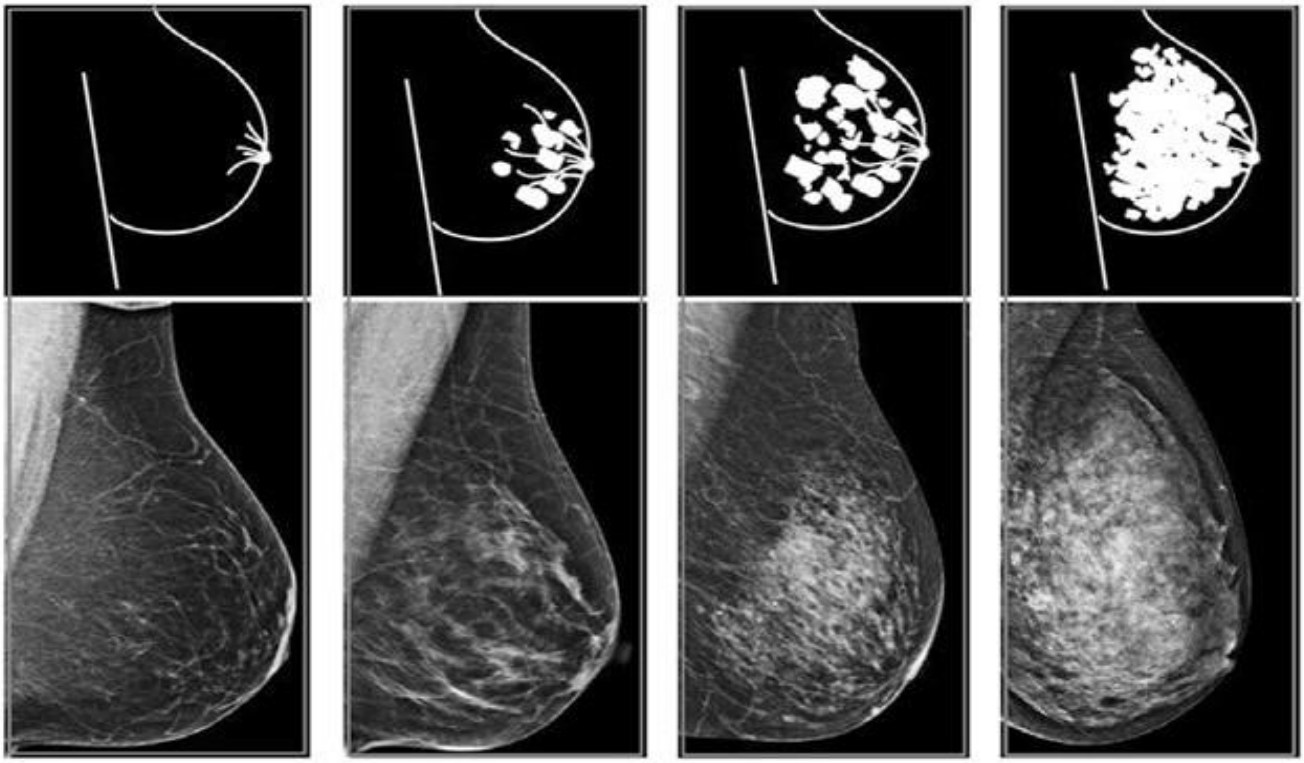

Figure 2.1: Mammograms related to different types of breast tissue according to their density, from fatty breast to fibrous and dense breast.

2.2.2 DIGITAL BREAST TOMOSYNTHESIS (DBT)

Digital Breast Tomosynthesis (DBT) is an advanced imaging technique that allows a volumetric reconstruction of the whole breast from an arbitrary number of low-dose two-dimensional (2D) images (projection) obtained by different X-ray tube angles, with a geometric principle very similar to that applied by stratigraphic technique (Figure 2.2). In addition to planar images, DBT allows to create and view thin-section reconstructed images that may decrease the lesion-masking effect of overlapping normal tissue and reveal the true nature of potential false-positive findings without the need for recall [30].

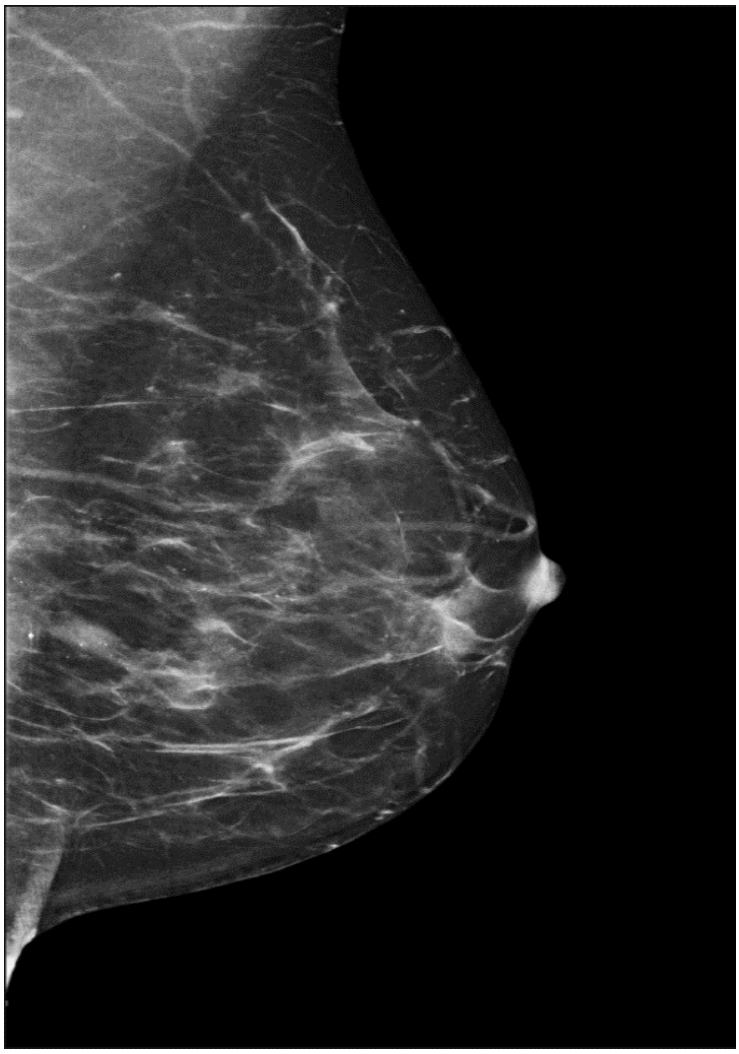

Figure 2.2: Digital Breast Tomosynthesis image

2.2.3 CONTRAST ENHANCED SPECTRAL MAMMOGRAPHY (CESM)

Contrast Enhanced Spectral Mammography (CESM) is an advanced imaging technique combining conventional mammography principles with an intravenous contrast agent injection, that allows a contrast enhancement evaluation of the breast. It highlights breast areas with contrast which is typical of cancer neo-angiogenesis. It is a bilateral examination that works with a double exposure (low and high energy acquisitions) in a single breast compression from which 4 standard projections are obtained (Figure 2.3).

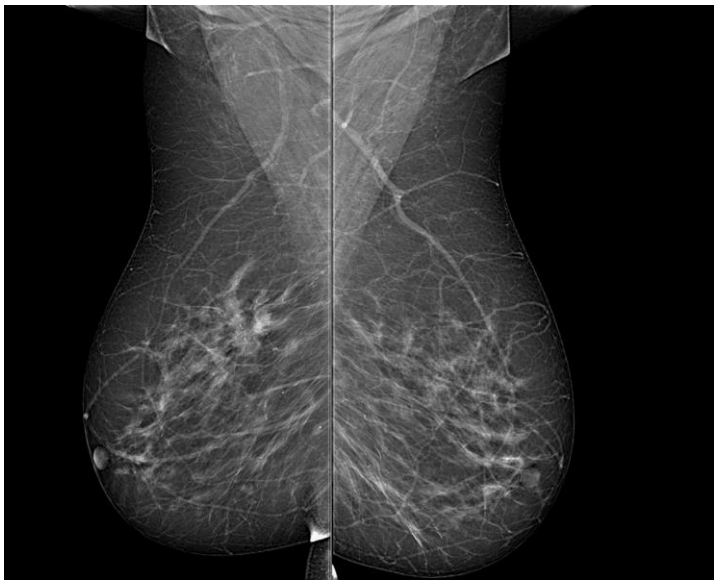

Figure 2.3: Bilateral CESM breast image

2.2.4 MAGNETIC RESONANCE IMAGING (MRI)

Breast MRI is a non-invasive and non-ionizing diagnostic imaging tool that employs low-energy radio frequency waves and a magnetic field to obtain detailed images of structures within the breast [30]. MRI can be used to measure the size of the cancer and look for metastasized tumors in women who have been previously diagnosed with breast cancer. Tumors with size less than or equal to 2 cm have been accurately identified and measured using MRI. It has the ability to detect neo-angiogenesis and suspected breast malignancies that often escape clinical, mammographic and ultrasound detection [31]. Breast MRI provides more detailed images because of its high spatial resolution (Figure 2.4) but, by another hand, it has too low specificity that results in further tests or biopsies, being expensive and not standardized [32].

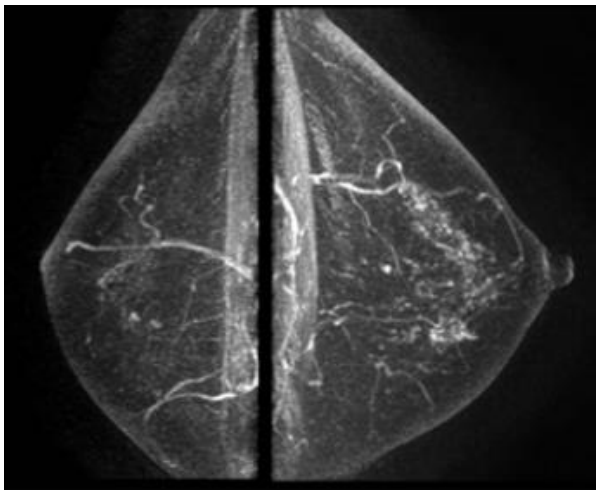

Figure 2.4: Breast MRI image

2.2.5 DYNAMIC CONTRAST ENHANCED MRI (DCE-MRI)

Dynamic Contrast Agent Enhanced breast MRI (DCE-MRI) works by analysing the temporal enhancement pattern of a tissue following the intravenous injection of a paramagnetic contrast agent. This non-invasive imaging technique quantitatively determines the extent of tissue vascularization, interstitial space composition, and differentiation of lesions [31]. This imaging modality is useful to depict tumour angiogenesis with overall recurrence and overall survival of breast cancer patients [32]. As a result, lymph node metastasis that occurs due to greater angiogenesis in breast cancer can also be predicted using this imaging modality. DCE-MRI is a non-invasive and three-dimensional (3D) technique, not limited by breast tissue density but, it has low specificity.

2.2.6 DIFFUSION-WEIGHTED IMAGING (DWI)

Diffusion Weighted Imaging (DWI) is a form of unenhanced MRI that uses the diffusion of water molecules to generate contrast in MR images to address additional information to the classical outcomes from regular breast MRI [33]. The potential benefits of DWI techniques include improved differentiation of benign and malignant breast lesions and assessment and prediction of therapeutic efficacy [34]. DWI has enabled the identification of breast cancer particularly in dense breasts.

2.2.7 MAGNETIC RESONANCE ELASTOGRAPHY (MRE)

Magnetic Resonance Elastography (MRE) can be used to obtain useful information about tissue mechanical properties in vivo [35]. After an application of an external stress, breast MRE, a non-invasive, non-ionizing, and cross-sectional imaging modality, can quantitate the viscoelastic properties of breast tissues. Breast cancers often have a higher stiffness due to the large number of cells, collagen and proteoglycans compared to the normal surrounding tissues and benign lesions [36].

2.2.8 MAGNETIC RESONANCE SPECTROSCOPY (MRS)

Magnetic Resonance Spectroscopy (MRS) can measure a chemical spectrum in the interested region using high magnetic field strengths, typically of 11–14 T, on body fluids, cell extracts, and tissue samples, providing additional information about the chemical content in the region [37]. MRS of the breast can be used to measure the level of choline, a metabolite elevated in breast malignancies and used as a diagnostic biomarker.

2.2.9 POSITRON EMISSION TOMOGRAPHY (PET)

PET imaging has been widely adopted as an important clinical modality for oncology. Even though many types of PET radiotracers have been developed to non-invasively interrogate in vivo tumor metabolism, deoxy-fluoro-D-glucose (FDG) is the most widely used because it takes the advantage of the enhanced glucose metabolism of cancer cells [24]. In fact, cancerous cells are highly proliferative and have a higher glucose metabolism rate than normal cells. FDG PET radiotracers enter cells via the glucose transporter and are, thus, taken up in greater amounts by tumor cells than by healthy cells. Both a PET and CT scan can be done simultaneously so that CT gives a detailed picture of that area which is highly radioactive by the PET analysis [38].

Actually, Positron Emission Mammography (PEM) is a newly developed imaging method for breast cancer detection similar to PET able to detect small clusters of cancer cells.

2.2.10 SINGLE-PHOTON COMPUTED TOMOGRAPHY (SPECT)

SPECT is a nuclear medicine method that generates images as three-dimensional reconstructions of rotating planar images acquired over a 180° or 360° arc around the patient. It also determines the three-dimensional radioactivity distribution resulting from the radiopharmaceutical uptake inside the patient. Typical radiopharmaceuticals used for breast cancer imaging include ^{99m}Tc-diphosphonates, thallium chloride, ^{99m}Tc-tetrofosmin and ^{99m}Tc-methoxyisobutylisonitrile [39].

2.2.11 BREAST SPECIFIC GAMMA IMAGING (BSGI)

Breast Specific Gamma Imaging (BSGI), also known as scintimammography, is a molecular breast imaging approach and a specialized nuclear medicine imaging test that allows detection of sub-centimeter and mammographically occult breast cancer with a sensitivity and specificity comparable to MRI [24]. In BSGI, a radiotracer such as Technetium Tc99m Sestamibi is injected into the patient's bloodstream. After injection, the radiotracer eventually accumulates in the breast, where it gives off energy in the form of gamma rays. This energy is detected by a gamma camera that measures the amount of radiotracer absorbed by the body and produces detailed pictures of organs and tissues structures and functions. Unlike mammography, BSGI is unaffected by breast density [40].

2.2.12 ULTRASOUND

Ultrasound is a supplemental tool that may be utilized to analyse some breast changes in women with dense breast tissues, as well as suspicious areas not seen on a mammogram [30]. A handheld device called transducer is put on the breast's upper skin, where a gel must be applied, and moved across the breast showing the structure of the underlying tissue. Thus, the sound waves are emitted though tissues and the generated echoes must be picked up by the transducer which bounces off the body tissues. The echoes are then converted into an image (Figure 2.5) where it is possible to see the internal structure of the breast according to specific orientation and location [40]. It is a non-invasive and safe exam because the patient is not exposed to ionizing radiations, but it has several limitations: it may fail to detect microcalcifications, it may miss some early signs of cancer, it has poor resolution, it depends upon the quality of the transducer used, skills or experience of the operator that is doing the scan and on his image interpretation.

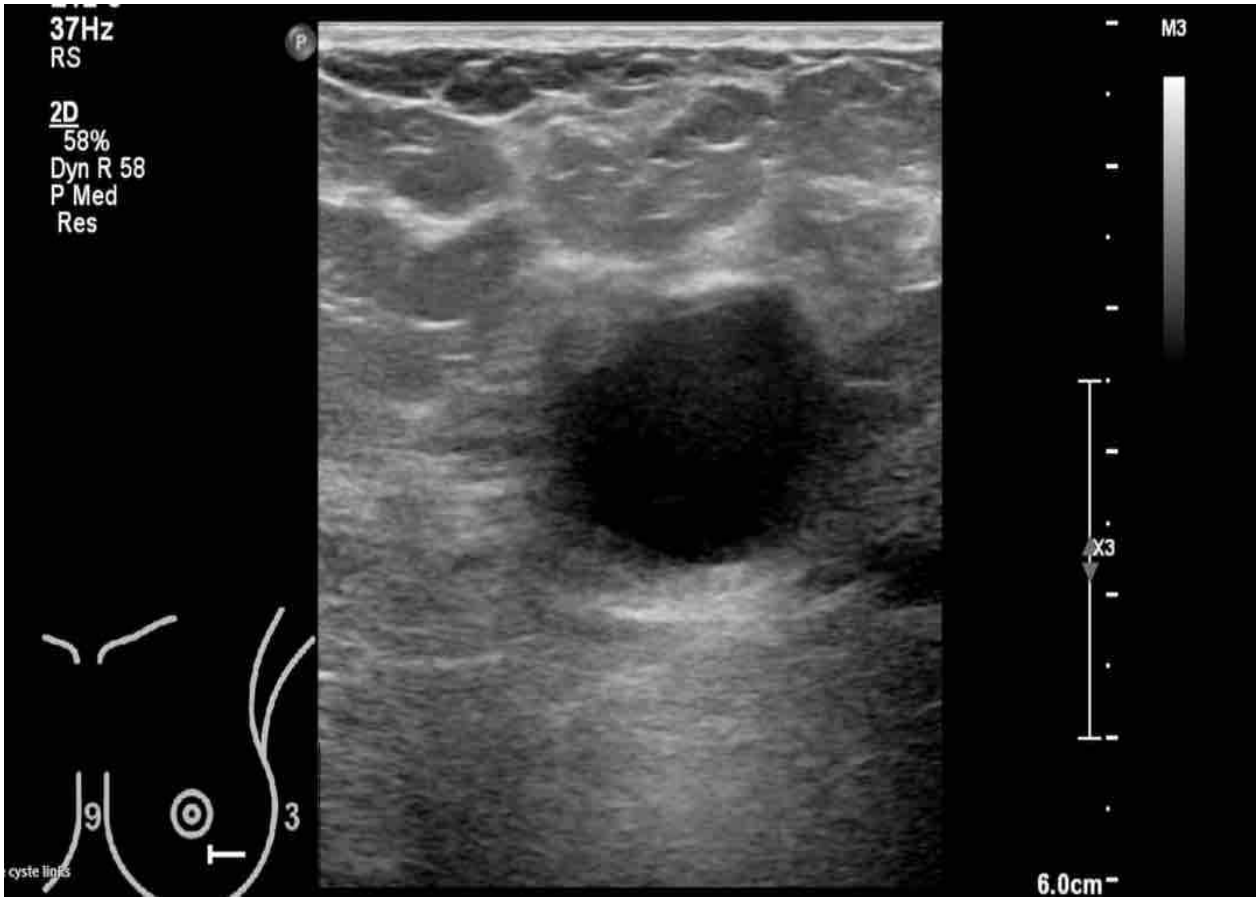

Figure 2.5: Breast ultrasound image

The characteristics of the principal imaging techniques for breast cancer detection are shown in Table I.

Table I: Principal breast cancer detection techniques

	MAMMOGRAPHY	ULTRASOUND	PET	MRI
DESCRIPTION	X-ray examination for evaluation and detection of the breast abnormalities	Sound waves are introduced into body and the echoes are used to produce the image	Based on glucose metabolism to detect cancer, a picture is obtained by the detection of radioactive areas of body from special PET scanner	Radio waves and a magnetic field are applied to alter the alignment of protons of hydrogen nuclei of cells in order to produce very detailed, cross-sectional image
ADVANTAGES	Low execution time, Good resolution, easily portable device	Wide availability, Non-invasive, lower costs, painless	Provides anatomical or functional information, good contrast	High spatial resolution, no radiation exposure, can better detect the intraductal spread of cancer
LIMITATIONS	Limited dynamic range, ionizing radiations are used, poor contrast, not good for young women	Poor resolution, poor contrast, dependent upon the skills/experience of physician	Non-portable, slow imaging time, uses ionizing radiations, expensive	Slow imaging time, non-portable, only lateral view is available, expensive, complex interpretation
SENSITIVITY	Low (women < 40 years old) High (women > 40 years old)	Low	High	Very High
SPECIFICITY	High	Moderate	Moderate	Low

2.3 BREAST CANCER DIAGNOSIS FROM MAGNETIC RESONANCE IMAGING

The role of MRI in the detection of breast cancer and the evaluation of breast lesions has become better defined in the last few years. Major improvements in MR technology and imaging techniques, coupled with the increased interest in developing ways that supplement mammography, as well as the ability to noninvasively differentiate benign from malignant lesions, resulted in a surge in

research using breast MRI for breast cancer detection and diagnosis. Its main indications are staging of known cancer, screening for breast cancer in women at increased risk, and evaluation of response to neoadjuvant chemotherapy. As opposed to mammography and US, MRI is a functional technique [41].

2.3.1 PHYSICS OF MAGNETIC RESONANCE IMAGING

Magnetic Resonance (MR) gives information on the molecular structure of the analysed tissues. It measures the absorption of electromagnetic radiation by molecules in the presence of a magnetic field. The nuclei of certain atoms such as hydrogen, phosphorus, and sodium have magnetic properties. If these atoms are exposed to an external magnetic field, their axes, which are normally randomly oriented, became aligned with the external magnetic field. Moreover, if radiofrequency (RF) energy is applied in a proper direction, it can be used to tip over and out of line the axis of these atomic magnets that start to precess. As they came back aligned with the magnetic field, they produce RF signals that can be collected by antennas and used to generate images of the tissues in which they reside. Protons can be kept precessing in a magnetic field if their precession is continually reinforced by pulses emitted a specific frequency, known as the resonance frequency [42]. This is defined by the Larmor equation, which states that the resonant frequency of the proton equals the strength of the magnetic field multiplied by the gyromagnetic constant of the proton. The gyromagnetic constant is influenced by the way the protons are organized into molecules. The protons in fat have a different gyromagnetic constant than the protons in water. This chemical difference can be used to enhance various aspects of an MR image. When a single pulse is given using the correct amount of energy, the protons can be tipped in different ways. If they are tipped of 90° respect to the direction of the main external magnetic field, this will maximize the amount of signal that they can return as they precess back into alignment. This realignment is called “free induction decay” (FID). Repetition Time (TR) is the amount of time between successive pulse sequences applied to the same slice and Time to Echo (TE) is the time between the delivery of the RF pulse and the receipt of the echo signal.

When the radiofrequency pulse is switched off, the magnetism relaxes with two characteristic relaxation times: T1 and T2. The time that it takes for the axis of the proton to return to alignment to the magnetic field is called T1 (actually T1 is 63% of the total time). Instead, if the protons are left alone, they begin to precess at different rates and, as they get out of phase, their signals begin to

cancel each other out. So, there is a point in time at which they will be completely out of phase and there will no longer be net signal from the protons. This time is measured as T2 (37% of its initial magnitude).

2.3.2 MAGNETIC RESONANCE IMAGING MODALITIES

There are different types of MR imaging modalities, each of which captures certain characteristics of the underlying anatomy. All these modalities differ in contrast and function. Three modalities of MR images are commonly referenced for clinical diagnosis: T1 (spin-lattice relaxation), T2 (spin-spin relaxation), and T2-Flair. T1-weighted images are produced by using short TE and TR times. The contrast and brightness of the image are predominately determined by T1 properties of tissue. They suppress the signal of the water and enhance the signal of different tissues as fluid (e.g. urine, cerebrospinal fluid) with low signal intensity (black), muscle with intermediate signal intensity (grey) and fat with high signal intensity (white). Thus, they are favourable for observing structures (for example, gray matter and white matter in the brain). Conversely, T2-weighted images are produced by using longer TE and TR times. In these images, the contrast and brightness are predominately determined by the T2 properties of tissue: they enhances the signal of the water and are utilized for locating tumors (Figure 2.6); Another modality is the Fluid Attenuated Inversion Recovery (Flair): The Flair sequence is similar to a T2-weighted image except that the TE and TR times are very long. T2-Flair images present the location of lesions with water suppression [43].

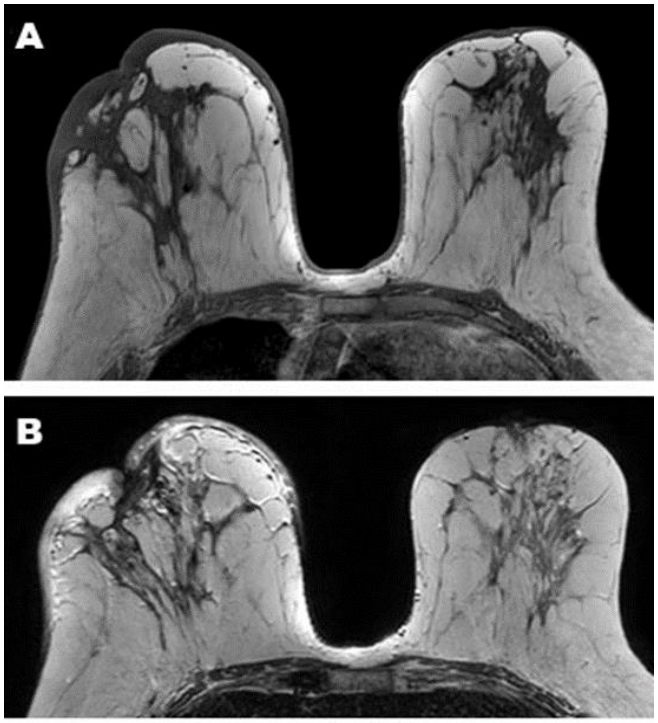

Figure 2.6: An example of breast MRI modalities: (A) T1, (B) T2

2.3.3 TISSUE CHARACTERIZATION AND LESION IDENTIFICATION IN MAGNETIC RESONANCE IMAGING

In MRI, the signal intensity depends indirectly on particular physical and chemical characteristics of the tissues being imaged. These tissue properties influence the behaviour of the nuclei undergoing resonance and their behaviour is what directly affects the MRI signal. In particular, the relaxation times, would provide very specific diagnostic information on the state of tissue within the body [44]. The T1 signal is based on the time it takes for the proton to return to its alignment to the main magnetic field. Protons in more tightly bound molecules such as fat have faster relaxation times (greater interaction between molecules), so their T1 signal is shorter than pure water. Thus, fat has the shortest T1 relaxation time. By convention, tissues with short T1 are presented as bright signals on T1 images.

Instead, in free water like breast cysts, the molecules and hydrogen nuclei are separated and do not interact rapidly with their surroundings; consequently, they lose their energy at a slower rate and get out of phase each other over a longer period of time. Thus, the water in breast cysts gives back a signal more slowly. It has a long T1 and a long T2. By convention, tissues with a long T2 are presented as bright signals on T2-weighted images. Thus, cysts (that contain fluid) with long T1 are darker on T1-weighted images and those with long T2 are brighter on T2-weighted images (Figure 2.7). Fibrous tissues, although largely made up of water, are denser and characterized by a greater intermolecular interaction, so they lose their energy more quickly, giving short T2 signals and are shown darker on T2 images.

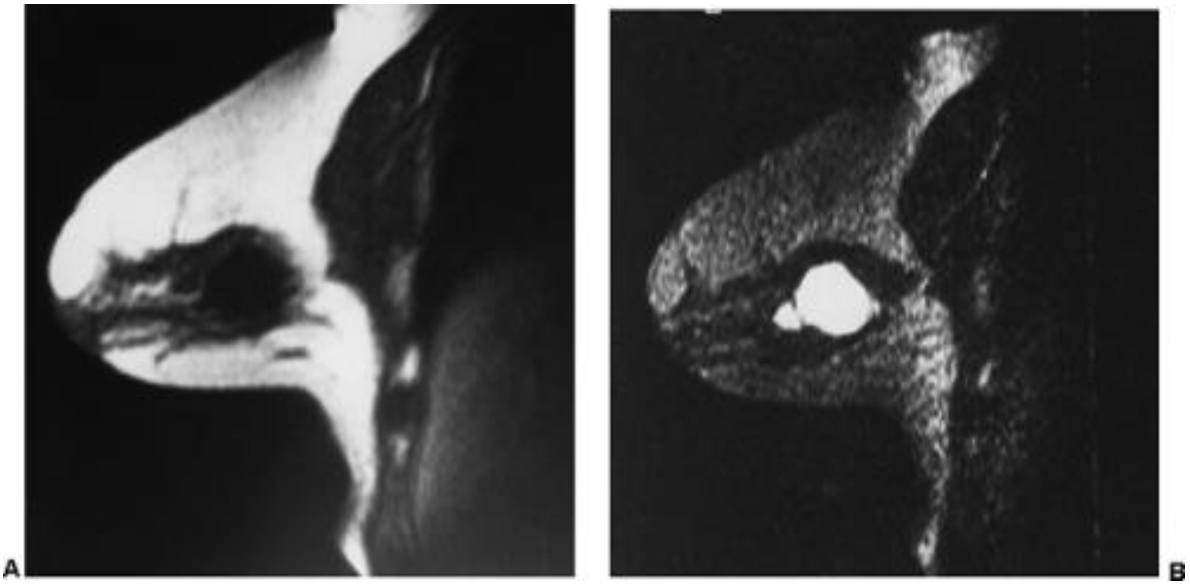

Figure 2.7: It is shown a cyst that has a characteristically low signal intensity (black) on the T1-weighted image (A) and a high signal intensity (white) on the T2-weighted image (B).

The strength of T1 and T2 signals can be used to evaluate some disease processes directly. However, in the breast it has been found that these signals are not sufficient for distinguishing malignant lesions from normal breast tissue and benign lesions. Thus, as in other organ systems, a contrast material is needed to distinguish cancers from the background breast tissue.

Breast lesion identification on MRI images depends on contrast-enhancement within the breast after intravenous injection of contrast material. As both normal breast tissue and breast lesions will enhance after contrast administration, the detection of a malignant lesion within normal breast tissue is based on the earlier and stronger enhancement on images. Differences in MR enhancement characteristics between benign and malignant breast lesions are believed to depend on differences in vascularity, vessel permeability, and extracellular diffusion space [45]. Moreover, particular emphasis is given to patient and tumor-related factors that influence image interpretation: the menstrual cycle, pevious surgery, radiation therapy and chemotherapy [46].

Lesions are categorized as foci (<5 mm of enhancement and by definition too small), masses (space-occupying lesions), and non-mass enhancement (NME), areas of enhancement without a clear space-occupying lesion. Masses are further characterized on the basis of their shape, margins and internal enhancement pattern [47]. Approximately two-thirds to three-quarters of cancers manifest as a mass, that has an irregular size and margin, heterogeneous enhancement patterns and show washout. Classic malignant areas of NME have a segmental distribution and a clumped or clustered ring pattern of internal enhancement. While most cancers are easily recognizable by their morphologic features alone, smaller lesions are more difficult to assess. In general, the features of NME are less specific than those of masses. Foci have a likelihood of malignancy of 2.9%–6% [48].

In general, lesions seen on MRI that are round, ovoid, or smoothly lobulated with sharply defined margins tend to be benign, while those that are irregular in shape or have ill-defined or spiculated margins tend to be malignant. Benign masses tend to have uniform enhancement, while malignant masses show heterogeneous enhancement and some of them exhibit brighter enhancement at their periphery, called rim enhancement (Figure 2.8).

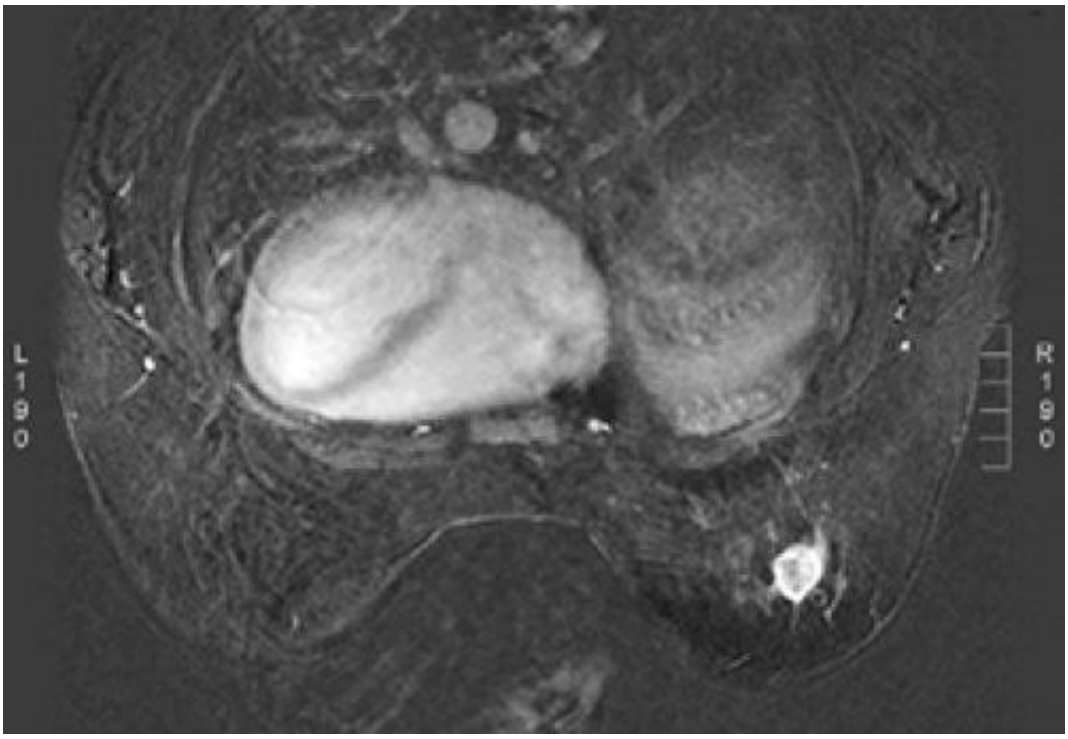

Figure 2.8: Rim enhancement. Many cancers display contrast enhancement at their periphery soon after the administration of contrast.

3. DEEP LEARNING

Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. Capabilities currently classified as AI include successfully understanding human speech, self-driving cars, intelligent routing in content delivery networks, military simulations, and interpreting complex data. In particular, Machine Learning (ML) is a branch of AI related to computers that learn from data using algorithms to perform tasks without being explicitly programmed.

Deep learning (DL), also known as deep structured learning, hierarchical learning or deep machine learning, is an AI method (Figure 3.1) that teaches computers to process data in a way that is inspired by the human brain. DL algorithms can be regarded both as a sophisticated and mathematically complex evolution of ML algorithms: they are composed of elements inspired by the structure and function of the brain called Artificial Neural Networks (ANN).

DL models can recognize complex patterns in pictures, texts, sounds, and other data to produce accurate insights and predictions. They can be used to automate tasks that typically require human intelligence, such as describing images or transcribing a sound file into text.

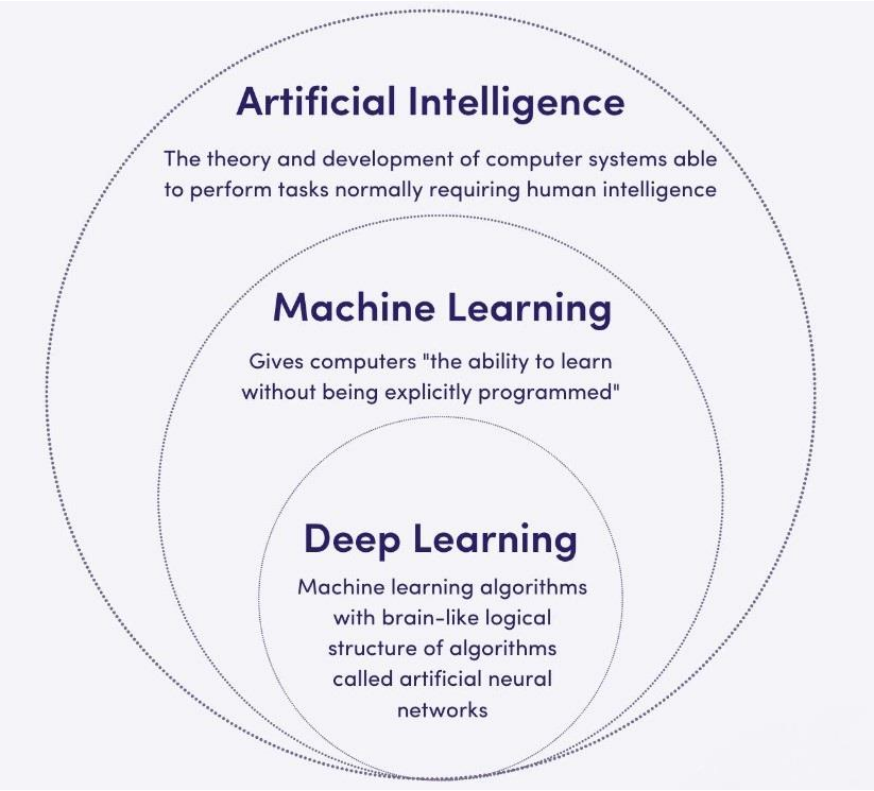

Figure 3.1: Hierarchical organization of AI and its sub-classes.

3.1 DEEP LEARNING OVERVIEW

Conventional ML techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transform the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector. The learning subsystem, often a classifier, can detect or classify patterns in the input just from the feature vector [49].

Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification. DL methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules. Each of them transforms the representation at one level, starting with the raw input, into a representation at a higher (slightly more abstract) level.

With the composition of enough such transformations, very complex functions can be learned. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations.

The key aspect of DL is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.

It has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition, it has overcome other ML techniques at predicting the activity of potential drug molecules, analysing particle accelerator data, reconstructing brain circuits, and predicting the effects of mutations in non-coding DNA on gene expression and disease [50]. Perhaps more surprisingly, DL has produced extremely promising results for various tasks in natural language understanding, classification, sentiment analysis, question answering and language translation [51].

Some of the most successful DL methods involve ANN and related ML algorithms that contain more than one hidden layer [52].

ANN are inspired by the 1959 biological model proposed by Nobel laureates David H. Hubel & Torsten Wiesel, who found two types of cells in the primary visual cortex in the human brain [53]: simple cells and complex cells (Figure 3.2).

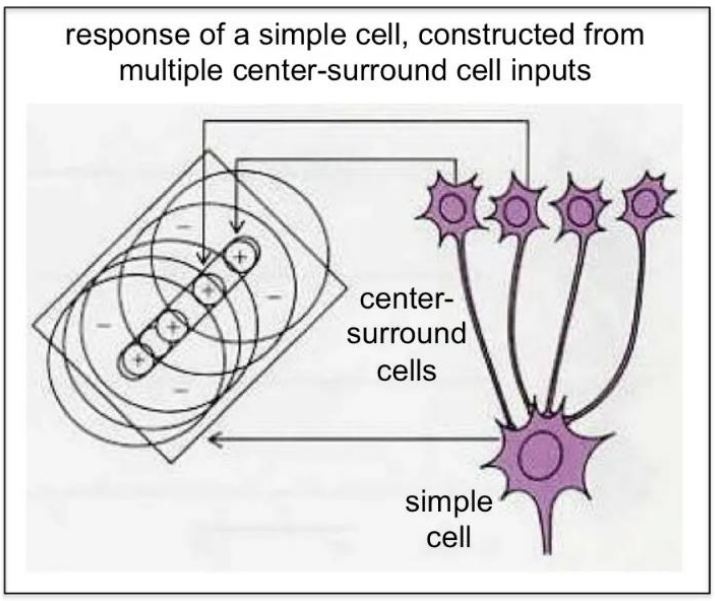

Figure 3.2: Model of cell responses proposed by David H. Hubel & Torsten Wiesel

Many artificial neural networks can be seen as cascading models [49] of cell types inspired by these biological observations.

These deep nets are part of the broader ML field of learning representations of data and use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input (Figure 3.3).

Networks are based on the learning of multiple levels of features or representations of the data: higher level features are derived from lower level features to form a hierarchical representation.

Moreover, they learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts. The neurons are the basic units of ANN and they are organized in layers. Each neuron processes the received input and send it to the neurons of the subsequent layer by means of weighted connections. The weights are adjusted during the training process to enhance the performance of the model. In a deep network, there are many layers between the input and the output, allowing the algorithm to use multiple processing layers, composed of multiple linear and non-linear transformations [54].

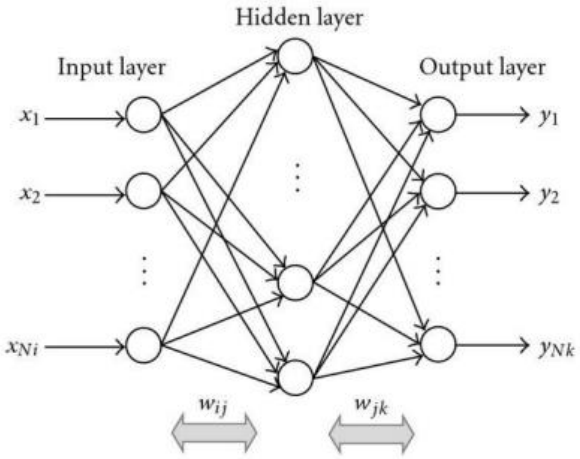

Figure 3.3: Multilayer neural network

3.2 LEARNING PROCESS

A neural network (NN) is made up of neurons connected to each other. Each connection is associated to a weight and a bias which are the parameters of the network. The weight (w_j) expresses the information being used by the net to solve a problem and the importance and the strength of the connection when multiplied by the input value. The Bias (b_j) represents specific characteristics of the neuron and its function is to shift the decision boundary. In addition, each neuron has an activation function ($a()$) that defines the output of the neuron and allows the network to learn complex patterns in data. The activation function is used to introduce non-linearity in the modelling capabilities of the network. In other words, the function is applied to all the inputs by the neurons following

$$a(w_1x_1 + \cdots + w_nx_n), \tag{1}$$

where w_1, …, w_n are the weights, and x_1, …, x_n are the inputs (Figure 3.4). The network's weights are not set by the designer of the system but, instead, are learned from examples during the training process. In fact, the model training requires labelled data consisting of the input x_i and the expected output y_i. The training procedure starts from a non-zero random set of weights and biases. This is called parameter initialization of the network. The network iterates through each input example and computes the output $\hat{y}_i$ based on the current weights. The loss function, also called cost

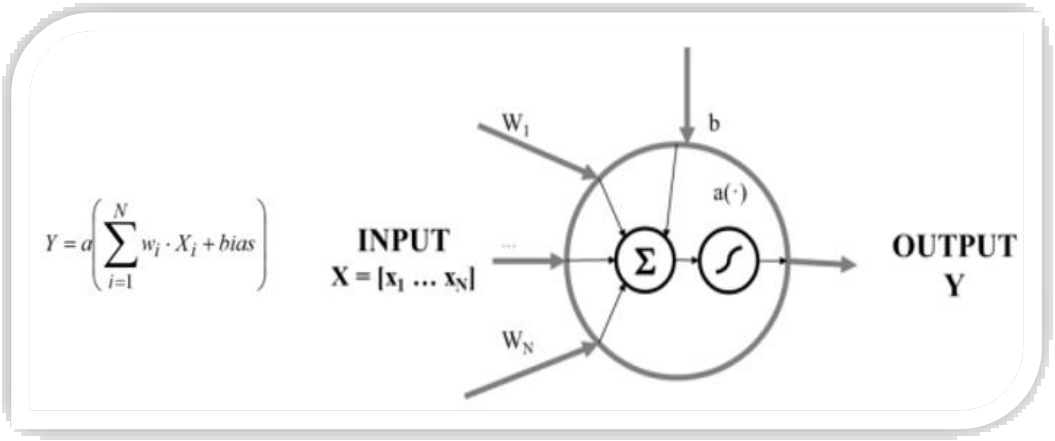

Figure 3.4: Basic structure of an artificial neuron

function, is then used to determine the error between the output of the network and the target value. There are many types of cost functions available and, all of them are utilized in optimization problems. The aim is to minimize the difference between model outputs and target values by updating the weights. So, a proper tuning of weights and biases is required to ensure lower error rates, making the model reliable by increasing its generalization capability. Among all the available methods, backpropagation is the most used one. It aims to minimize the cost function by adjusting network's weights and biases. The level of adjustment is determined by the gradients of the cost function with respect to those parameters [56].

DL can be used for supervised, semi-supervised, unsupervised as well as reinforcement machine learning:

- **Supervised Machine Learning**: It is the ML technique in which the NN learns to make predictions or classify data with a labelled dataset. Both input features with and the target variables are given in input to the network. The neural network learns to make predictions based on the cost or error that comes from the difference between the predicted and the actual target. DL algorithms like Convolutional Neural Networks (CNN), Recurrent neural Neural networks Networks (RNN) are used for many supervised tasks like image classifications and recognition, sentiment analysis, language translations, etc.
- **Unsupervised Machine Learning**: It is the ML technique in which the NN learns to discover the patterns or to cluster the dataset with unlabelled data. Here the model has to self-determine the hidden patterns or the relationships within the data. Popular techniques

include self-organizing maps, Nearest-neighbor mapping, Kmeans clustering and Singular value decomposition. Algorithms like autoencoders and generative models are used for clustering, dimensionality reduction and anomaly detection.

- **Semi-supervised learning**: It is used for the same applications as supervised learning, but it uses both labelled and unlabelled data for training. Typically, the amount of unlabelled data is higher than that of labelled data because they are less expensive and require less effort to be acquired. This type of learning can be used for many tasks such as classification, regression and prediction. Semi-supervised learning is useful when the cost associated with data labelling is too high to allow for a fully labelled training process [53].
- **Reinforcement Machine Learning**: It is the ML technique in which an agent learns to make decisions in an environment to maximize a reward signal. The agent interacts with the environment by taking action and observing the resulting rewards. The objective is to choose actions that maximize the expected reward over a given amount of time. DL can be used to learn policies, or a set of actions, that maximizes the cumulative reward over time [53].

3.3 CONVOLUTIONAL NEURAL NETWORKS

The most widely used architectures (Figure 3.5) in DL are Feedforward Neural Networks (FNNs) (Single Layer or Multilayer), CNNs and RNNs:

- FNN is the simplest type of ANN with a linear flow of information through the network.
- CNN is specific for image and video recognition tasks. These types of networks are able to automatically learn features from the images, which makes them well-suited for tasks such as image classification, object detection, and image segmentation.
- RNN is a type of NN that is able to process sequential data, such as time series and natural language. RNNs are able to maintain an internal state that captures information about the previous inputs.

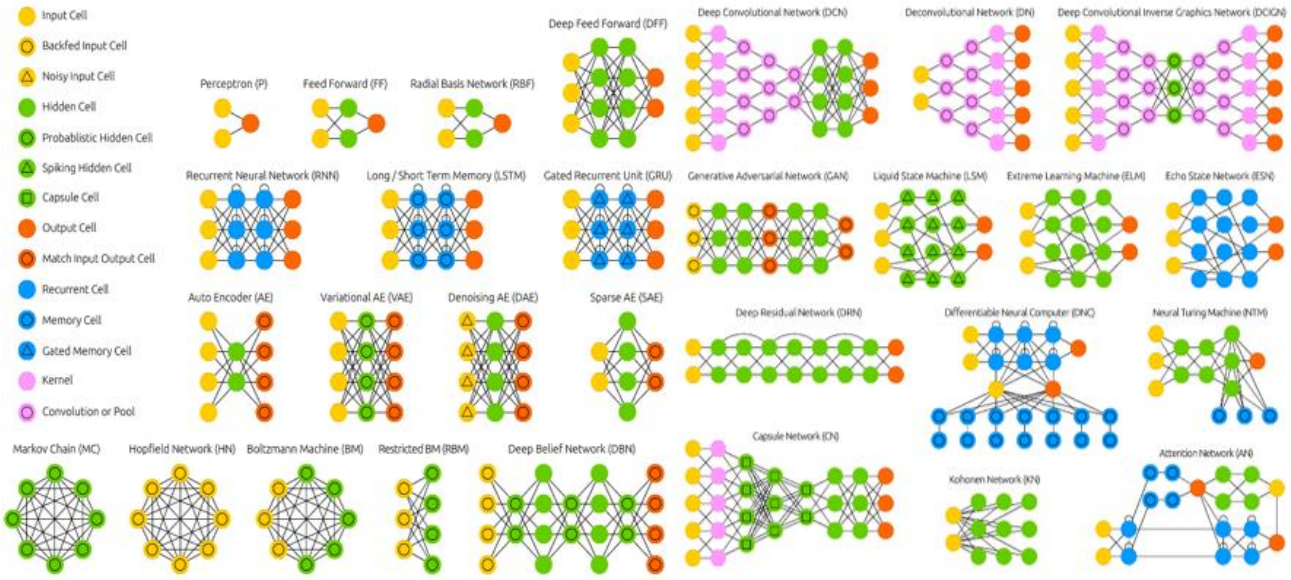

Figure 3.5: Different types of NN

CNN is a well-known DL architecture inspired by the natural visual perception mechanism of the living creatures, mostly used to analyse images. There are different types of structures of CNN but their basic components are very similar: they are mainly formed by three types of layers, namely convolutional, pooling, and fully-connected layers (Figure 3.6).

The convolutional layer extracts feature representations of the inputs with a tuneable number of filters. This layer performs convolutional operations between the image and the filters [57]. It is composed of several convolution kernels used to generate different feature maps. The operation of convolution is mathematically a dot product between two matrices:

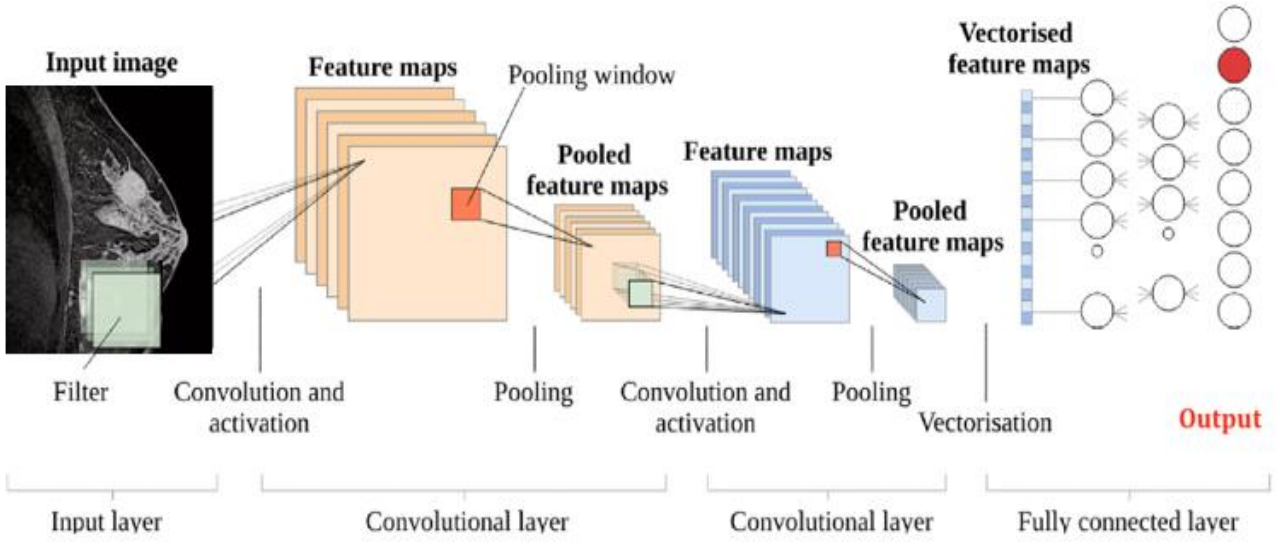

Figure 3.6: CNN Architecture

- Kernel: set of learnable parameters.
- Receptive Field: portion of analysed image whose dimension are tuneable.

For example, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels. During the forward pass, the kernel slides across the height and width of the image-producing the image representation of that receptive region. This produces a 2D representation of the image named feature map that gives the response of the kernel at each spatial position. The sliding size of the kernel is called stride [58].

A convolution operation on small regions of input is introduced to reduce the number of free parameters and to improve generalization. One major advantage of CNN is the use of shared weights in convolutional layers, which means that for each pixel it is used the same filter; this reduces memory footprint and improves performance.

The convolutional layer is to detect local conjunctions of features from the previous layer, while the role of the pooling layer is to merge semantically similar features into one (Figure 3.7).

Pooling layers are used to reduce the dimensions of the feature maps: it reduces the number of parameters to learn and the amount of computation performed in the network.

In fact, this type of layer summarises the features present in a region of the feature map generated by a convolution layer, performing feature selection. So, further operations are performed on summarised features instead of precisely positioned features generated by the convolution layer. This makes the model more robust to variations in the position of the features in the input image.

The pooling layer works by dividing the input feature map into a set of non-overlapping regions, called pooling regions. Each pooling region is then transformed into a single output value, which represents the presence of a particular feature in that region. The most common types of pooling operations are max pooling and average pooling.

In max pooling, the output value for each pooling region is simply the maximum value of the input values within that region. This has the effect of preserving the most salient features in each pooling region, while discarding less relevant information. Max pooling is often used in CNNs for object recognition tasks, as it helps to identify the most distinctive features of an object, such as its edges and corners.

In average pooling, the output value for each pooling region is the average of the input values within that region. This has the effect of preserving more information than max pooling, but may also dilute the most salient features. Average pooling is often used in CNNs for tasks such as image segmentation and object detection, where a more fine-grained representation of the input is

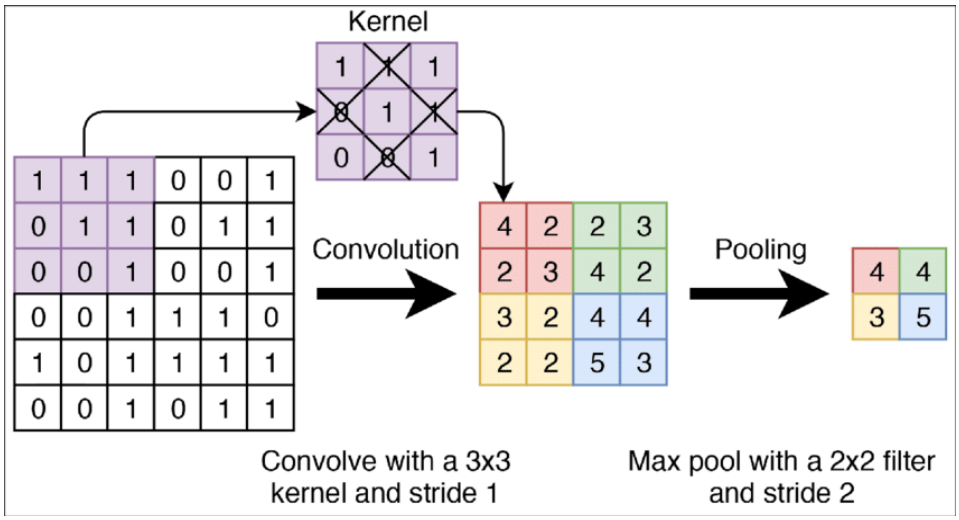

Figure 3.7: An example of schematic representation of a convolution and pooling layer in a CNN.

required. The Fully Connected (FC) layer is usually placed before the output layer and form the last few layers of CNN. The input to the fully connected layer is the flattened output from the final Pooling or Convolutional Layer. In this stage, the classification process takes place [59].

3.4 BREAST MAGNETIC RESONANCE IMAGES CLASSIFICATION FOR TUMOR DETECTION

The early detection of breast cancer is fundamental for a better prognosis. Breast MRI has been a very fast developing tool for the assessment of breast cancer and quickly moved from research to clinical settings. It offers functional methods for aid treatment response assessment that better reflect the viability of tumor and tumor burden versus just size changes. Understanding the indications for breast MRI, diagnostic criteria utilized to detect and characterize breast cancer and technical challenges are important in both clinical and research settings [60]. Utilizing additional methods to the traditional ones to increase diagnosis efficiency and reduce the false prediction rate is always necessary. In recent years, AI technology has made great progress in automatically analysing medical images for anomaly detection. In comparison with manual inspection, automated image analysis using AI reduces the time and effort needed for manual image screening and more efficiently captures valuable and relevant information from massive image collections [61]. Computer-aided methods have been widely applied to diagnose lesions on breast magnetic

resonance imaging (MRI): a key component is the selection of an appropriate classifier responsible for separating malignant and benign lesions.

Computer-Aided Diagnosis (CAD) system automates and speeds up image processing and analysis functions, and detects breast lesions by using an enhancement threshold. In particular, signal intensities of DCE MRI images are analysed by a CAD system within each voxel obtained during the dynamic sequences. The American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) MR lexicon defined a persistent kinetic pattern in MRI images as a continuous increase in signal over time. A plateau is defined as a signal intensity that does not change over time, after an initial rise. When the signal intensity decreases by more than 10%, after its highest point following its initial rise, it is defined as a washout [62]. Usually, specific colours such as blue, yellow and red are assigned to each pixel of the image of interest for the different types of tissue enhancement, such as persistent, plateau and washout, respectively (Figure 3.9). This is a very powerful tool for tumor detection, given that the system provides both morphologic and kinetic features of the breast lesion, and also provides quantitative information, such as, lesion dimensions and maximum tumor volumes [63].

Moreover, CAD for MRI is used to distinguish as non-invasive and invasive breast lesions, invasive cancers without lymph node (LN) metastasis, and invasive breast cancers with LN metastasis.

Most breast MR image analysis algorithms rely on the full-breast MRI protocol, including the temporal information from the late-phase scans and the morphological information from the early-phase scans.

According to CNN DL techniques, features that are extracted can describe multi-level tumour information from low-level (visual characteristics) to high-level (more abstract features). In a conventional image classification task, features are extracted locally using specific rules. Local texture and statistical features are universally accepted as the more important features for breast image classification. However, most current state-of-the-art techniques generally extract the features globally using kernels. In order to improve the detection efficiency of breast tumours and reduce the cost of detection, it's important to observe the breast parenchyma, fibrous gland and Background Parenchymal Enhancement (BPE) in the structured MRI. In particular, BPE is the normal manifestation of enhanced fibrous gland tissue during DCE-MRI examination. Studies have shown that the incidence of breast cancer increases with the increase of BPE. Because breast MRI examination involves images with very high resolution of soft tissue, it can clearly distinguish the breast skin, subcutaneous fat gap, normal glands and diseased tissue. In fact, these high-resolution

images enable researchers to distinguish between the different types of breast lesions. In addition, they led to identify specific features through the analysis of multiple dynamic parameter changes in the regions of interest of each mammary gland background or to extract image features based on identification of specific pixels in the segmentation. Thus, MRI images are extremely useful for diagnosis and prediction of breast tumours according to the difference in the spatial morphology of the enhanced part of the lesion [64].

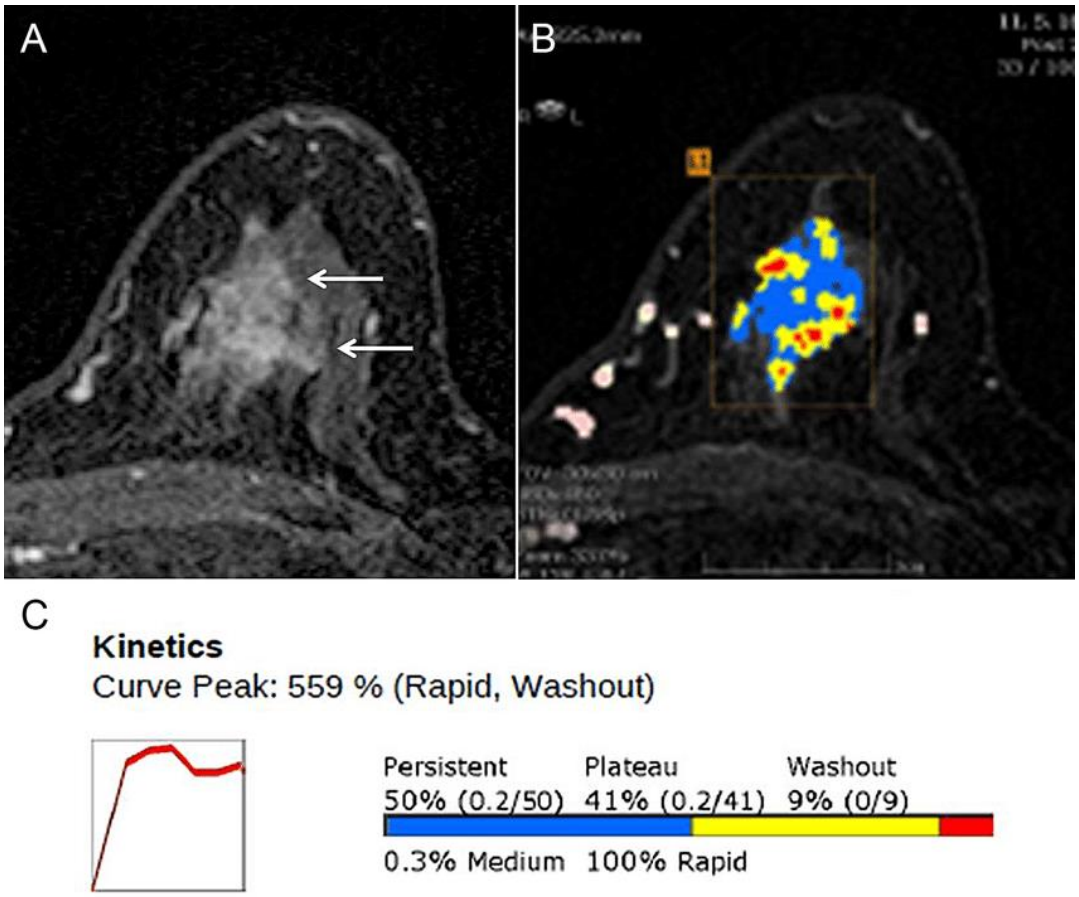

Figure 3.9: An example of MR images with computer-aided detection (CAD) showing an invasive ductal carcinoma in the left breast. (A) Axial T1-weighted MR image shows a 26-mm irregular mass (arrows). (B) Axial maximum-intensity-projection MR image shows CAD colour overlay over the breast mass. Areas in red, yellow, and blue indicate a rapid washout-type delayed enhancement, plateau-type delayed enhancement, and persistent-type delayed enhancement pattern, respectively. (C) Kinetic curve graph showing rapid initial enhancement and rapid washout-type curve.

3.4.1 VGG-16

VGG-16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition”. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes [65]. The creators of this model evaluated the networks and increased the depth using an architecture with very small 3×3 convolution filters, which showed a significant improvement on the prior configurations (Figure 3.10). The image given in input to the VGG-16 is passed through the first stack of 2 convolution layers of receptive size of 3x3, followed by Rectified linear unit (ReLU) activations. Each of these two layers contains 64 filters. The convolution stride is fixed at 1 pixel, and the padding is 1 pixel. This configuration preserves the spatial resolution, and the size of the output activation map is the same as the input image dimensions. The activation maps are then passed through spatial max pooling over a 2x2 pixel window, with a stride of 2 pixels. These halves the size of the activations. The activations then flow through a similar second stack, but with 128 filters as against 64 in the first one. This is followed by the third stack with three convolutional layers and a Max Pooling layer. The number of filters applied here are 256. This is followed by two stacks of three convolutional layers, with each containing 512 filters. The stacks of convolutional layers are followed by three fully connected layers with a flattening layer in-between. The first two have 4096 neurons each, and the last fully connected layer serves as the output layer and has 1000 neurons corresponding to the 1000 possible classes for the ImageNet dataset. The output layer is followed by the Softmax activation layer used for categorical classification.

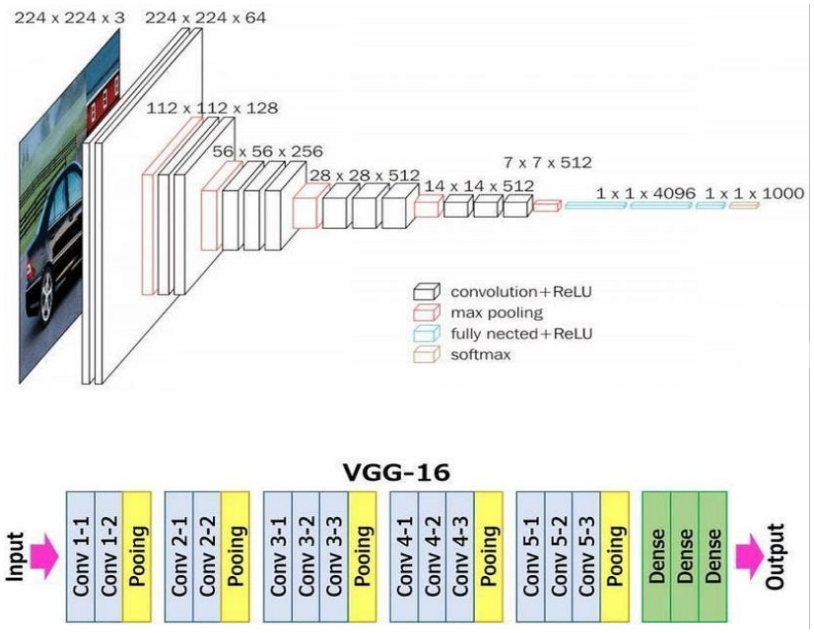

Figure 3.10: VGG-16 Network Architecture

3.4.2 VGG-19

VGG-19 is a variant of the VGG model that consists of 19 layers (16 convolution layers, 3 Fully connected layer, 5 Max Pooling layers and 1 SoftMax layer). This means that VGG-19 has three more convolutional layers than VGG-16 (Figure 3.11).

The image given in input to the VGG-16 is passed through the first stack of convolution, followed by ReLU activations.

It uses kernels of 3 x 3 size with a stride size of 1 pixel that enable them to cover the whole notion of the image and spatial padding is used to preserve the spatial resolution of the image. Moreover, max pooling is performed over a 2 x 2-pixel windows with stride of 2 and it is followed by Rectified linear unit to introduce non-linearity, to make the model classify better and to improve computational time. Lastly, it implements: three fully connected layers from which the first two are of size 4096 and after that, a layer with 1000 channels for 1000-way ILSVRC classification and the final layer is a Softmax function.

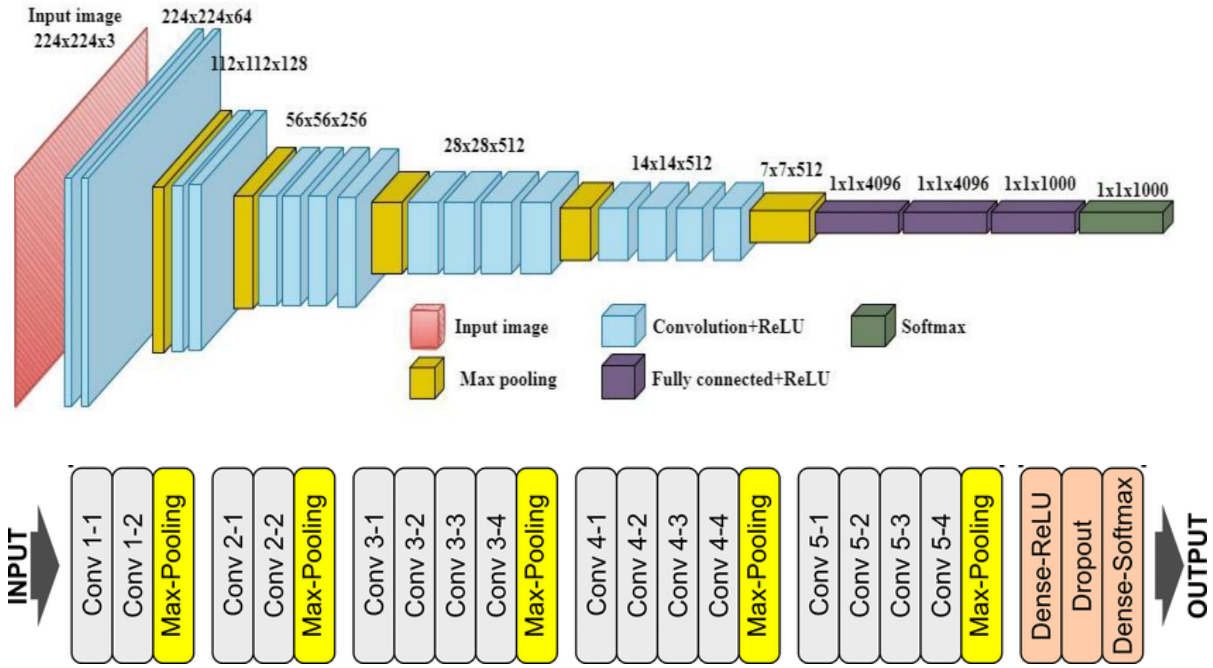

Figure 3.11: VGG-19 Network Architecture

3.4.3 RESNET

Residual Network (ResNet) architecture is a type of ANN that allows the model to skip layers without affecting performance. It deals with the problem of vanishing gradients that often occur in very deep neural networks. When a neural network has many layers, the gradients can become very small, which makes it difficult for the network to learn. ResNet solves this problem by introducing shortcut connections, also known as skip connections, that allow the network to skip over certain layers and directly propagate the input to deeper layers. This helps to preserve the gradient and make it easier for the network to learn. The maximum depth of the network has reached 152 layers, breaking through the depth of all previous designs of convolutional neural networks. When the network performance drops, it promotes the deepening of the CNN network layer. Therefore, this becomes an important module for the convolutional neural network in the process of deepening the number of network layers. The purpose of the ResNet is to increase the information transmission process: the information of the previous layer of the network is directly transmitted to the design of the next layer. Part of the information will be lost, so the shortcut added by ResNet is directly transmitted to the current layer through the previous layer network signal, which solves the process of network deepening [66].

The problem of training very deep networks has been solved by the introduction of Residual blocks (Figure 3.12) in the ResNet model. Skip connections are the base of the Residual block: the output is not the same due to this skip connection. Without the skip connection, the input x gets multiplied by the weights (w) of the layer followed by adding a bias term and, with the application of the activation function f(), we get the output as:

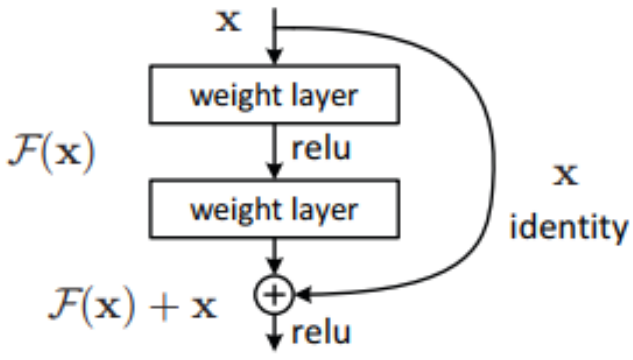

Figure 3.12: ResNet Residual block

$$H(x)=f(wx + b) \quad (2)$$

Now, with the introduction of a new skip connection technique, the output H(x) is changed to:

$$H(x)=f(x) + x \quad (3)$$

But the dimension of the input may be varying from that of the output and depends if it is working with a convolutional layer or pooling layers. Hence, this problem can be handled with two approaches:

· Zero is padded with the skip connection to increase its dimensions.

· 1×1 convolutional layers are added to the input to match the dimensions. In such a case, an additional parameter w_1 is added against the first approach and the output is [65]:

$$H(x)=f(x)+w_1 x \quad (4)$$

Regarding the structure of the network, ResNets can have variable sizes, depending on the dimension of each layer of the model and how many layers it has: ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164, ResNet-1202 etc. Every ResNet architecture performs the initial convolution and max-pooling using 7×7 and 3×3 kernel sizes respectively. Afterward, Stage 1 of the network starts and it has 3 Residual blocks containing 3 layers each. The size of kernels used to perform the convolution operation in all 3 layers of the block of Stage 1 are 64, 64 and 128 respectively. The curved arrows shown in Figure 3.13 refer to the identity connection. The dashed connected arrow represents that the convolution operation in the Residual Block is performed with stride 2, hence, the size of input will be reduced to half in terms of height and width but the channel width will be doubled. As we progress from one stage to another, the channel width is doubled and the size of the input is reduced to half. For deeper networks like ResNet50, ResNet152, etc, bottleneck design is used. For each residual function F, 3 layers are stacked one over the other. The three layers are 1x1, 3x3, 1x1 convolutions. The 1×1 convolution layers are responsible for reducing and then restoring the dimensions. The 3×3 layer is left as a bottleneck with smaller input/output dimensions. Finally, the network has an Average Pooling layer followed by a fully connected layer having 1000 neurons.

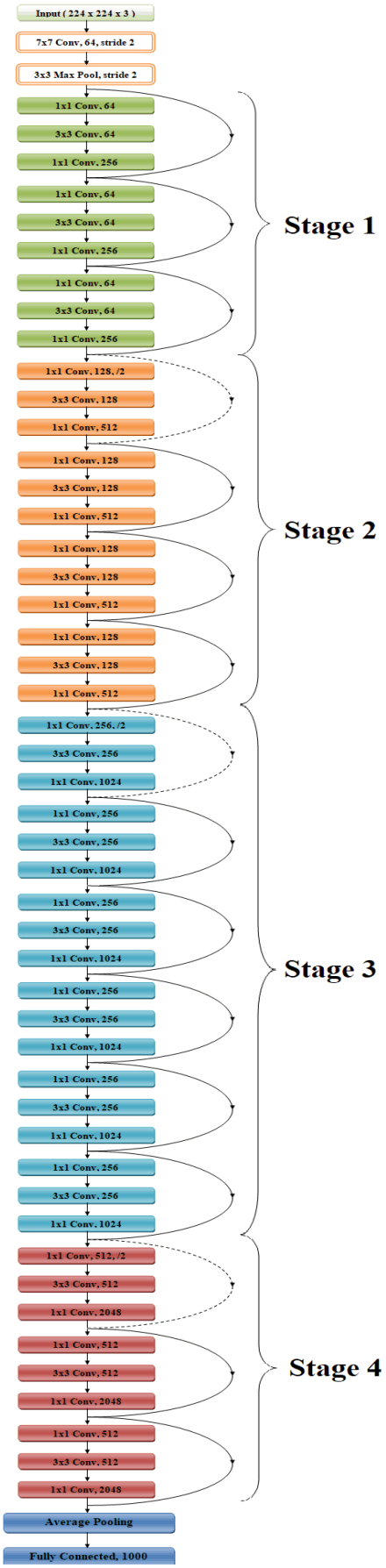

Figure 3.13: ResNet-50 network Architecture

4. EVALUATION OF THERAPEUTIC EFFICACY IN CANCER SUBJECTS THROUGH DEEP-LEARNING/MACHINE LEARNING TECHNIQUES ON RADIOLOGICAL IMAGES: A LITERATURE REVIEW

In cancer subjects that are undergoing therapy, some achieve a pathologic complete response (pCR), some achieve a partial response and others do not respond at all, or even progress. Accurate prediction of treatment response has the potential to improve patient care by improving prognostication, enabling reduction of toxic treatment that has little benefit, facilitating upfront use of novel targeted therapies and avoiding delays to surgery. However, visual inspection of a patient's tumor on radiological images is insufficient to predict the specific response to therapy [67]. ML and DL approaches, using both qualitative and quantitative features, have recently been applied to predict early treatment response in the course of, or even before, the start of any type of therapy. This is a novel field but the data published so far has shown promising results.

4.1 INTRODUCTION

The clinical response to treatment is an important indicator of the therapeutic effect of anticancer agents. Its value and interpretation have to be carefully considered within the context that it is used. In general, response assessment is combined with other indicators of the patient's condition to contribute to the decision-making process [68]. In clinical practice, the imaging evaluation of treatment response in cancer patients is based on dimensional changes of tumor lesions in sequential scans. Clinically relevant features, difficult to perceive for the human eye, can be extracted from radiological images with a novel feature transformation method called radiomics. Moreover, radiomics-based approaches have gained attention due to their high prediction power for response to chemotherapy in various types of tumors. In particular, radiological texture analysis is useful for diagnosing, staging and assessing therapy response in several cases. In order to facilitate the detection, ML and DL are increasingly investigated with the aim of improving patient diagnosis, treatment options and outcomes. DL models offer the opportunity to automatically extract imaging features to maximize model performance for the task of interest [69]. Specifically, deep neural networks enable the development of predictive models by performing all the processing steps usually involved in the design of a classic ML model, including feature extraction and learning and,

in contrast to supervised ML, deep learning CNN can operate on the whole images without requiring radiologists to manually contour the tumor on images.

Moreover, CAD systems assist doctors in the interpretation of medical images, giving a second-opinion and aiding the final diagnosis decision. Computerized diagnosis assesses the knowledge that a person or a computer has and offers an outcome to decide what kind of lesion is present and, in that case, if it is cancerous or not [60]. Medical imaging technology with applied CAD-based Machine Learning Techniques (MLTs) is becoming a powerful tool for cancer diagnosis and detection. Segmenting structures, detecting abnormalities and extracting characteristics of malignancies are some of the tasks that a standard CAD system performs. In Figure 4.1 it is reported an example of the steps required in a CAD system for breast cancer segmentation, as acquisition of the image, pre-processing, segmentation, extracting features, classification and evaluation. Firstly, the dataset needs to be prepared for the following phases, by means of a set of image pre-processing operations such as: smoothing, sharpening, noise removal and edge detection. Subsequently, the process of segmentation involves splitting an image into several areas that share common characteristics including contrast, brightness, texture, colour and grey level. Segmentation aims to perform manipulation of an image towards easier analysis and improved meaningful content by extracting Region Of Interest (ROI). In feature extraction, lesions and normal tissue that are represented by certain features are taken for evaluations. In particular, texture has traditionally been a significant diagnostic feature since its analysis is a relevant method for lesion identification and disease diagnosis [60]. Moreover, textural and geometric features' values are utilized to proceed with classification: extracted features are categorized into classes of malignant and benign disease. Finally, an algorithm will be used on suspected lesions to evaluate the classified features exploiting relevant methodologies: the goal is to identify those with a high likelihood of being correctly identified and the lowest risk of leading to diagnostic errors. The evaluation step is critical and sometimes computerized systems are required in cases where the human vision is limited and cannot distinguish a problem. As such, any evaluation algorithms for CAD systems must consider sensitivity, specificity, and evaluation of positive predictions. Thus, the aim of this systematic review is to provide an overview of the principal ML and DL methods and techniques applied on radiological images that have been proposed in literature until now for the evaluation of the therapeutic efficacy in cancer subjects.

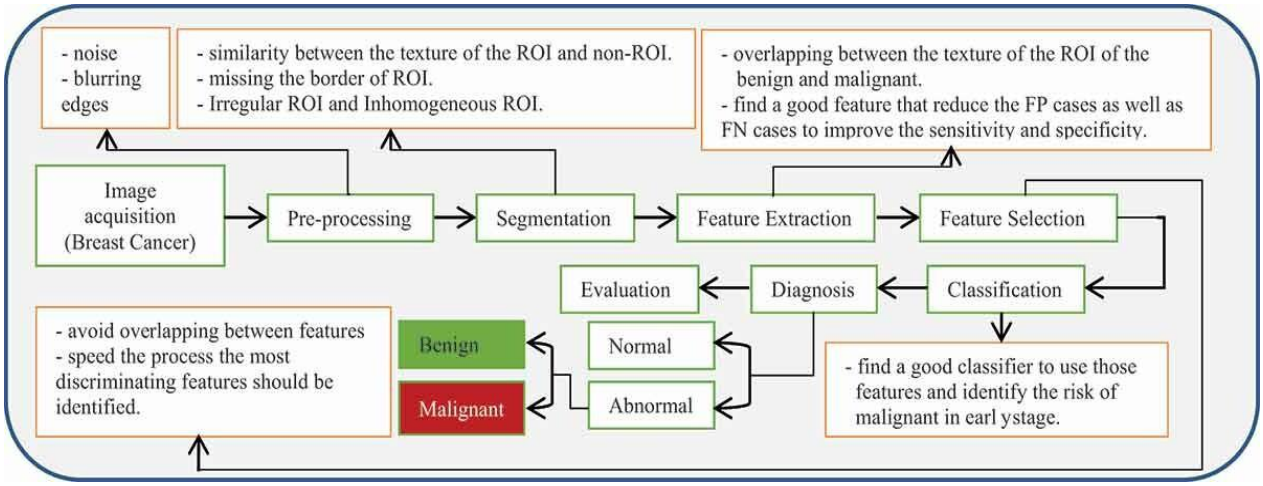

Figure 4.1: Block diagram of a CAD system

4.2 METHODS

The literature search and study selection were performed according to PRISMA guidelines for systematic reviews [70]. The literature search was conducted in PubMed, Scopus and Web of Science (WoS) during the period April 2023-May 2023.

In order to perform the search, 'Machine Learning' and 'Deep learning' and 'ML' and 'DL' terms were combined as keywords in the query with the Boolean operator 'OR'. In addition, searches for studies on tumor response detection on radiological images were performed using the terms 'cancer' and 'tumor', 'magnetic resonance' and 'computed tomography' and at last, 'therapy response' and 'treatment response' respectively, combined with the Boolean operator 'OR'. Then, the four queries were combined using the Boolean operator 'AND'. 'Title' and 'Abstract' were used to limit the search of field in PubMed, while 'Title', 'Abstract' and 'Keyword' were used to limit the search of field in Scopus and the search on Web of Science was limited only with 'Abstract'. English language, Open Access and, in order to select the most innovative proposals, Publication year greater or equal to 2015, were used as limits to filter the documents. The outcomes of the literature search were imported in Zotero reference management system to remove duplicates. Documents that were not available were excluded; documents that were reviews or databases were removed as well. Then, titles and abstracts were examined to select only documents of interest. Finally, a full-text analysis was carried out to discard documents according to these exclusion criteria: proposal of articles that not propose machine learning or deep learning techniques for prediction of therapeutic response,

proposal of articles that perform analysis through histological images or clinical features and documents that lack of evaluation of therapeutic efficacy.

4.3 RESULTS

The literature search resulted in 159 documents in PubMed, 23 in Scopus and 219 in Web of Science. 17 documents were identified as duplicates, so from 401 articles, 384 were left for the analysis (Figure 4.2). After the elimination of reviews and databases and articles that were not available, 256 documents were obtained. Moreover, analysing documents according to title and abstract, 112 were removed because out of topic, obtaining 144 articles. Finally, performing a full-text examination, 10 documents were removed because machine learning or deep learning methods

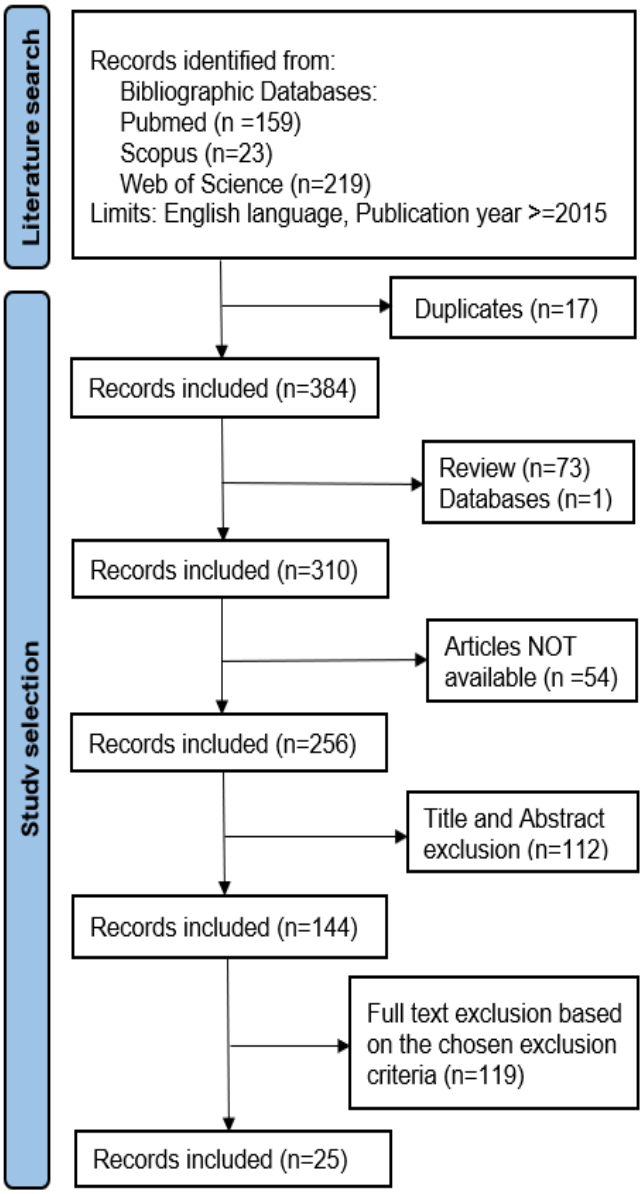

Figure 4.2: Process of literature search and study selection

were not proposed for treatment response prediction, 45 articles were eliminated because proposed studies based on histological images or clinical features and 64 documents were excluded due to the lack of the evaluation of the therapeutic efficacy. So, according to these exclusion criteria, 25 articles were finally obtained [71-95]. Figure 4.2 shows the flow-chart of the entire process of literature search and study selection and Table II summarize and gives a general overview of the methods proposed in literature for the evaluation of therapeutic efficacy in cancer subjects through deep learning/machine learning techniques on radiological images.

Table II: Overview of the methods proposed in literature

STUDY	AUTHORS	YEAR	IMAGE TYPE	METHOD	FEATURES SELECTED FOR PREDICTION	RESULTS
[71]	S. S. Alahmari, D. Cherezov, D. B. Goldgof, L. O. Hall, R. J. Gillies and M. B. Schabath	2018	CT	ML (Random Forest)	Conventional + Delta features	Risk prediction: -AUC: 0.822 -Specificity: 0.930 -Sensitivity: 0.490
[72]	Davide Bellini, Iacopo Carbone, Marco Rengo, Simone Vicini, Nicola Panvini et al	2022	MRI	ML (Decision tree)	Texture Analysis (TA): entropy, kurtosis, skewness and MPP	pCR: -Sensitivity: 0.815 -Specificity: 0.615
[73]	Cain, E.H., Saha, A., Harowicz, M.R. et al.	2019	DCE-MRI	ML (Logistic regression+ Support vector machine)	Tumor-based features and FGT enhancement	64/288 pCR,224/288 pNR: -AUC (LR): 0.707 -AUC(SVM): 0.705
[74]	Marco Caballo, Wendelien B. G. Sanderink PhD, Luyi Han MSc, Yuan Gao MSc, Alexandra Athanasiou MD, MSc, Ritse M. Mann MD	2022	MRI	ML (logistic regression model): 1.univariate analysis 2.multivariate analysis	Texture temporal variation and enhancement kinetics heterogeneity	59/251 pCR: -AUC: 0.707

[75]	Cha, K.H., Hadjiiski, L., Chan, HP. et al.	2017	CT	DL CNN, Random forest on segmented lesions (RF-SL) and Random forest on ROI (RF-ROI)	Gray-level, Texture features, morphological features and gradient field features	pCR, pNR: 1.DL-CNN: -AUC: 0.730 -Sensitivity: 0.500 -Specificity: 0.810 2.RF-SL: -AUC:0.770 -Sensitivity: 0.500 -Specificity: 0.79 3.RF-ROI: -AUC: 0.79 -Sensitivity: 0.67 -Specificity: 0.55
[76]	Chang, R., Qi, S., Wu, Y. et al.	2022	CT	DL	-	26 pCR, 27 pNR: -AUC: 0.940 -Accuracy: 0.880 -Sensitivity:0.870 -Specificity:0.860 -F1-score:0.880
[77]	Na Lae Eun, Daesung Kang, Eun Ju Son, Jeong Seon Park, Ji Hyun Youk, Jeong-Ah Kim, Hye Mi Gweon	2019	MRI	ML (Random forest)	Texture features	pCR: -AUC: 0.820 -Accuracy: 0.830 -Sensitivity: 0.620 -Specificity: 0.910 -PPV: 0.750 -NPV: 0.850
[78]	Jin, C., Yu, H., Ke, J. et al.	2021	MRI	DL	-	pCR: -AUC: 0.920 -Sensitivity: 0.900 -Specificity: 0.920 -PPV: 0.830 -NPV: 0.960
[79]	Langenhuizen, Patrick P. J. H., Zinger, Svetlana, Leenstra, Sieger, Kunst, Henricus P. M., Mulder, Jef J. S., Hanssens et al.	2020	MRI	Supervised ML (sML)	First-order statistics (FOS) features, Minkowski functionals (MFs), gray-level co-occurrence matrix (GLCM) features, gray-level size zone matrix (GLSZM) features	True tumor progression and Long-term tumor control: -AUC: 0.990 -Accuracy: 0.830 -Sensitivity: 0.830 -Specificity: 0.820
[80]	Lu, L., Dercle, L., Zhao, B. et al.	2021	CT	DL	Tumor burden	41 % pCR: -AUC: 0.760 -OS: 18 months -HR: 0.490

[81]	Ahmad Maaref, Francisco Perdigon Romero, Emmanuel Montagnon, Milena Cerny, Bich Nguyen et al.	2020	CT	DL (Deep CNN)	Textural features	pCR: -AUC: 0.880 -Accuracy: 0.760 -Sensitivity: 0.980 -Specificity: 0.540
[82]	Mehta R, Cai K, Kumar N, Knuttinen MG, Anderson TM, Lu H, Lu Y.	2017	PET/ CT	ML (multinomial naive Bayes classifier)	SUV_{max} (S), Tumor volume (V), 3D gray level co-occurrence matrix (C), WKS (W), 3D Zernike descriptor (Z)	pCR: -recall (PPV): 0.820 -precision: 0.840
[83]	Moghadas-Dastjerdi H, Rahman SH, Sannachi L, Wright FC, Gandhi S, Trudeau ME et al.	2021	CT	ML (AdaBoost-DT classifier)	Textural and second derivative textural (SDT) features	pCR: -AUC: 0.880 -Accuracy: 0.850 -Sensitivity: 0.870 -Specificity: 0.750 -precison: 0.920 -F-score: 0.890
[84]	Noémie Moreau,Caroline Rousseau,Constance Fourcade,Gianmarco Santini,Aislinn Brennan,Ludovic Ferrer,Marie Lacombe et al.	2022	PET/ CT	DL	SULpeak, TLG, PET Bone Index (PBI), PET Liver Index (PLI)	pCR: -AUC: 0.890 -Sensitivity: 0.870 -Specificity: 0.870
[85]	Nasief, H., Zheng, C., Schott, D. et al.	2019	CT	ML	Delta Radiomic Features (kurtosis, skewness, coarseness, NESTD, IDN, mean, and contrast)	pCR: -AUC: 0.960 -Accuracy: 0.940
[86]	Pang Xiaolin, Wang Fang, Zhang Qianru, Li Yan, Huang Ruiyan, Yin Xinke, Fan Xinjuan	2021	MRI	DL (SVM)	Intensity, texture and shape features	pCR: -AUC: 0.810 -Accuracy: 0.850 -Sensitivity: 0.50 -Specificity: 0.930 -F1 score: 0.650
[87]	Peng Jie, Huang Jinhua,	2021	CT	ML+DL (RF+DL)	Gray Level Variance, Large Area Low Gray	pCR: -AUC: 0.990

	Huang Guijia, Zhang Jing				Level Emphasis, Coarseness, Strength	- Sensitivity: 0.930 - Specificity: 1.000
[88]	Peng, J., Kang, S., Ning, Z. et al.	2020	CT	DL (ResNet50)	-	pCR(CR+PR): -AUC: 0.950
[89]	Peng, S.; Chen, L.; Tao, J.; Liu, J.; Zhu, W.; Liu, H.; Yang, F.	2021	MRI	ML	Zone Entropy, Gray Level NonUniformity, Small Dependence Emphasis, Correlation, Cluster Shade	44 pCR, 26 non-pCR: -AUC: 0.919 -Sensitivity: 0.880 -Specificity: 0.860 -Accuracy: 0.850
[90]	Lucie Petrova, Panagiotis Korfiatis,Ondra Petr, Daniel H., LaChance, Ian Parney, Jan C. Buckner, Bradley, J. Erickson	2019	MRI	ML (SVM)	Apparent Diffusion Coefficient (ADC) and dynamic-susceptibility contrast (DSC) perfusion measurements	23 responders, 31 non responders: -Accuracy: 0.820 -Sensitivity: 0.610 -Specificity: 0.970 OS (22 long, 32 short): -Accuracy: 0.780 -Sensitivity: 0.620 -Specificity: 0.880
[91]	Qu Hui, Zhai Huan, Zhang Shuairan, Chen Wenjuan, Zhong Hongshan, Cui Xiaoyu	2023	CT	ML (LDA)	Standard discrete (SD) feature, Discrete change (DC) feature, Relative change rate (RCR), Relative average change rate (RACR), Ploy (P) feature	Objective response (OR)= CR+PR: -AUC: 0.950 -Accuracy: 0.850
[92]	Sharaby I, Alksas A, Nashat A, Balaha HM, Shehata M, Gayhart M, Mahmoud A, Ghazal M, Khalil A. et al.	2023	CT	ML (SVM)	Shape, functionality, and Textural features	pCR(regression ≥30 %): -Accuracy: 0.950 -Sensitivity: 0.960 -Specificity: 0.940 -F1 score: 0.970
[93]	Yan, M., Wang	2020	CT	ML (SVM)	Flatness and Coefficient of variation	Tumor response scores (TRS): -TRS 0=“progressive disease” or “stable disease” -TRS 1=“partial response” or

						"complete response": -AUC: 0.910 -Accuracy:0.880 -Recall:0.880 -Precision:0.890 -F1 score:0.880
[94]	Yoo J, Lee J, Cheon M, Woo S-K, Ahn M-J, Pyo HR, Choi YS, Han JH, Choi JY	2022	PET/ CT	ML (RF)	tumor heterogeneity, fractal dimension, tumor shape and proliferation	pCR: -AUC: 0.970 -Accuracy: 0.930 -Sensitivity: 0.940 -Specificity: 0.920 -PPV: 0.940 -NPV: 0.930
[95]	Zhigang Yuan, Marissa Frazer, Anupam Rishi, Kujtim Latifi, Michal R., Tomaszewski, Eduardo G. Moros et al.	2021	CT + PET	ML (Logistic Regression)	LoG.5Skewness, LoG2CoeffVari, RL-HGRE, RL-LRHGE, SZ-LAE, and SZ-LIE	Tumor regression grade (TRG) and Neoadjuvant rectal (NAR) score: 1. TRG 0 vs. TRG 1–3: -AUC: 0.860 -Accuracy: 0.880 2. TRG 0–1 vs. TRG 2–3: -AUC: 0.950 -Accuracy: 0.910 3.low vs. intermediate vs. high NAR scores: -AUC: 0.810 -Accuracy: 0.670

4.3.1 BLADDER CANCER TREATMENT RESPONSE ASSESSMENT IN CT USING RADIOMICS WITH DEEP-LEARNING

In this study, radiomics-based predictive models using pre and post-treatment computed tomography (CT) images are analysed to verify their ability to distinguish between bladder cancers with and without complete chemotherapy responses. Three unique radiomics-based predictive models are assessed, each of which employs different fundamental design principles ranging from a pattern recognition method via deep-learning convolution neural network (DL-CNN), to a more deterministic radiomics feature-based approach and then a bridging method between the two, utilizing a system which extracts radiomics features from the image patterns. The study is performed

with the help of two experienced radiologists as references for comparison with the three predictive models. The study indicates that the computerized assessment using radiomics information from the pre and post-treatment CT of bladder cancer patients has the potential to assist in assessment of treatment response [75]. A DL-CNN is trained in order to distinguish between bladder lesions that were diagnosed as stage T0 post-treatment (no residual tumor) and those that were greater than stage T0 (any residual tumor). For predictive model, a radiomics-feature-based analysis is applied to the segmented lesions (RF-SL). A random forest classifier is trained to use these features to distinguish between lesions that fully responded to treatment and those that did not. The performance of the DL-CNN is compared against a radiomics feature-based method, where the percent change in the features extracted from the segmented lesions pre- and post-treatment is used (RF-SL), and against a third method extracting radiomics features from the paired region of interest (ROIs) used by the DL-CNN (RF-ROI). For the RF-SL, five features are consistently selected which included a contrast feature and four run length statistics texture features. For the RF-ROI, the grey-level average, the skewness of the grey-level histogram, and two run length statistics texture features are consistently selected. These results show that the texture, which characterizes the heterogeneity of the bladder lesions, is an important indicator for the estimation of full responders to chemotherapy. All three methods perform comparably in terms of AUC to the two expert radiologists. The RF-SL performs slightly better than the DL-CNN; however, the RF-ROI method results in worse performance compared to the DL-CNN, indicating that the DL-CNN is able to better characterize the paired ROIs to identify full responders compared to extracting features from the ROIs and using the random forest classifier. Table III shows the results obtained by the study.

Table III: Results of the three predictive models compared to the radiologists in terms of AUC, Complete Response (sensitivity) and Non-Complete response (specificity).

	DL-CNN	**RF-SL**	**RF-ROI**	**RADIOLOGIST 1**	**RADIOLOGIST 2**
AUC	0.730 ± 0.080	0.770 ± 0.080	0.690 ± 0.080	0.760 ± 0.080	0.770 ± 0.070
Complete Response (Sensitivity)	6/12 (50.0%)	6/12 (50.0%)	8/12 (66.7%)	11/12 (91.7%)	11/12 (91.7%)
Non-Complete response (Specificity)	34/42 (81.0%)	33/42 (78.6%)	23/42 (54.8%)	18/42 (42.9%)	16/42 (38.1%)

4.3.2 DEEP MULTIPLE INSTANCE LEARNING FOR PREDICTING CHEMOTHERAPY RESPONSE IN NON-SMALL CELL LUNG CANCER USING PRETREATMENT CT IMAGES

In this study, a deep multiple instance learning (DMIL) model is proposed to predict the chemotherapeutic response in non-small cell lung cancer (NSCLC) patients using pre-treatment CT images [76]. Two datasets of NSCLC patients treated with chemotherapy as the first-line treatment are collected from two hospitals. Dataset 1 (163 response and 138 nonresponse) is used to train, validate, and test the DMIL model and dataset 2 (22 response and 20 nonresponse) is used as the external validation cohort. The proposed DMIL model consists of three main modules: Pre-processing module, Feature extraction module and Feature representation module.

Five different pre-trained models are trained in the Feature extraction Module (AlexNet, VGG16, ResNet34, DenseNet, and MobileNet_v2) as the backbone module to extract features from CT slices. VGG16 shows better predictive performance than the other models. By comparing CT images before and after chemotherapy, all cases are categorized as “response” or “nonresponse.” The response group includes CR (complete response) and PR (partial response), whereas the nonresponse group includes PD (progressive disease) and SD (stable disease).

Results shows that the model correctly predicts 27 of 30 nonresponse and 26 of 30 response patients in the test cohort. Performance measures as accuracy, AUC, sensitivity, specificity, and F1-score give values of 0.883, 0.982, 0.871, 0.867, and 0.885, respectively, whereas the cut-off value was 0.833

4.3.3 PREDICTING TREATMENT RESPONSE FROM LONGITUDINAL IMAGES USING MULTI-TASK DEEP LEARNING

In this study, a multi-task deep learning approach is proposed to predict treatment response and test the model in multi-institution cohorts (internal and external) of rectal cancer patients. The multi-task deep learning network consists of two subnetworks (Figure 4.3): one for feature extraction and tumor segmentation, and one for response prediction [78]. The feature extraction and segmentation subnetwork consists of two identical 3D U-net with shared parameters. The response prediction subnetwork combines the extracted image features from three different network layers via depth-wise convolution.

For response prediction, the proposed 3D RP-Net achieves consistently high accuracy across the training and two validation cohorts. The AUC is 0.95 (95% CI: 0.91–0.98) and 0.92 (95% CI: 0.87–0.96) in the internal and external validation cohorts, respectively. At the optimal cut-off point, the 3D RP-Net shows sensitivity at 93% and 91%, specificity at 94% and 92% for predicting pCR for the two validation cohorts.

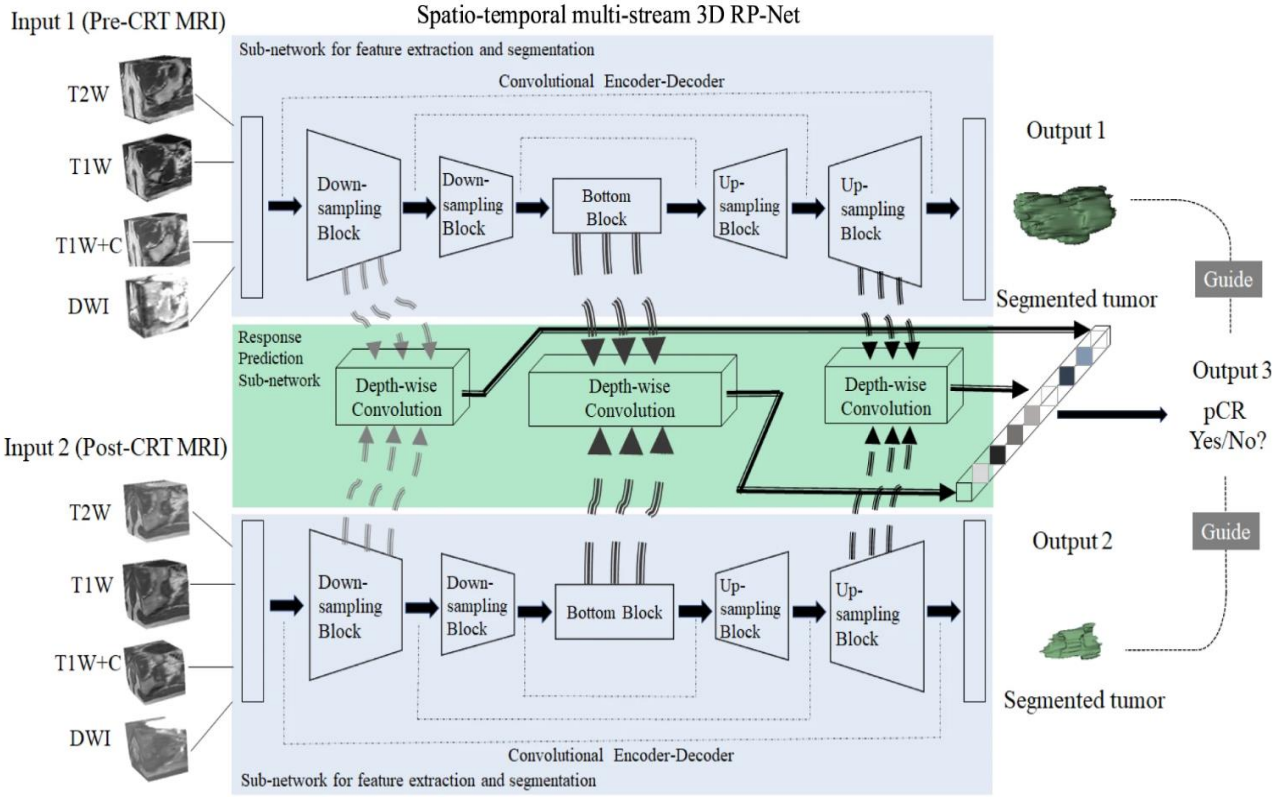

Figure 4.3: Schematic representation of the multi-task deep learning network [78]

4.3.4 DEEP LEARNING FOR THE PREDICTION OF EARLY ON-TREATMENT RESPONSE IN METASTATIC COLORECTAL CANCER FROM SERIAL MEDICAL IMAGING

The DL method used in this study deploys two types of networks, the CNN and the RNN. CNN is basically the type of network that is specified for image analysis and computer vision, while RNN is designed for doing prediction based on temporal sequence. RNN allows automated learning of time-dependent relation between features rather than using human curate modelling (e.g. the tumor growth inhibition modelling).

For automated response assessment, CNN is utilized to extract image features from CT scan at each time point, while RNN is utilized to build up a time-dependent network from the image features extracted by CNN at the series of time points.

The validation of prognostic performance of DL prediction score consists of four steps: First, reproducibility analysis; Second, classification of patients into high- or low-risk groups is determined according to the Youden Index on the receiver operator characteristic curve (ROC); Third, in the test set, Kaplan-Meier survival analysis is performed to assess the association between DL-based stratification (using the cut-off determined in the tuning set) and patient's Overall Survival (OS). Finally, the DL prediction score was compared to the Response evaluation criteria in solid tumours (RECIST) and early tumour shrinkage (ETS) criteria [80].

Results of the study shows that the percentage of responders defined by DL criteria vs. RECIST criteria is 41% vs. 6.7%. The AUC (95% CI) of DL prediction score on the tuning set to classify DL-responder/ non-responder is 0.76 (95% CI: 0.72,0.80). The DL-responders has a significant better OS than the DL-non-responders, with median OS 18.0 vs.10.4 months, hazard ratio (HR) (95% CI) = 0.49 (0.40,0.61).

4.3.5 PREDICTING THE RESPONSE TO FOLFOX-BASED CHEMOTHERAPY REGIMEN FROM UNTREATED LIVER METASTASES ON BASELINE CT: A DEEP NEURAL NETWORK APPROACH

In this paper, a fully automated framework based on deep convolutional neural networks (DCNN) is proposed to differentiate treated and untreated lesions and to identify new lesions appearing on CT scans. Moreover, a fully connected neural network is presented to predict the response of patients with colorectal liver metastases (CLM) undergoing FOLFOX with Bevacizumab chemotherapy, from untreated lesions in pre-treatment computed tomography (CT).

The CNN with fivefold cross-validation is trained for prediction purposes and the layer architecture that has been used in order to perform classification includes four convolutional 2D layers, followed by a rectified linear unit (ReLU), two 2D Max Pooling layers, two drop out, and a flatten layer. The following four experiments are performed: bounding box on the largest untreated lesion per patient, segmentation of the largest untreated lesion per patient, bounding box on all untreated lesions per patient and segmentation of all untreated lesions per patient [81].

The proposed deep neural network based on Inception-Inspired-CNN shows high classification performance discriminating treated vs untreated lesions in CT with an AUC of 0.97, sensitivity of 90% (95% confidence interval, 86–93), a specificity of 91% (95% CI, 85–94), and an overall accuracy

of 0.91 (95% CI, 88–93). In order to evaluate response prediction, all the four different scenarios are analysed. The first experiment based on the bounding box over the largest untreated lesion achieves an AUC of 0.66 (95% CI, 0.61–0.75), accuracy of 61% (95% CI, 54–73), sensitivity of 59% (95% CI, 53–68), and a specificity of 65% (95% CI, 62–77). The second experiment based on the segmentation of the largest untreated lesion achieves AUC of 0.68 (95% CI, 0.63–0.78), accuracy of 66% (95% CI, 62–75), sensitivity of 69% (95% CI, 62–80), and specificity of 60% (95% CI, 54–71). For the third experiment, bounding boxes on all lesions yield an AUC of 0.83 (95% CI, 0.78–0.87), accuracy of 78% (95% CI, 74–83), sensitivity of 97% (95% CI, 94–99), and specificity of 59% (95% CI, 52–68). Finally, an AUC of 0.88 (95% CI, 0.85–0.94), accuracy of 76% (95% CI, 71–82), sensitivity of 98% (95% CI, 96–99), and specificity of 54% (95% CI, 50–60) are obtained with the fourth configuration, based on segmentations of all lesions. Results demonstrate improved performance with the third and fourth experiments where all lesions have been applied for training the CNN. From the 34 poor responders in the test set, the proposed model is able to predict response correctly in 27 cases. In conclusion, this study shows that fully connected neural networks are more efficient in predicting highly heterogeneous CLM treated with TACE compared with traditional classification methods based on hand-crafted texture feature analysis methods for conventional machine learning.

4.3.6 AUTOMATIC SEGMENTATION OF METASTATIC BREAST CANCER LESIONS ON 18F-FDG PET/CT LONGITUDINAL ACQUISITIONS FOR TREATMENT RESPONSE ASSESSMENT

The aim of this work is to propose deep learning neural networks used to segment breast cancer metastatic lesions on longitudinal whole-body PET/CT and extract imaging biomarkers from the segmentations to evaluate their potential to determine treatment response. Baseline and follow-up PET/CT images of 60 patients are used to train two deep-learning models to segment breast cancer metastatic lesions: One applied on baseline images and one on follow-up images. From the automatic segmentations, four imaging biomarkers were computed and evaluated: SULpeak, Total Lesion Glycolysis (TLG), PET Bone Index (PBI) and PET Liver Index (PLI) [84].

SULpeak is used to assess metabolic change between two acquisitions, Total Lesion Glycolysis (TLG) assesses both metabolic activity and lesion volume, PET Bone Index (PBI) indicates bone metastases and PET Liver Index (PLI) individuates liver metastases.

To evaluate the potential of the biomarkers to assess treatment response, changes between the baseline and follow-up images are analysed with each imaging biomarker and in order to determine the best biomarker to assess treatment response, a Receiver Operating Characteristic (ROC) curve

and its Area Under the Curve (AUC) are computed. PERCIST responses assessed by medical experts are binarized as responders for subjects with CR and PR and non-responders for subjects with SD or PD.

The ROC analysis evaluates each difference measured between baseline and follow-up acquisitions as a potential threshold for a binary prediction of treatment response.

Results shows that the highest AUC score is obtained for ΔSULpeak at 0.89, followed by ΔTLG at 0.80, ΔPBI at 0.72 and ΔPLI at 0.54. The AUCs for ΔSULpeak, ΔTLG and ΔPBI are not statistically different (p-values ≥ 0.001) but ΔPLI has significantly lower predictive value than the other biomarkers (p-value ≤ 0.001). The optimal cutoff values to classify patients as responders or non-responders are−32%, −43%, −8% and 0% for ΔSULpeak, ΔTLG, ΔPBI and ΔPLI respectively. According to the Mann-Whitney U test, the responder/non-responder groups defined by each biomarker are statistically different (p-value ≤ 0.001) except for the ΔPLI (p-value = 0.062).

4.3.7 A PIPELINE FOR PREDICTING THE TREATMENT RESPONSE OF NEOADJUVANT CHEMORADIOTHERAPY FOR LOCALLY ADVANCED RECTAL CANCER USING SINGLE MRI MODALITY: COMBINING DEEP SEGMENTATION NETWORK AND RADIOMICS ANALYSIS BASED ON "SUSPICIOUS REGION"

This study proposes a pipeline of pCR prediction using a combination of DL and radiomics analysis. In order to improve the efficiency for clinical application, the pipeline only included a post-nCRT T2-weighted MRI [86]. Unlike other studies that attempted to carefully find the ROI using a pre-nCRT MRI as a reference, in this case it is considered the ROI on a "suspicious region", which is a continuous area that has a high possibility to contain a tumor or fibrosis as assessed by radiologists. A deep segmentation network, termed the two-stage rectum-aware U-Net (tsraU-Net), is designed to segment the ROI to substitute for a time-consuming manual delineation of the rectum region. This is followed by a radiomics analysis model based on the ROI to extract the hidden information and predict the pCR status. Patients dataset is separated into three groups: Dataset Seg-T consisted of patients to train deep networks for ROI segmentation; dataset Rad-T consisted of patients to build the radiomics model for predicting the pCR status; and dataset In-V for internal validation. In addition, dataset Ex-V is used as an external validation set. The "suspicious region" is defined as a continuous region containing 129 abnormal intensity signals compared to a normal rectal wall, which are highly suspected to be cancer or fibrosis according to clinical experience. The following

analysis extracts a great number of radiomics features, including texture, first-order statistics, and shape, on the ROI and its wavelet decompositions to represent certain properties. Statistical techniques are later applied to select the most representative features and construct a final model to predict the pCR status. The support vector machine (SVM) is applied to predict the pCR status and for the pCR status prediction evaluation, five metrics are applied: the area under receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, specificity, and at least the F-score, that is a weighted harmonic mean that comprehensively considers sensitivity and specificity. Table IV shows the results of the study: the "suspicious region" is capable to be the ROI in this research, which means a single post-nCRT T2-w MRI has the ability to predict the pCR status without the help of a pre-nCRT MRI or other post-nCRT modalities.

Table IV: Results of the pCR status predicted performance on datasets Rad-T, In-V and Ex-V, in terms of AUC, accuracy, sensitivity, specificity, F0.5-score, F1-score, F1.5-score

DATASET	AUC	ACCURACY	SENSITIVITY	SPECIFICITY	$F_{0.5}$-SCORE	F_1-SCORE	$F_{1.5}$-SCORE
Rad-T	0.924	0.860	0.861	0.859	0.860	0.860	0.860
In-V	0.829	0.804	0.750	0.816	0.802	0.792	0.769
Ex-V	0.815	0.853	0.500	0.929	0.793	0.650	0.583

4.3.8 PREDICTING THE INITIAL TREATMENT RESPONSE TO TRANSARTERIAL CHEMOEMBOLIZATION IN INTERMEDIATE-STAGE HEPATOCELLULAR CARCINOMA BY THE INTEGRATION OF RADIOMICS AND DEEP LEARNING

In this study, it is aimed to develop radiology-based models for the preoperative prediction of the initial treatment response to transarterial chemoembolization (TACE) in patients with hepatocellular carcinoma (HCC) with the integration of radiomics and DL. Five radiomics ML models (linear, logistic, GBM, SVM, and RF) and a DL model are developed to precisely predict the initial treatment response to TACE in the training and validation cohorts. The deep learning framework includes two convolutions, two max-poolings and one dense layer. The final output layer was a Softmax classifier [87]. According to the modified Response Evaluation Criteria in Solid Tumors, the initial response to TACE is classified as CR, PR, SD, or PD. Initial treatment response and non-response are strictly defined as CR + PR and SD + PD after the first course of TACE therapy. The

performance of each radiomics model is evaluated in the training and validation cohorts using receiver operating characteristic (ROC) analysis and the optimal cutoff value for predicting treatment response is defined using the Youden index. A total of 1167 features are extracted from the hepatic-arterial 3D-CT images. A total of 457 pyradiomics features are eliminated in the Interclass Correlation Coefficient (ICC) analysis. To acquire robust features, the remaining 710 features are subjected to feature selection using the Recursive Feature Elimination (RFE) algorithm. Based on 5-fold cross-validation, 14 radiomics features are finally selected and used to build the five ML models. All machine learning models have significantly high prediction in the training and validation cohorts (each $P < 0.001$). The simple linear model shows the lowest accuracy in the two cohorts (AUC = 0.784; 95% CI: 0.707–0.860, $P < 0.001$ vs AUC = 0.763, 95% CI: 0.693–0.833, $P < 0.001$). The logistic model is superior to the linear model in the training and validation groups (AUC = 0.801 vs 0.784 and 0.781 vs 0.763, respectively). For the three nonlinear models, RF shows better predictive accuracy than SVM and GBM in the two cohorts (AUC = 0.967 vs 0.841 and 0.839; and 0.964 vs 0.765 and 0.810, respectively). The three nonlinear models (SVM, GBM, and RF) have better predictive ability than the two linear models (linear and logistic). In the above engineered features analysis, the DL model demonstrates high accuracy in the training and validation cohorts (AUC = 0.981, 95% CI: 0.964–998, $P < 0.001$ vs AUC = 0.972, 95% CI: 0.951–0.993, $P < 0.001$).

4.3.9 RESNET FOR PREDICTING RESPONSE OF TRANSARTERIAL CHEMOEMBOLIZATION IN HEPATOCELLULAR CARCINOMA FROM CT IMAGING

The aim of this study is to train and validate (in different cohorts) a DL model on CT images for the preoperative prediction of the response of patients' cohorts with intermediate-stage hepatocellular carcinoma (HCC) undergoing TACE [88]. It is presented a predictive model from the outputs using the transfer learning techniques of a residual convolutional neural network (ResNet50). A series of blocks consisting of three convolutional layers (fc1000, fc1000_softmax, and classification layers_fc1000) are replaced by new layers (fc4, fc4_softmax, and classification layers_fc4) to extract deep residual features and transmit features from the front layer to the latter one. At the end of the network, a full-connection layer is used to perform classification. Based on the radiology evaluation in patients after the first TACE therapy, the different responses on hepatic-arterial CT images are determined by modified RECIST, including CR, PR, SD, and PD. The objective response is defined as CR+PR and the non-response as SD+PD. The training cohort exhibits an average accuracy

of 84.0% and a low error of 16% of predicting CR, PR, SD, and PD in TACE therapy. The independent validation cohorts 1 and 2 show an average accuracy of 85.1% and 82.8% and low errors of 14.9% and 17.2%, respectively. Interestingly, it is found that the accuracy for the training cohort was lower than for the validation cohort (84.3% vs. 85.1%). Misclassified CR patches by the deep learning model are more observed in PR patches than in SD and PD patches in the training cohort (1.5%) and validation cohorts 1 (1.7%) and 2 (1.1%). Meanwhile, misclassified PD patches are more frequently found in PR patches than in SD and CR patches. The precision probability of preoperatively predicting the four therapy responses (i.e., CR, PR, SD, and PD) via each ROI patch is calculated and found useful in individualized clinical treatment.

5. TUMOR-INFILTRATING LYMPHOCYTES EVALUATION ON BREAST MRI: A REAL DEEP LEARNING APPLICATION

Recent evidence has suggested that TILs have prognostic and predictive capabilities for breast cancers. TIL assessments are clinically useful for risk predictions, adjuvant and neoadjuvant chemotherapy decisions, and more recently, immunotherapy. TIL evaluations are being included in clinical trials and diagnostic assessments, which has raised concerns regarding the existence of a standardized methodology for its evaluation [96]. Therefore, recommendations and guidelines for visual TIL assessment (VTA) in invasive breast carcinoma patients have been recently developed by the International Immuno-Oncology Biomarker Working Group (or TILs-WG) [97]. TIL populations are quantified by determining how much of a demarcated area of stroma or tumor visible on a slide is infiltrated by immune cells (average TIL%).

In this study it is proposed the use of three different DL models for the assessment of TIL from MRI breast cancer images.

5.1 DATASET

The Dataset TCGA-BH has been acquired from the Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) data collection. It is part of a larger effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA). Clinical, genetic, and pathological data resides in the Genomic Data Commons (GDC) Data Portal while the radiological data is stored on The Cancer Imaging Archive (TCIA). Matched TCGA patient identifiers allow researchers to explore the TCGA/TCIA databases for correlations between tissue genotype, radiological phenotype and patient outcomes. Tissues for TCGA were collected from many sites all over the world in order to reach their actual targets, usually around 500 specimens per cancer type. For this reason, the image data sets are also extremely heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. In most cases the images were acquired as part of routine care and not as part of a controlled research study or clinical trial.

5.1.1 RADIOLOGICAL DATA FROM THE CANCER IMAGING ARCHIVE

TCIA is a service which de-identifies and hosts a large publicly available archive of medical images of cancer. It is funded by the Cancer Imaging Program (CIP), a part of the United States National Cancer Institute (NCI) and is managed by the Frederick National Laboratory for Cancer Research (FNLCR). The imaging data are organized as "collections" defined by a common disease (e.g. lung cancer, breast cancer), image modality or type (MRI, CT, digital histopathology, etc) or research focus. Digital Imaging and Communications in Medicine (DICOM), a worldwide standard for the storage and transmission of medical imaging, is the primary file format used by TCIA for radiology imaging; A DICOM file consists of a header and image data sets packed into a single file. The information within the header is organized as a constant and standardized series of tags. By extracting data from these tags one can access important information regarding the patient demographics, study parameters, etc. An emphasis is made to provide supporting data related to the images such as patient outcomes, treatment details, genomics and expert analyses. In this study, the NBIA Data Retriever was used to download from TCIA the radiological images that have the following characteristics (Table V):

Table V: Detailed description of radiological images used in the study

IMAGE STATISTICS	RADIOLOGY IMAGING STATISTICS
Modalities	MR, MG
Number of Participants	139
Number of Studies	164
Number of Series	1877
Number of Images	230167
Images Size (GB)	88.1

5.1.2 TUMOR INFILTRATING LYMPHOCYTE VALUE FROM GENOMIC DATA COMMONS DATA PORTAL

GDC Data Portal is an interactive data system for researchers to search, download, upload, and analyse harmonized cancer genomic data sets, including TCGA (The Cancer Genome Atlas) and Therapeutically Applicable Research to Generate Effective Therapies (TARGET). It contains information about projects, primary sites, cases, files, genes and mutations. This database allows to browse using Projects, Exploration and analysis.

The Exploration Page allows users to explore data in the GDC using advanced filters, which includes those on a gene and mutation level. Users can choose filters on specific Cases, Genes, and/or Mutations and it's possible to visualize these results. The Gene or Mutation data for these visualizations comes from the Open-Access MAF files on the GDC Data Portal. In this study, the TIL percentages for each patient have been obtained from the exploration page in the slide details section (Figure 5.1) where the user can visualize the histological image features. There is also a Clinical tab with filters that apply specifically to clinical data: The Pathology Report of each case has been downloaded in order to assess the tumor characteristics and to identify the cancer position (left breast, right breast).

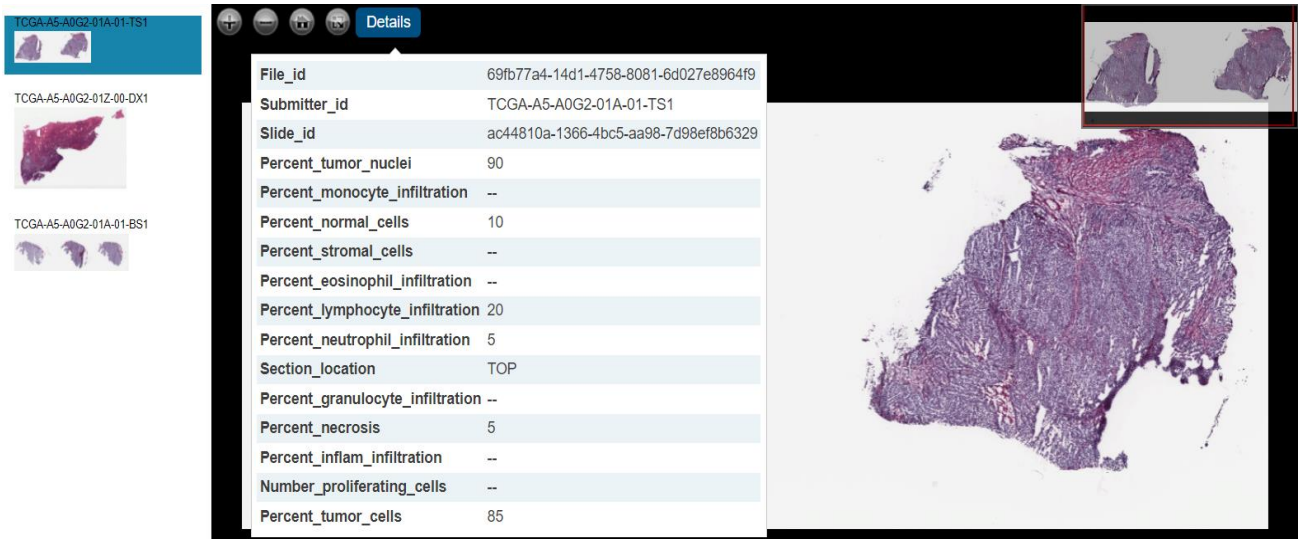

Figure 5.1: An example of histological slide details section from which is possible to retrieve TIL percentage

5.2 DATA VISUALIZATION AND SELECTION

For the visualization of the acquired radiological images, different software have been taken into account and at least the MicroDicom DICOM Viewer was chosen in order to select the images that were more clinically relevant for the study. MicroDicom is an application supported by Windows for primary processing and preservation of medical images in DICOM format: it is equipped with the most common tools for manipulation of DICOM images and it has an intuitive user interface.

All the images acquired from The Cancer Imaging Archive were uploaded on the software and displayed in order to perform a specific selection. The TCGA-BH sagittal vibrant images (Figure 5.2) were chosen for the study.

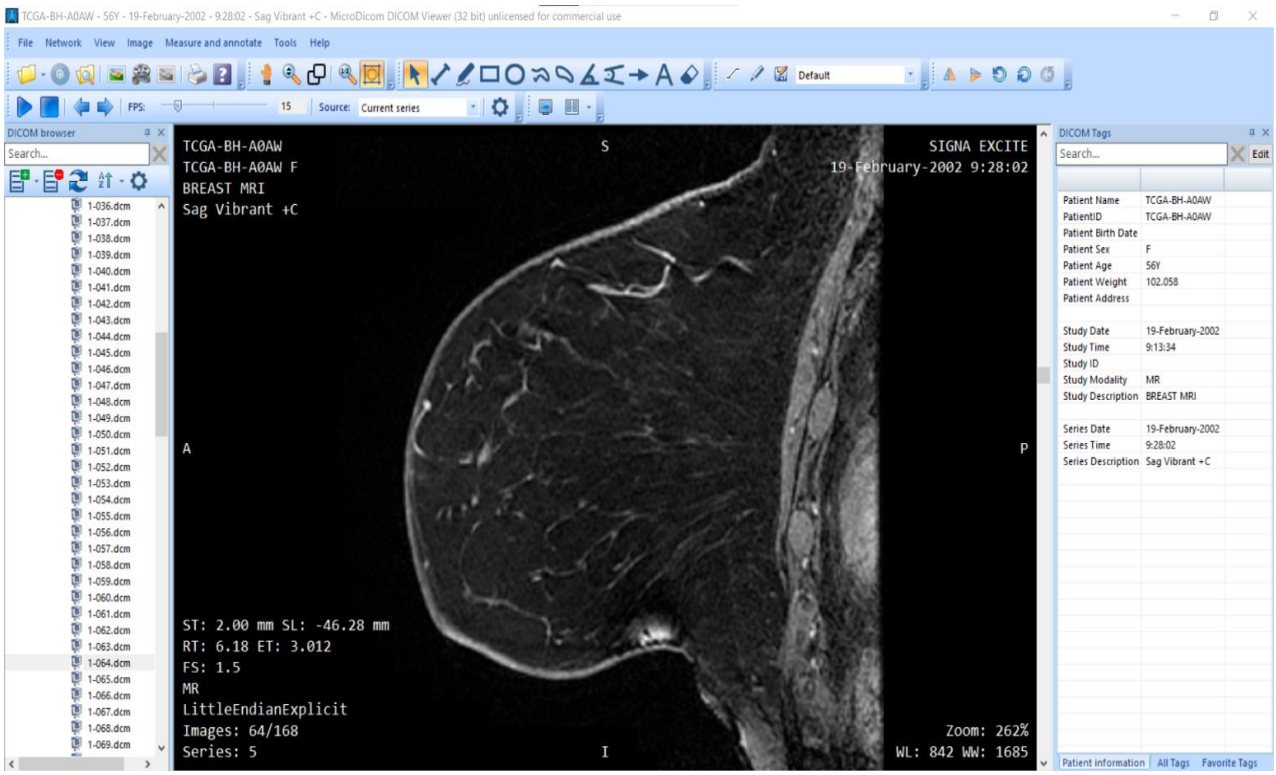

Figure 5.2: An example of one slice of the selected sagittal vibrant breast MRI

Specifically, 'vibrant' stands for simultaneous, high-definition fat-suppressed bilateral breast imaging technique. The sagittal plane shows the best visualization for breast anatomy and tumor detection. After that, the principal characteristics as tumor location (left or right breast), extracted from the Pathology Report, and TIL percentage were collected for each cancer subject and saved on a table in Excel (Table VI). Subsequently, the obtained TCA-BH sagittal vibrant dataset, composed of images related to a total number of 46 patients, was split in two different datasets: one related to patients that have tumor on the left breast (25 patients) and one related to patients with lesion on the right breast (21 patients).

In addition, the original dataset was divided in others two folders: one that has low TIL percentage, as TIL<=10, and one that has High TIL percentage (TIL > 10). Files related to patients that have not a specific value for the TIL percentage were removed as well as additional subfolders that were not useful for the study (8 patients). So finally, 19 files were obtained both for Low and High TIL folders.

5.3 DATA PRE-PROCESSING

After data selection, all the breast MRI images were pre-processed in Colab. Colaboratory, or Colab for short, is an on-cloud product from Google Research. It allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to ML and DL, data analysis and education. Its main advantage is the on-cloud nature. Figure 5.3 shows all the steps followed to perform the dataset pre-processing.

First, all the DICOM data were uploaded from the source folder and converted into the Nifti format. After, all the breast MRI images were subjected to some of the most common pre-processing steps in radiological image analysis: normalization, bias field correction and resampling.

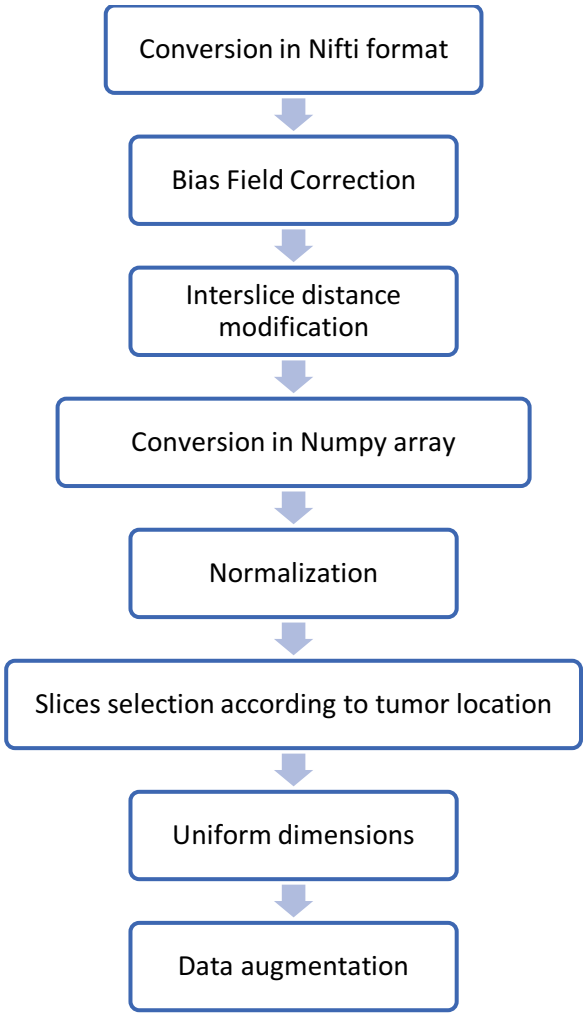

Figure 5.3: Dataset pre-processing flow chart

All the MRIs were first normalized between 0 and 255 to make images more comparable and consistent in terms of image scale, intensity values and orientation.

Then, bias field correction, also known as intensity non-uniformity correction or shading correction, was used to address the problem of non-uniform intensity across each slice. Such inhomogeneity is caused by a combination of factors including variations in the sensitivity of the imaging equipment, imperfections in the magnetic field, patient-specific effects and other technical factors.

Finally, the interslice distance modification was used as resampling technique to standardize the spatial representation of the MRI. The interslice distance was set to 1 mm. The dataset was then converted in Numpy array.

The folders that contains tumor location data were processed and halved in order to remove slices that do not contain useful information for the analysis. According to the direction of the acquisition of the images (from left to right on the sagittal plane) data related to patients that have tumor on the left breast were processed so that only the first half of the data was maintained; instead, for the folder that contains data of patients with right breast tumor, only the second half of data was preserved.

Moreover, in order to get uniform MRIs for the classification, all the images were processed to get the size (256,256,100): a total number of 100 slices for each MRI scan were obtained by performing data augmentation.

Data augmentation is an artificial processing technique that increases the original dataset by creating modified copies of data using the pre-existing data. In this case only geometrical transformations (rotation and scaling) were applied to increase the size and diversity of the original set. The augmentation was made by randomly selecting different slices to augment, in order to avoid image redundancy that could bias the model leading to overfitting. Once obtained a pre-processed and standardized dataset (Figure 5.4), all the images associated with tumor location were grouped in a single folder and finally divided into two distinct subfolders based on TIL percentage (High/Low), as previously outlined in section 5.2. This final step was performed to categorize patient data for classification purposes: the Low TIL dataset was labelled as 0, while the High TIL dataset was labelled as 1.

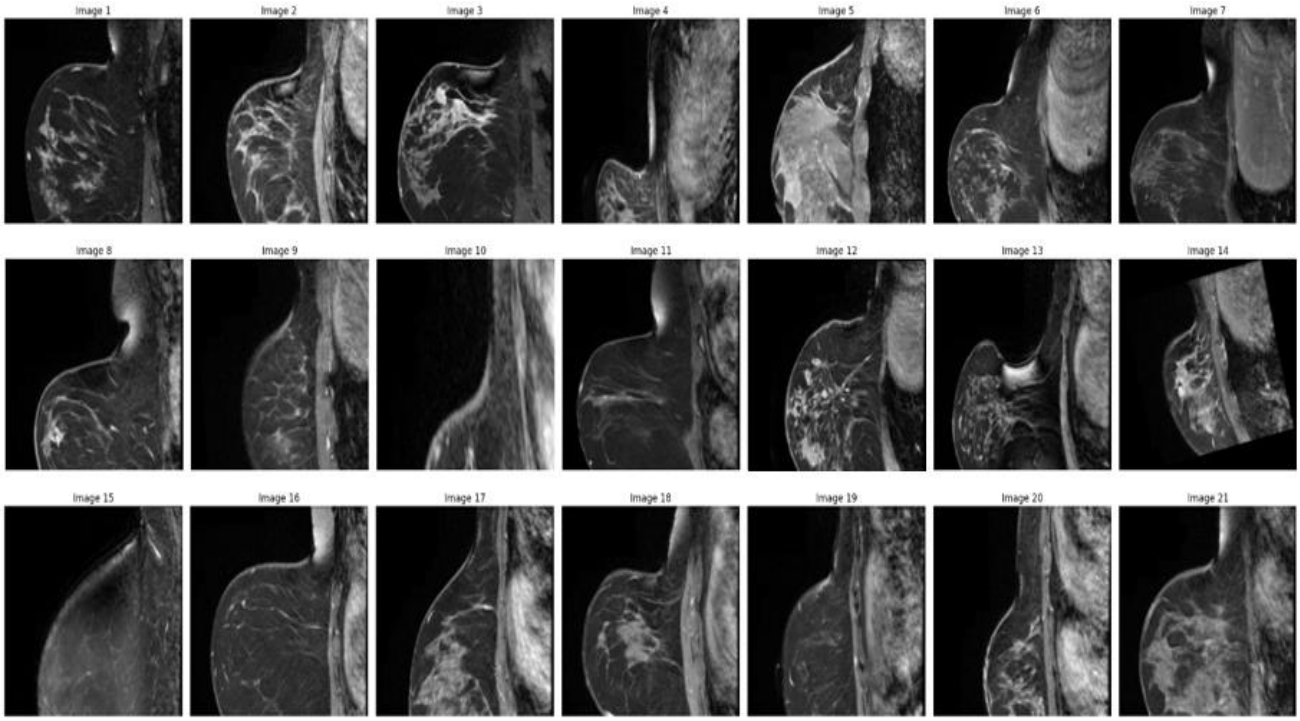

Figure 5.4: Pre-processed MRI images related to patients with tumor on the right breast

5.4 HYPER-PARAMETER TUNING

In general, a DL model has variable parameters, called Hyper-parameters, that are not learned from the data but must be defined before the training process. These values affect the results of the model and its performance. Hyper-parameter tuning is the process of finding the optimal values for these hyper-parameters in order to maximize the model accuracy and efficiency on a specific task. It involves testing various combinations of values and returning the one that produces the best performance. In this study the hyper-parameters tuned are:

- **Learning Rate**: It represents the step size taken by the optimization algorithm during the weight update process. It directly influences the convergence speed and stability of the training process. A higher learning rate may lead to faster convergence but risks overshooting the optimal solution, while a lower rate may improve stability but prolong training.
- **Batch Size**: It determines the number of data samples processed in a single forward and backward pass during each training iteration. It impacts computational efficiency and memory usage. Smaller batch sizes introduce more frequent weight updates but may require more epochs for convergence, while larger batch sizes can accelerate training but might lead to less effective generalization.
- **Dropout Rate**: It is a regularization technique that involves randomly deactivating a fraction of neurons during each training iteration. This helps prevent overfitting by promoting robustness and reducing the network's reliance on individual neurons.

The Hyper-parameter tuning was performed using the Gaussian Prosses (GP) Minimization algorithm. It is a probabilistic optimization technique that is particularly useful when the evaluation of the objective function (in this case, the performance metric of the DL model) is expensive or time-consuming. Instead of exhaustively searching the hyperparameter space, GP Minimization constructs a surrogate model, a Gaussian Process, that approximates the relationship between hyperparameters and the performance metric (e.g., accuracy, loss) of the DL model. The optimization process begins with an initial set of hyperparameter configurations. The objective function is evaluated for each of these configurations; this process involves training and validating the DL model. During the procedure, the algorithm continues to suggest new hyperparameter configurations based on the surrogate model's predictions and uncertainties. Thus, the optimization process continues for a predefined number of iterations or until a stopping criterion is met. This

procedure was executed through 11 trials, for 4 epochs each. The initial set of hyper-parameters was selected by a pre-determined range of values. Specifically, the hyper-parameters tuned were: two dropout rate (dropout1_rate and dropout2_rate), learning rate (learning_rate) and batch size (batch_size). The model was fed with 2D slices resized to dimensions (128,128) to reduce the computational costs. At the end of the tuning process, the optimal values for the hyper-parameters were obtained and selected in order to perform model training and testing.

5.5 MODEL TRAINING AND TESTING

In this study, labelled data were divided into training and testing datasets with a split ratio of 0.3. Since the original dataset was imbalanced, weights associated with each class were calculated and the resulting datasets were used for training and testing. After, K-fold cross-validation was performed to train and test the models. This technique was used to assess the performance and the generalization capabilities of the model avoiding overfitting (that might be caused by the relatively small dataset).

In this study, three experiments were performed: three different models were trained and tested using the selected hyper-parameters. Each model's ability to classify data in High/Low TIL was finally assessed based on accuracy, sensitivity, specificity, F_1-score metrics and ROC curve (AUC).

5.5.1 EXPERIMENT 1

In the first experiment, the VGG-16 model was used to train and test the data. It is a 16-layer DCNN, with 3 fully connected layers and 13 convolutional layers. The outline of the model architecture is further explained in the subchapter 3.4.1. In this case, three splits of 5 epochs each were performed for training and testing the VGG-16 for the classification analysis. The loss function and the optimizer are categorical cross entropy and Stochastic Gradient Descent (SGD), respectively.

5.5.2 EXPERIMENT 2

In the second experiment, a simple CNN was tested. The network has an input layer, followed by a 2D convolutional layer, characterized by a kernel of size (3x3). The convolutional layer is followed by ReLU activation function.

Subsequently, a Max-Pooling layer is added to reduce the spatial dimensions of the image. The output is then flattened into a one-dimensional vector by means of a flattening layer.
After that, a fully connected (dense) layer with 128 units and ReLU activation is added. This layer is followed by two dropout layers used to prevent overfitting during training by randomly dropping out units.
Finally, there is an output layer with 2 units for the binary classification and a Softmax activation layer. The model is compiled using categorical cross entropy. SGD is chosen as optimizer.
The model was trained and tested three times to achieve the best performance on the specific classification task.

5.5.3 EXPERIMENT 3

In the third experiment, training and testing were performed using a CNN that is less complex respect to VGG-16 but, more sophisticated respect to the model outlined in subchapter 5.5.2.
The model architecture (Figure 5.5) presents an input layer, followed by two consecutive convolutional layers, each characterized by 32 filters, a (3x3) kernel size and a ReLU activation. Batch normalization is applied after each convolutional layer, enhancing both convergence and generalization of the model. Subsequently, a Max-Pooling layer with a (2x2) pool size is introduced to down sample the spatial dimensions of the data. Dropout, set at a rate of 0.25, is employed after the Max-Pooling layers to reduce the risk of overfitting. Moreover, two additional convolutional layers are introduced, featuring 64 and 128 filters, respectively, both with a (3x3) kernel size, ReLU activation and batch normalization. Max pooling and dropout layers are added again to enhance the model's learning capabilities. The output from these convolutional layers is flattened into a one-dimensional vector, transitioning to a dense layer with 512 units and a ReLU activation. Following, batch normalization and dropout are implemented. Subsequently, another dense layer with 128 units is introduced, followed by ReLU activation, batch normalization, and dropout. The final layer is a dense layer with 2 units for binary classification and utilizes the Softmax activation for probability distribution over the two classes. The model uses SGD as optimizer. Categorical cross entropy is chosen as the loss function and accuracy is the selected metric for model evaluation. Five splits were performed for this model.

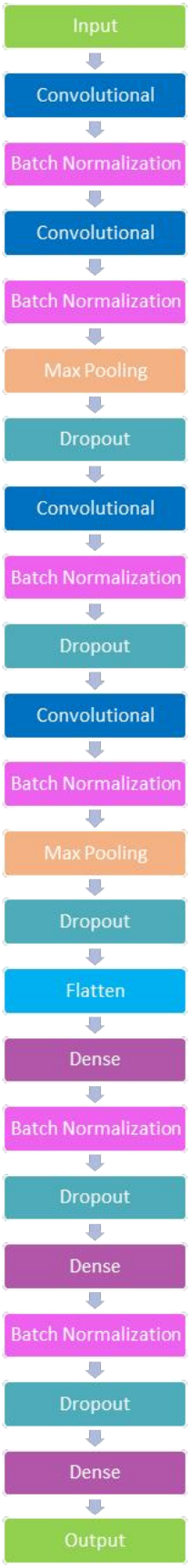

Figure 5.5: Experiment 3 CNN Architecture

6. RESULTS

Once performed all the steps required for the classification analysis and after having trained and tested different models in three experiments, results obtained to classify the processed data into High/Low TIL are as follows.

6.1 RESULTS OF EXPERIMENT 1

Results related to Hyper-parameter tuning are shown in Table VII and the optimal values are reported in Table VIII. Weights associated with each class obtained to get a balanced dataset were 0.863 for Low TIL and 1.187 for High TIL. Model's ability to classify data in all the three splits based on accuracy, sensitivity, specificity and F_1-score metrics are reported in Table IX, while ROC curve (AUC) is shown in Figure 6.1.

Table VII: Hyper-parameter tuning results for each trial and evaluation in terms of validation accuracy

TRIALS	Dropout1_rate	Dropout2_rate	Learning_rate	Batch_size	Validation Accuracy
1	0.5	0.5	0.00010	10	60.02506
2	0.3	0.4	0.00010	23	60.02506
3	0.1	0.4	0.00010	126	60.02506
4	0.3	0.2	0.00100	232	60.02506
5	0.1	0.3	0.00001	35	60.02506
6	0.2	0.7	0.00001	34	60.02506
7	0.5	0.3	0.00100	25	60.02506
8	0.4	0.3	0.00010	86	60.02506
9	0.4	0.6	0.00010	23	60.02506
10	0.2	0.1	0.00010	221	60.02506
11	0.5	0.6	0.00001	28	60.02506

Table VIII: Hyper-parameters optimal values

	Dropout1_rate	Dropout2_rate	Learning_rate	Batch_size
OPTIMAL VALUE	0.5	0.5	0.0001	10

Table IX: Classification evaluation results for each split in terms of accuracy, specificity, sensitivity and F_1-score

	ACCURACY	TIL	SPECIFICITY	SENSITIVITY	F_1-SCORE
SPLIT 1	0.58	HIGH	0.58	1.00	0.73
		LOW	0.00	0.00	0.00
SPLIT 2	0.58	HIGH	0.58	1.00	0.73
		LOW	0.00	0.00	0.00
SPLIT 3	0.58	HIGH	0.58	1.00	0.73
		LOW	0.00	0.00	0.00

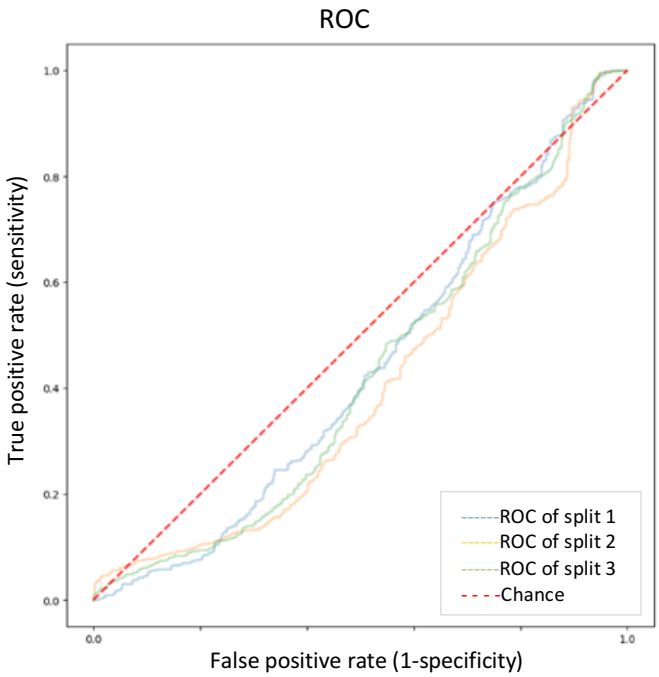

Figure 6.1: ROC curve (AUC) related to the three splits performed

6.2 RESULTS OF EXPERIMENT 2

Results related to Hyper-parameter tuning are shown in Table X and the optimal values are reported in Table XI. Weights associated with each class obtained to get a balanced dataset were 0.863 for Low TIL and 1.187 for High TIL. Model's ability to classify data in all the three splits based on accuracy, sensitivity, specificity and F_1-score metrics are reported in Table XII, while ROC curve (AUC) is shown in Figure 6.2.

Table X: Hyper-parameter tuning results for each trial and evaluation in terms of validation accuracy

TRIALS	Dropout1_rate	Dropout2_rate	Learning_rate	Batch_size	Validation Accuracy
1	0.5	0.5	0.000100	10	59.774435
2	0.8	0.2	0.000100	11	59.774435
3	0.3	0.4	0.000100	85	59.774435
4	0.3	0.6	0.000100	250	59.774435
5	0.4	0.8	0.000100	203	59.774435
6	0.2	0.3	0.000100	172	59.774435
7	0.2	0.2	0.000100	23	59.774435
8	0.7	0.6	0.000010	43	59.774435
9	0.7	0.3	0.000001	31	48.308271
10	0.8	0.7	0.000001	12	47.368422
11	0.3	0.5	0.001000	220	59.774435

Table XI: Hyper-parameters optimal values

	Dropout1_rate	Dropout2_rate	Learning_rate	Batch_size
OPTIMAL VALUE	0.5	0.5	0.0001	10

Table XII: Classification evaluation results for each split in terms of accuracy, specificity, sensitivity and F_1-score

	ACCURACY	TIL	SPECIFICITY	SENSITIVITY	F_1-SCORE
SPLIT 1	0.58	HIGH	0.58	1.00	0.73
		LOW	0.00	0.00	0.00
SPLIT 2	0.58	HIGH	0.58	1.00	0.73
		LOW	0.00	0.00	0.00
SPLIT 3	0.58	HIGH	0.58	1.00	0.73
		LOW	0.00	0.00	0.00

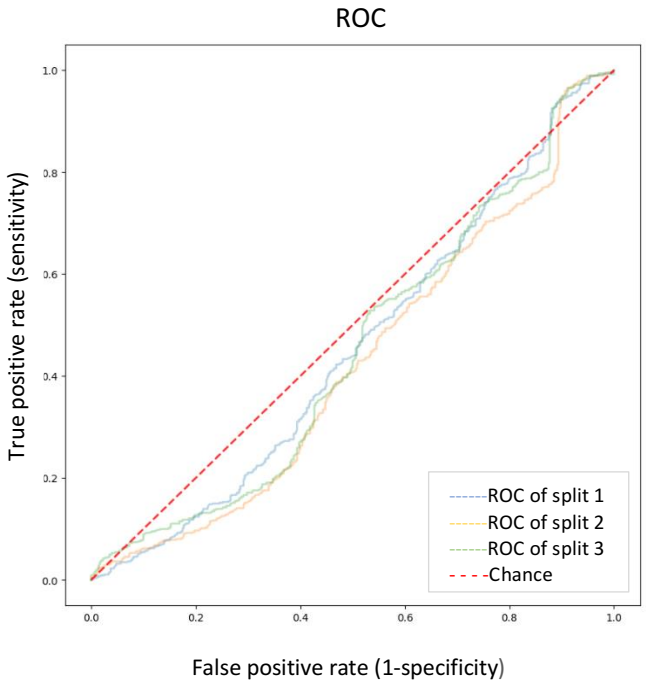

Figure 6.2: ROC curve (AUC) related to the three splits performed

6.3 RESULTS OF EXPERIMENT 3

Results related to Hyper-parameter tuning are shown in Table XIII and the optimal values are reported in Table XIV. Weights associated with each class obtained to get a balanced dataset were 0.863 for Low TIL and 1.187 for High TIL. Model's ability to classify data in all the five splits based on accuracy, AUC, sensitivity, specificity and F_1-score metrics are reported in Table XV, while ROC curve is shown in Figure 6.3.

Table XIII: Hyper-parameter tuning results for each trial and evaluation in terms of validation accuracy

TRIALS	Dropout1_rate	Dropout2_rate	Learning_rate	Batch_size	Validation Accuracy
1	0.5	0.5	0.00010	10	71.804512
2	0.1	0.2	0.00010	26	41.165414
3	0.5	0.3	0.00001	247	40.225562
4	0.5	0.7	0.00001	107	40.225562
5	0.6	0.7	0.01000	111	40.225562
6	0.6	0.6	0.00010	192	59.774435
7	0.2	0.1	0.00001	19	57.706767
8	0.5	0.4	0.00100	247	40.225562
9	0.2	0.2	0.00001	63	40.225562
10	0.2	0.2	0.01000	129	40.225562
11	0.2	0.7	0.00001	163	59.774435

Table XIV: Hyper-parameters optimal values

	Dropout1_rate	Dropout2_rate	Learning_rate	Batch_size
OPTIMAL VALUE	0.5	0.5	0.0001	10

Table XV: Classification evaluation results for each split in terms of accuracy, AUC, specificity, sensitivity and F_1-score

	ACCURACY	AUC	TIL	SPECIFICITY	SENSITIVITY	F_1-SCORE
SPLIT 1	0.85	0.9431	HIGH	0.84	0.92	0.88
			LOW	0.88	0.76	0.81
SPLIT 2	0.86	0.9625	HIGH	0.97	0.79	0.87
			LOW	0.77	0.96	0.86
SPLIT 3	0.75	0.9905	HIGH	0.98	0.59	0.73
			LOW	0.63	0.98	0.77
SPLIT 4	0.88	0.9852	HIGH	0.99	0.81	0.89
			LOW	0.79	0.99	0.88
SPLIT 5	0.95	0.9886	HIGH	0.97	0.93	0.95
			LOW	0.91	0.96	0.94

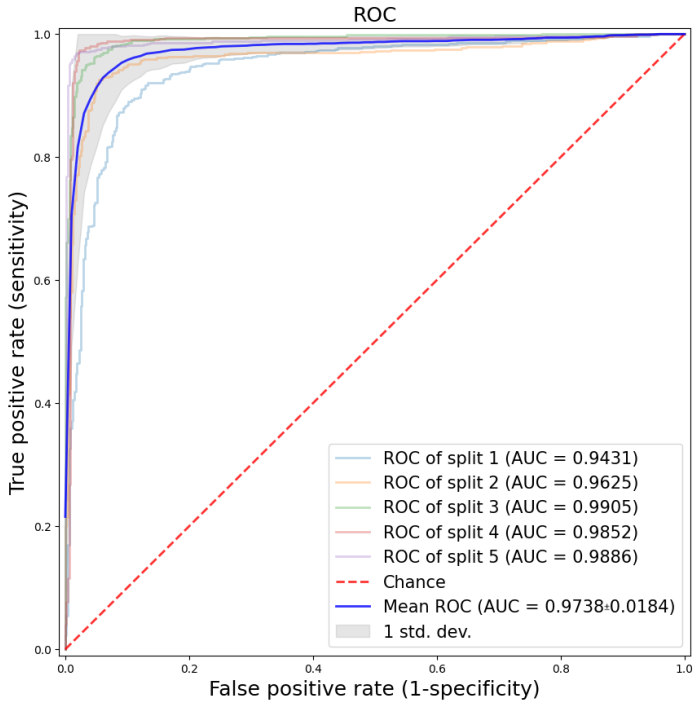

Figure 6.3: ROC curve (AUC) related to the five splits performed

DISCUSSION AND CONCLUSION

Based on the analysis presented in this study, it was possible to assess TIL from MRI breast cancer images with good accuracy in Experiment 3. MRI Images acquired from TCIA were initially pre-processed to obtain a standardized and uniformed dataset with dimensions (256, 256, 100). The folders that contains tumor location data were halved in order to remove slices that not contain useful information for the analysis: this step was crucial because it allowed the classifier to focus only on images where the tumor was present, thereby facilitating its identification. Subsequently, the study relied on three experiments performed using three different classifiers: VGG-16 and two simpler CNNs, one with a single convolutional layer and the other with four convolutional layers. Before executing training and testing for each specific model, all images were resized to have 2D slices with dimensions (128, 128).

In each experiment, the algorithm iteratively performed hyper-parameter tuning for eleven trials. Subsequently, weights associated with each class were calculated because the original dataset resulted to be imbalanced. Considering the results obtained from the three experiments, the VGG-16 model proved to be too much complex. In fact, by changing iteratively the parameters set in each trial during hyper-parameter tuning, the performance in terms of validation accuracy remains exactly the same (Table VII). This became evident during model training and testing, where the classifier yielded values close to 0 for the Low TIL class in terms of specificity, sensitivity, and F_1-score (Table IX). So, the model was not able to predict the 'Low TIL' class at all. In addition, the ROC curve, that represents the trade-off between sensitivity (True Positive Rate) and specificity (True Negative Rate), was found to be close to the diagonal. This indicates that the model was not able to generalize and discriminate between classes because the number of trainable parameters was too high with respect to the number of breast MRI images used for training.

In the second experiment, the model which is significantly simpler, achieved outcomes very similar to the ones associated to VGG-16.

In fact, results from hyper-parameter tuning in terms of validation accuracy remain more or less unchanged across calls (Table X) and, after training and testing, the classifier showed to be not able to discriminate between the two classes, obtaining the same values for accuracy, specificity, sensitivity, F1 score (Table XII) and ROC curve of the previous model.

In the third experiment instead, the model, characterized by an intermediate level of complexity compared to the first two, yielded good results.

This was evident as observed among all the eleven tuning trials: the results exhibit considerable variability in terms of validation accuracy (Table XIII). Moreover, during both training and testing, the model shows good results in distinguishing between the two classes, obtaining values for accuracy, AUC, sensitivity, specificity and F1-score close to 1 (Table XV). This was further confirmed by the ROC curve, which was significantly distant from the diagonal.

In general, this study obtained positive outcomes, demonstrating that the model with intermediate level of complexity emerged as the most reliable for classifying the available dataset into High/Low TIL.

In scientific literature, there are few sources regarding the prediction of therapy outcomes in breast cancer combined with DL applications, especially concerning TILs. For this reason, the literature search was performed focusing more in general on the evaluation of therapeutic efficacy in different types of cancers through ML and DL techniques. Most of the studies focus on the prediction of therapy response in a 'blind manner'.

DL models are simply fed with MRI and CT images and predict the response without searching for some motivations. This study represents an innovative approach because it is the first one that estimate a parameter quantifiable from histological images by using MRIs.

Moreover, the classifier was able to achieve good results even though tumor segmentation was not applied to the breast cancer images.

However, the findings of this study have to be seen in light of some limitations. Firstly, all the breast cancer MRI images were obtained from the same institution, leading to the issue of variability.

In the classification analysis, the original dataset was highly heterogeneous and too small for the purpose of the study. Image pre-processing required extensive work, and data augmentation was necessary to obtain a sufficiently informative dataset.

In addition, the study for the assessment of TIL has been performed on 2D MRI breast cancer images. So, in future, further investigations, as 3D analysis, will be required to preserve volumetric information of the tumor. This is an important aspect because the future application of DL techniques on cancer radiological images for the TIL assessment could make a positive impact on clinical applications and enhance patients care.

BIBLIOGRAPHY

[1] Akram, M., Iqbal, M., Daniyal, M. et al. Awareness and current knowledge of breast cancer. Biol Res. 2017

[2] Kornelia Poyak J, Breast cancer: origins and evolution. Clin Invest. 2007

[3] SEER, Breast Cancer Module Map. U. S. National Institutes of Health. National Cancer Institute. 2021

[4] Khan YS, Sajjad H. Anatomy, Thorax, Mammary Gland. StatPearls. Treasure Island (FL): StatPearls Publishing. 2023

[5] Gusterson B, Warburton MJ, Mitchell D, Ellison M, Neville AM, Rudland PS. Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases. Cancer Res. 1982

[6] Evan G, Vousden K. Proliferation, cell cycle and apoptosis in cancer. Nature. 2001

[7] Hartwell L, Kastan M. Cell cycle control and cancer. Sci. 1994

[8] Leighton J, Kalla R, Turner J, Fennell R. Pathogenesis of tumor invasion. Cancer Res. 1960

[9] Jarde T, Perrier S, Vasson M, Caldefie-Chezet F. Molecular mechanisms of leptin and adiponectin in breast cancer. Eur J Cancer. 2011

[10] Hanahan D, Weinberg R. The hallmarks of cancer. Cell. 2000

[11] Jain R. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science. 2005

[12] West A, Wullkopf L, Christensen A, Leijnse N, Tarp JM, Mathiesen J, et al. Division induced dynamics in non-Invasive and invasive breast cancer. Biophys J. 2017

[13] Cserni, Gábor & Chmielik, Ewa & Cserni, Bálint & Tot, Tibor. The new TNM-based staging of breast cancer. Virchows Archiv. 2018

[14] Bednarek A, Sahin A, Brenner A, Johnston D, Aldaz C. Analysis of telomerase activity levels in breast cancer: positive detection at the in situ breast carcinoma stage. Clin Cancer Res. 1997

[15] Jacquillat C, Weil M, Baillet F, Borel C, Auclerc G, Maublanc M. Results of neoadjuvant chemotherapy and radiation therapy in the breast-conserving treatment of 250 patients with all stages of infiltrative breast cancer. Cancer. 1990

[16] Neuman H, Morrogh M, Gonen M. Stage IV breast cancer in the Era of targeted therapy, Does surgery of the primary tumor matter. Cancer. 2015

[17] S. Demaria, E. Pikarsky, M. Karin, et al. Cancer and inflammation: promise for biologic therapy.

J Immunother. 2010

[18] R. Salgado, C. Denkert, S. Demaria, N. Sirtaine et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. 2015

[19] Baisheng Lin, Likun Du, Hongmei Li, Xiao Zhu, Liao Cui, Xiaosong Li, Tumor-infiltrating lymphocytes: Warriors fight against tumors powerfully. Biomedicine & Pharmacotherapy. 2020

[20] Carsten Denkert, Sibylle Loibl, Aurelia Noske et al. Tumor-Associated Lymphocytes as an Independent Predictor of Response to Neoadjuvant Chemotherapy in Breast Cancer. Journal of Clinical Oncology. 2010

[21] Stanton SE, Disis MLClinical significance of tumor-infiltrating lymphocytes in breast cancer. Journal for ImmunoTherapy of Cancer. 2016

[22] Gu-Trantien C, Loi S, Garaud S, Equeter C et al. $CD4^+$ follicular helper T cell infiltration predicts breast cancer survival. J Clin Invest. 2013

[23] Mahdiyeh Shamsi, Jalil Pirayesh Islamian, Breast cancer: early diagnosis and effective treatment by drug delivery tracing. Nucl. Med. Rev 2017

[24] Bhushan A, Gonsalves A, Menon JU. Current State of Breast Cancer Diagnosis, Treatment, and Theranostics. Pharmaceutics. 2021

[25] Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 200

[26] Mainiero MB, Moy L, Baron P, Didwania AD, diFlorio RM, Green ED, Heller SL, et al. Expert Panel on Breast Imaging. ACR Appropriateness Criteria Breast Cancer Screening. J Am Coll Radiol. 2017

[27] Berg WA. Tailored supplemental screening for breast cancer: what now and what next? AJR. 2009

[28] Graves, M.J., Zhu, C. Basic Principles of Magnetic Resonance Imaging. In 3D Imaging Technologies in Atherosclerosis. Springer: Boston, USA. 2015

[29] Torrisi, Restuccia, et al. Study of gold nanoparticles for mammography diagnostic and radiotherapy improvements. Rep. Pract. Oncol. Radiother. J. Greatpoland Cancer Cent. Pozn. Pol. Soc. Radiat. Oncol. 2019

[30] M. Sehgal et al. A review of breast ultrasound. J Mammary Gland Biol. Neoplasia. 2006

[31] Rahbar, Partridge, S.C. Multiparametric MR Imaging of Breast Cancer. Magn. Reson. Imaging Clin. North Am. 2016

[32] Lee, Cho, Kim, Cha, Cho, Ko,Moon. Correlation between high resolution dynamic MR features and prognostic factors in breast cancer. Korean J. Radiol.,2008

[33] Malayeri, El Khouli, Zaheer, Jacobs, Corona-Villalobos,Kamel, Macura. Principles and Applications of Diffusion-weighted Imaging in Cancer Detection, Staging, and Treatment Follow-up. RadioGraphics .2011

[34] Durur-Subasi, I. DW-MRI of the breast: A pictorial review. Insights Into Imaging. 2019

[35] Glaser, Manduca, Ehman, R.L. Review of MR elastography applications and recent developments. J. Magn. Reson. Imaging 2012

[36] Hawley, Kalra, Mo, Raterman, Yee, Kolipaka, A. Quantification of breast stiffness using MR elastography at 3 Tesla with a soft sternal driver: A reproducibility study. J. Magn. Reson. Imaging 2017

[37] Bolan, Meisamy, Baker, Lin, Emory, Nelson, Everson, Yee, Garwood, In vivo quantification of choline compounds in the breast with 1H MR spectroscopy. Magn. Reson. Med. 2003

[38] Jaglan, P., Dass, R. & Duhan, M. Breast Cancer Detection Techniques: Issues and Challenges. J. Inst. Eng. India Ser. 2019

[39] Bénard, F., Turcotte, É. Imaging in breast cancer: Single-photon computed tomography and positron-emission tomography. Breast Cancer Res. 2005

[40] Zhang, Wang, Wang, Yu, Zhu, Zhan, Chen, Li, Huang, J. Breast-specific gamma imaging or ultrasonography as adjunct imaging diagnostics in women with mammographically dense breasts. Eur. Radiol. 2020

[41] Candelaria, R., Fornage, B.D. Second-look US examination of MR-detected breast lesions. J. Clin. Ultrasound. 2011

[42] Ritse M. Mann, Nariya Cho, and Linda Moy, Breast MRI: State of the Art. Radiology. 2019

[43] Yang, Q., Li, N., Zhao, Z. et al. MRI Cross-Modality Image-to-Image Translation. Sci Rep. 2020

[44] Ortendahl, D.A., Hylton, N.M., Kaufman, L.Tissue Characterization with MRI: The Value of the MR Parameters. Higer, H.P., Bielke, G, Berlin. 1990

[45] Kawahara D, Nagata Y. T1-weighted and T2-weighted MRI image synthesis with convolutional generative adversarial networks. Rep Pract Oncol Radiother. 2021

[46] Tofts PS, Berkowitz B, Schnall MD. Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumors using a permeability model. Magn. Reson. Med. 1995

[47] Kinkel, K. and Hylton, N.M., Challenges to interpretation of breast MRI. J. Magn. Reson. 2001

[48] Machida Y, Shimauchi A, Kuroki Y, et al. Single focus on breast magnetic resonance imaging: diagnosis based on kinetic pattern and patient age. Acta Radiol. 2017

[49] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015

[50] Leung, M. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. Deep learning of the tissue-regulated splicing code. Bioinformatics. 2014

[51] Sutskever, I. Vinyals, O. & Le. Q. V. Sequence to sequence learning with neural networks. Advances in Neural Information Processing Systems. 2014

[52] Deng, L., Yu, D. "Deep Learning: Methods and Applications". Foundations and Trends in Signal Processing. 2014

[53] Pariwat Ongsulee, Artificial Intelligence, Machine Learning and Deep Learning, Fifteenth International Conference on ICT and Knowledge Engineering. Department of Computer Science Faculty of Science, Siam University. 2017

[54] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press. 2016

[55] Tal Linzen and Marco Baroni. Syntactic Structure from Deep Learning. Annual Review of Linguistics. 2021

[56] Simeon Kostadinov. Understanding Backpropagation Algorithm. Towards Data Science. 2019

[57] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma et al. Recent advances in convolutional neural networks. Pattern Recognition. 2018

[58] Mayank Mishra. Convolutional Neural Networks. Explained. Towards Data Science. 2020

[59] MK Gurucharan. Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network. 2022

[60] Dilovan Asaad Zebari, Dheyaa Ahmed Ibrahim at al. Systematic Review of Computing Approaches for Breast Cancer Detection Based Computer Aided Diagnosis Using Mammogram Images. Applied Artificial Intelligence. 2021

[61] Wang L, Zhang M, He G, Shen D, Meng M. Classification of Breast Lesions on DCE-MRI Data Using a Fine-Tuned MobileNet. Diagnostics. 2023

[62] Nam, Sang & Ko, Eun Sook & Lim, Yaeji & Han, Boo-Kyung & Ko, Eun & Choi, Ji & Lee, Jeong. Preoperative dynamic breast magnetic resonance imaging kinetic features using computer-aided diagnosis: Association with survival outcome and tumor aggressiveness in patients with invasive breast cancer. Department of Radiology, Gachon University of Medicine. 2018

[63] Song, S.E., Seo, B.K., Cho, K.R. et al. Computer-aided detection (CAD) system for breast MRI in assessment of local tumor extent, nodal status, and multifocality of invasive breast cancers: preliminary study. Cancer imaging. 2015

[64] Yin, X., Yin, L.; Hadjiloucas, S. Pattern Classification Approaches for Breast Cancer Identification via MRI: State-Of-The-Art and Vision for the Future. Appl. Sci. 2020

[65] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, Computer Vision and Pattern Recognition. 2015

[66] Jing Lu, Yan Wu, Mingyan Hu et al., Breast Tumor Computer-Aided Detection System Based on Magnetic Resonance Imaging Using Convolutional Neural Network Department of Radiology. Fourth Medical Center of PLA General Hospital. Beijing. China. 2021

[67] Roberto Lo Gullo, Sarah Eskreis-Winkler, Elizabeth A. Morris, Katja Pinker. Machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy. Department of Radiology. 2019

[68] Therasse P. Measuring the clinical response. What does it mean? Eur J Cancer. 2002

[69] F. Pesapane, M. Codari, F. Sardanelli. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 2018

[70] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, and S.E. Brennan. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. 2020.

[71] S. S. Alahmari, D. Cherezov, D. B. Goldgof, L. O. Hall, R. J. Gillies and M. B. Schabath. Delta Radiomics Improves Pulmonary Nodule Malignancy Prediction in Lung Cancer Screening. IEEE Access. 2018

[72] Bellini, D., Carbone, I., Rengo, M., Vicini, S., Panvini, N., Caruso, D., Iannicelli, E., Tombolini, V., Laghi. A. Performance of Machine Learning and Texture Analysis for Predicting Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer with 3T MRI. Tomography. 2022

[73] Cain, E.H., Saha, A., Harowicz, M.R. et al. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. Breast Cancer Res Trea. 2019

[74] Marco Caballo, Wendelien B. G. Sanderink PhD, Luyi Han MSc, Yuan Gao MSc, Alexandra Athanasiou MD, MSc, Ritse M. Mann MD. Four-Dimensional Machine Learning Radiomics for the Pretreatment Assessment of Breast Cancer Pathologic Complete Response to Neoadjuvant Chemotherapy in Dynamic Contrast-Enhanced MRI. Journal of Magnetic resonance imaging. 2022

[75] Cha, K.H., Hadjiiski, L., Chan, HP. et al. Bladder Cancer Treatment Response Assessment in CT using Radiomics with Deep-Learning. Sci Rep. 2017.

[76] Chang, R., Qi, S., Wu, Y. et al. Deep multiple instance learning for predicting chemotherapy response in non-small cell lung cancer using pretreatment CT images. Sci Rep. 2022

[77] Na Lae Eun, Daesung Kang, Eun Ju Son, Jeong Seon Park et al, Texture Analysis with 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. Department of Radiology. Gangnam Severance Hospital. 2019

[78] Jin, C., Yu, H., Ke, J. et al. Predicting treatment response from longitudinal images using multi-task deep learning. Nat Commun. 2021

[79] Langenhuizen, Patrick P. J. H., Zinger, Svetlana; Leenstra, Sieger; Kunst, Henricus P. M.; et al. Radiomics-Based Prediction of Long-Term Treatment Response of Vestibular Schwannomas Following Stereotactic Radiosurgery. Otology & Neurotology. 2020

[80] Lu, L., Dercle, L., Zhao, B. et al. Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging. Nat Commun. 2021.

[81] Maaref, A., Romero, F.P., Montagnon, E. et al. Predicting the Response to FOLFOX-Based Chemotherapy Regimen from Untreated Liver Metastases on Baseline CT: a Deep Neural Network Approach. J Digit Imaging. 2020

[82] Mehta R, Cai K, Kumar N, Knuttinen MG, Anderson TM, Lu H, Lu Y. A Lesion-Based Response Prediction Model Using Pretherapy PET/CT Image Features for Y90 Radioembolization to Hepatic Malignancies. Technol Cancer Res Treat. 2017

[83] Moghadas-Dastjerdi, H., Rahman, S.-E.-T.H., Sannachi, L., Wright, F.C., Gandhi, S., Trudeau, M.E., Sadeghi-Naini, A., Czarnota, G.J. Prediction of chemotherapy response in breast cancer patients at pre-treatment using second derivative texture of CT images and machine learning. Department of Medical Biophysics. University of Toronto. 2021

[84] Moreau, Rousseau, Fourcade, Santini, Brennan, Ferrer, Lacombe, Guillerminet, Colombié, Jézéquel, et al. Automatic Segmentation of Metastatic Breast Cancer Lesions on ^{18}F-FDG PET/CT Longitudinal Acquisitions for Treatment Response Assessment. Cancers. 2022

[85] Nasief, H., Zheng, C., Schott, D. et al. A machine learning based delta-radiomics process for early prediction of treatment response of pancreatic cancer. Precis. Onc. 2019

[86] Pang Xiaolin, Wang Fang, Zhang Qianru, Li Yan, Huang Ruiyan, Yin Xinke, Fan Xinjuan. A Pipeline for Predicting the Treatment Response of Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer Using Single MRI Modality: Combining Deep Segmentation Network and Radiomics Analysis Based on “Suspicious Region”. Front. Oncol. 2021

[87] Peng Jie, Huang Jinhua, Huang Guijia, Zhang Jing. Predicting the Initial Treatment Response to Transarterial Chemoembolization in Intermediate-Stage Hepatocellular Carcinoma by the Integration of Radiomics and Deep Learning. Front. Oncol. 2021

[88] Peng, J., Kang, S., Ning, Z. et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. Eur Radiol,2020.

[89] Peng, Chen, Tao, Liu, Zhu, Liu, Yang, F. Radiomics Analysis of Multi-Phase DCE-MRI in Predicting Tumor Response to Neoadjuvant Therapy in Breast Cancer. Diagnostics 2021

[90] Lucie Petrova, Panagiotis Korfiatis, Ondra Petr, Daniel H., LaChance, Ian Parney, Jan C. Buckner, Bradley, J. Erickson. Cerebral blood volume and apparent diffusion coefficient - Valuable predictors of non-response to bevacizumab treatment in patients with recurrent glioblastoma. Journal of Neurogical Science. 2019

[91] Qu, H., Zhai, H., Zhang, S., Chen, W., Zhong, H., Cui, et al. Dynamic radiomics for predicting the efficacy of antiangiogenic therapy in colorectal liver metastases. College of Medicine and Biological Information Engineering. Northeastern University. 2023

[92] Sharaby, Alksas, Nashat, Balaha, Shehata, Gayhart, Mahmoud, Ghazal,Khalil, Abouelkheir, et al. Prediction of Wilms' Tumor Susceptibility to Preoperative Chemotherapy Using a Novel Computer-Aided Prediction System. Diagnostics. 2023

[93] Yan, M., Wang, W. Radiomic Analysis of CT Predicts Tumor Response in Human Lung Cancer with Radiotherapy. J Digit Imaging. 2020

[94] Yoo, Lee, J, Cheon, M., Woo, et al. Predictive Value of 18F-FDG PET/CT Using Machine Learning for Pathological Response to Neoadjuvant Concurrent Chemoradiotherapy in Patients with Stage III Non-Small Cell Lung Cancer. Cancers. 2022

[95] Zhigang Yua, Marissa Fraze, Anupam Rish, Kujtim Latifi, Pretreatment CT and PET Radiomics Predicting Rectal Cancer Patients in Response to Neoadjuvant Chemoradiotherapy. Rep. Pract. Oncol. Radiother. 2021

[96] Peng Sun, Jiehua He, Xue Chao, Keming Chen et al. A Computational Tumor-Infiltrating Lymphocyte Assessment Method Comparable with Visual Reporting Guidelines for Triple-Negative Breast Cancer. Department of Pathology, Sun Yat-sen University Cancer Center. China. 2021

[97] Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, Wienert S,Van den Eynden G et al. Evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann. Oncol. 2015

APPENDIX

Table VI: Collection of the principal cancer subject characteristics for the analysis. The rows highlighted in yellow belong to those patients which are excluded because of missing TIL information.

PATIENT	TUMOR LOCATION	% TIL
TCGA-BH-A0AW	Left	20
TCGA-BH-A0AZ	Right	2
TCGA-BH-A0B1	Right	-
TCGA-BH-A0B3	Right	15
TCGA-BH-A0B5	Left	2
TCGA-BH-A0B6	Left	35
TCGA-BH-A0BG	Left	40
TCGA-BH-A0BJ	Left	26
TCGA-BH-A0BM	Right	10
TCGA-BH-A0BQ	Left	30
TCGA-BH-A0BT	Right	5
TCGA-BH-A0C0	Left	30
TCGA-BH-A0DE	Left	5
TCGA-BH-A0DG	Right	20
TCGA-BH-A0DH	Left	3
TCGA-BH-A0DI	Right	10
TCGA-BH-A0DK	Left	15
TCGA-BH-A0DT	Right	22
TCGA-BH-A0DV	Left	7
TCGA-BH-A0DX	Left	40
TCGA-BH-A0DZ	Right	5
TCGA-BH-A0E0	Left	5
TCGA-BH-A0E1	Left	8
TCGA-BH-A0E2	Right	2
TCGA-BH-A0E9	Right	1
TCGA-BH-A0EI	Right	25
TCGA-BH-A0GY	Left	-
TCGA-BH-A0GZ	Left	-
TCGA-BH-A0H3	Left	20
TCGA-BH-A0H5	Right	50
TCGA-BH-A0H6	Right	7
TCGA-BH-A0H7	Right	36
TCGA-BH-A0H9	Right	18
TCGA-BH-A0HA	Right	12
TCGA-BH-A0HI	Right	-
TCGA-BH-A0HX	Left	4
TCGA-BH-A0HY	Left	3
TCGA-BH-A0RX	Left	3
TCGA-BH-A0W3	Left	-
TCGA-BH-A0W5	Left	100
TCGA-BH-A18F	Right	3
TCGA-BH-A18H	Right	10
TCGA-BH-A18I	Left	-
TCGA-BH-A28Q	Right	40
TCGA-BH-A201	Left	-
TCGA-BH-A202	Left	-

RINGRAZIAMENTI

Desidero ringraziare in primis il relatore, il correlatore di questa tesi e la mia tutor, per la disponibilità e per avermi guidato ed aiutato nella realizzazione dell'elaborato.
Ringrazio ulteriormente la Dottoressa Berardi del reparto di Oncologia dell'ospedale Torrette di Ancona per la collaborazione.

Grazie ai miei genitori, che da sempre mi supportano e mi sostengono in ogni mio percorso e scelta di vita, senza di loro non sarei arrivata fin qui.

Grazie alle mie coinquiline, alle amiche di sempre, agli amici e a tutti coloro che mi sono stati vicino, nei momenti di gioia e spensieratezza, ma anche in quelli di difficoltà.

Grazie a chi non c'è più, ovunque tu sia, spero di averti reso orgoglioso di me.

E, per finire,
Grazie a me.
Nonostante il periodo in continua lotta con me stessa,
la perseveranza, la costanza e la tenacia mi hanno permesso di raggiungere questo traguardo.